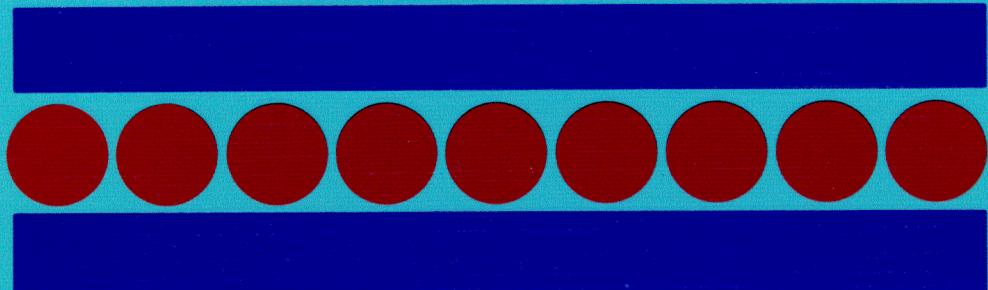


NAZIOARTEKO ESTATISTIKA  
MINTEGIA EUSKADIN

1992

SEMINARIO INTERNACIONAL  
DE ESTADISTICA EN EUSKADI

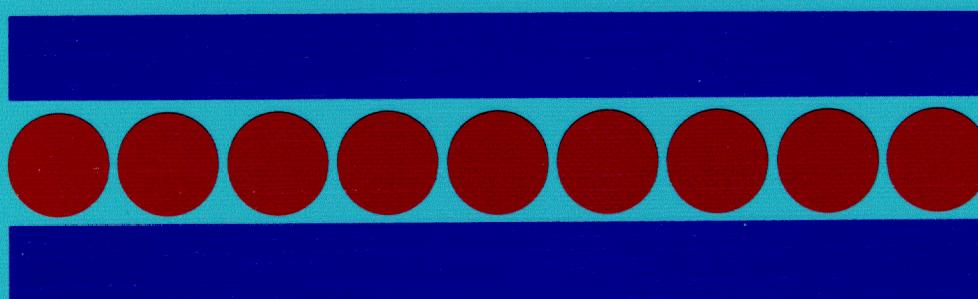


## METHODOLOGICAL ISSUES IN FAMILY EXPENDITURE SURVEYS

### GASTU FAMILIARREI BURUZKO INKESTETAKO ALDERDI METODOLOGIKOAK

### ASPECTOS METODOLOGICOS EN LAS ENCUESTAS DE GASTOS FAMILIARES

MALKA KANTOROWITZ



NAZIOARTEKO ESTATISTIKA  
MINTEGIA EUSKADIN

**1992**

SEMINARIO INTERNACIONAL  
DE ESTADISTICA EN EUSKADI

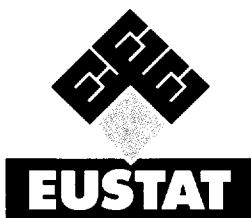
**METHODOLOGICAL ISSUES IN FAMILY  
EXPENDITURE SURVEYS**

**GASTU FAMILIARREI BURUZKO  
INKESTETAKO ALDERDI  
METODOLOGIKOAK**

**ASPECTOS METODOLOGICOS EN LAS  
ENCUESTAS DE GASTOS FAMILIARES**

**MALKA KANTROWITZ**

**CUADERNO 25 KOADERNOA**



**Lanketa / Elaboración:**

Euskal Estatistika-Erakundea /  
Instituto Vasco de Estadística

**Argitalpena / Edición:**

Euskal Estatistika-Erakundea /  
Instituto Vasco de Estadística  
C/Dato 14-16 01005 Vitoria-Gasteiz  
© Euskadiko K.A.ko Administrazioa  
Administración de la C.A. de Euskadi

**Botaldia / Tirada**

1.000 ejemplares  
X-1992

**Inprimaketa eta koadernaketa /**

Impresión y encuadernación:  
ITXAROPENA, S.A.  
Araba kalea, 45 - Zarautz (Gipuzkoa)

**Lege-gordailua / Depósito legal: S.S. 878/92**

ISBN: 84-7542-127-X Obra completa  
ISBN: 84-7749-124-0

## AURKEZPENA

Estatistikako Mintegi Internazionalak sustatzean, hainbat xederekin bete nahi luke Euskal Estatistika-Erakundeak, hala nola:

- Unibertsitatearekiko eta, Estatistika Sailarekiko lankidetza bultzatu.
- Funtzionari, irakasle, ikasle eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta abangoardiako ikerlari diren pertsonaiak Euskadira ekarri, guzti horrek zuzeneko harremanei eta esperientzien ezagupenei dagokienez suposatzen duen ondorio positiboarekin.

Iharduketa osagarri bezala eta interesaturik leudekeen ahalik eta pertsona eta Erakunde gehienetara iristearren, Ikastaro hauetako txostenak argitaratzea erabaki da, beti ere txostenemaiaren jatorrizko hizkuntza errespetatuz, horrela gure Herrian gai honi buruzko ezagutza zabaltzen laguntzeko asmoarekin.

Vitoria-Gasteiz, 1992ko Urria  
JOSE MARIA AGIRRE ESKISABEL  
Zuzendari Orokorra

## PRESENTATION

In promoting the International Statistical Seminars, the Basque Statistical Institute wishes to achieve several aims:

- Encourage collaboration with the University, especially with the Statistics Departments.
- Facilitate the professional recycling of civil servants, university teachers, students and whoever else may be interested in the statistical field.
- Bring to the Basque Country illustrious professors and investigators in the vanguard of statistical subjects, on a worldwide level, with the subsequent positive effect of encouraging direct relationships and sharing knowledge of experiences.

As a complementary activity and in order to reach as many interested persons and institutions as possible, it has been decided to publish the papers of these Courses, always respecting the original language of the writer, to contribute in this way towards the growth of knowledge concerning this subject in our country.

Vitoria-Gasteiz, October 1992  
JOSE MARIA AGIRRE ESKISABEL  
General Director

## **PRESENTACION**

Al promover los Seminarios Internacionales de Estadística, el Instituto Vasco de Estadística pretende cubrir varios objetivos:

- Fomentar la colaboración con la Universidad y en especial con los Departamentos de Estadística.
- Facilitar el reciclaje profesional de funcionarios, profesores, alumnos y cuantos puedan estar interesados en el campo estadístico.
- Traer a Euskadi a ilustres profesores e investigadores de vanguardia en materia estadística, a nivel mundial, con el consiguiente efecto positivo en cuanto a la relación directa y conocimiento de experiencias.

Como actuación complementaria y para llegar al mayor número posible de personas e Instituciones interesadas, se ha decidido publicar las ponencias de estos Cursos, respetando en todo caso la lengua original del ponente, para contribuir así a acrecentar el conocimiento sobre esta materia en nuestro País.

Vitoria-Gasteiz, Octubre de 1992  
JOSE MARIA AGIRRE ESKISABEL  
Director General

## **OHAR BIOGRAFIKOAK**

Malka Kantorowitz Zientzia Matematiko eta Fisikoetan lizentziatu zen 1963. urtean. 1964.ean Estatistikaren inguruko lizentziatura ikasketak bukatu zituen eta 1966. urtean Doktoregoa lortu zuen Estatistikan. Hori guztia Jerusalemgoko Unibertsitate Hebraitarrean.

1961ean Israeleko Central Bureau of Statistics-ean sartu zen eta 1970az geroztik Metodo Estatistikoen Arloko Zuzendaria izan da. 1988tik aurrera Ikerketa eta Garapen Saileko Zuzendariorde-buru gisa aritu da lanean. Bureau horretako zentsu eta inkesta asko eta askotako (familiak, nekazaritza, komertzioa etab.) laginketa eta estimazio metodoen, ikerketa prozeduren eta ebaluazio azterketen eremuan egin du bere lana.

Malka Kantorowitz-ek inkesta metodologikoari buruzko eta ebaluazio azterketen inguruko zenbait dokumentu argitaratuak ditu.

Laginketa metodoei buruzko, inkestaren diseinuaren metodologiari buruzko eta ebaluazio prozedurei buruzko ikastaroak ere eman izan ditu Israeleko zenbait Unibertsitatetan, batipat Jerusalemgoko Unibertsitate Hebraitarrean.

Malka Kantorowitz Israeleko Estatistika Elkartea sortu zutenetako bat dugu eta Elkarteko lehen Idazkari Orokorr gisara aritu zen lanean. Inkesta-Estatistikarien Nazioarteko Kontseiluko kide izan zen (1983-1987) eta bere herrialdeko ordezkarria da Elkarte horrentzat.

## **BIOGRAPHICAL SKETCH**

Malka Kantorowitz completed her B.Sc. in mathematics and physics in 1963; in 1964 she completed her studies for B.Sc. in statistics; and in 1966 obtained her M.Sc. in statistics, all at the Hebrew University of Jerusalem.

In 1961 she joined the Israel Central Bureau of Statistics and since 1970 she has been the Director of the Statistical Methods Division. From 1988 she has also served as a Senior Deputy Director, Research and Development Department. Her work has been primarily in fields of sampling and estimation methods, investigation procedures and evaluation studies for a wide spectrum of the Bureau's surveys and censuses (household, agriculture, business, etc.).

Malka Kantorowitz has published several papers on various issues of survey methodology and on evaluation studies.

She has taught courses in sampling methods, survey design methodology and evaluation procedures at several Israeli universities, primarily at the Hebrew University of Jerusalem.

Malka Kantorowitz is a founding member of the Israeli Statistical Association and served as the Association's first general secretary. She was a member of the Council of the International Association of Survey Statisticians (1983-1987) and is the country representative of this Association.

## **NOTAS BIOGRAFICAS**

Malka Kantorowitz completó su licenciatura en Ciencias Matemáticas y Físicas en 1963; en 1964 completó sus estudios de licenciatura en Estadísticas; y en 1966 obtuvo su doctorado en Estadísticas. Todo en la Universidad Hebrea de Jerusalem.

En 1961 se unió al Central Bureau of Statistics de Israel y desde 1970 ha sido Directora de la División de Métodos Estadísticos. A partir de 1988, también ha trabajado como Subdirectora Jefa del Departamento de Investigación y Desarrollo. Su trabajo se ha basado principalmente en los campos de métodos de muestreo y estimación, procedimientos de investigación y estudios de evaluación para una amplia gama de censos y encuestas del Bureau (familias, agricultura, comercio, etc.).

Malka Kantorowitz ha publicado varios documentos sobre diferentes temas de metodología de encuestas y sobre estudios de evaluación.

Ha impartido cursos de métodos de muestreo, metodología de diseño de encuestas y procedimientos de evaluación en varias universidades israelíes, principalmente en la Universidad Hebrea de Jerusalem.

Malka Kantorowitz es miembro fundadora de la Asociación Estadística Israeli y trabajó como primera Secretaria General de la Asociación. Fue miembro del Consejo de la Asociación Internacional de Estadísticos de Encuestas (1983-1987) y es la representante de su país para esta Asociación.



# AURKIBIDEA

|                        |    |
|------------------------|----|
| <b>LABURPENA .....</b> | 13 |
|------------------------|----|

## CONTENTS

|                           |    |
|---------------------------|----|
| <b>INTRODUCTION .....</b> | 15 |
|---------------------------|----|

|                                    |    |
|------------------------------------|----|
| <b>CHAPTER 1. BACKGROUND .....</b> | 17 |
|------------------------------------|----|

|  |    |
|--|----|
| 1.1 Purposes and Uses of FES's .....   | 17 |
| 1.2 The Survey Target Population ..... | 18 |
| 1.3 The Population Units .....         | 19 |
| 1.4 The Investigating Variables .....  | 20 |

|   |    |
|---|----|
| <b>CHAPTER 2. DATA COLLECTION .....</b> | 23 |
|---|----|

|  |    |
|--|----|
| 2.1 Questionnaires and Diaries .....       | 23 |
| 2.2 Data Collection Methods in FES's ..... | 24 |
| 2.3 Examples from Various Countries .....  | 25 |

|                                   |    |
|-----------------------------------|----|
| <b>CHAPTER 3. THE DIARY .....</b> | 28 |
|-----------------------------------|----|

|   |    |
|---|----|
| 3.1 Length of the Recording Period .....                | 28 |
| 3.2 Variation Over the Recording Period .....           | 29 |
| 3.3 Sampling Variances of Different Diary Lengths ..... | 35 |
| 3.4 Types of Diaries .....                              | 39 |

|   |    |
|---|----|
| <b>CHAPTER 4. THE QUESTIONNAIRE .....</b> | 41 |
|---|----|

|                                |    |
|--------------------------------|----|
| 4.1 The Reference Period ..... | 41 |
| 4.2 Recall Effects .....       | 42 |

|  |            |
|--|------------|
| 4.3 Length of the Reference Period and Sampling Variances .....    | 48         |
| 4.4 Optimal Length of the Reference Period .....                   | 52         |
| 4.5 Variability Estimation .....                                   | 54         |
| <b>CHAPTER 5. SAMPLING .....</b>                                   | <b>55</b>  |
| 5.1 Sampling in Two Dimensions .....                               | 55         |
| 5.2 Sampling Frames and Sampling Designs .....                     | 56         |
| 5.3 Sampling of Households .....                                   | 58         |
| 5.4 Non-Proportional Sampling of Households .....                  | 60         |
| 5.5 Allocating the Sample Over Time .....                          | 64         |
| <b>CHAPTER 6. NON-RESPONSE .....</b>                               | <b>66</b>  |
| 6.1 Effects on the Survey Estimates .....                          | 66         |
| 6.2 Non-Response Rates .....                                       | 67         |
| 6.3 Characteristics of Non-Response .....                          | 69         |
| 6.4 Incentive Effects .....  | 73         |
| <b>CHAPTER 7. WEIGHTING .....</b>                                  | <b>78</b>  |
| 7.1 Weighting in FES's .....                                       | 78         |
| 7.2 Effects of Post-stratification .....                           | 78         |
| 7.3 Classification Trees of Homogeneous Sub-groups .....           | 81         |
| 7.4 Illustration of Classification Trees .....                     | 83         |
| 7.5 Efficiency of Using Classification Trees .....                 | 88         |
| 7.6 External Weights of the Households .....                       | 92         |
| 7.7 Controlling Person Distribution with Household Weighting ..... | 93         |
| <b>CHAPTER 8. CODING .....</b>                                     | <b>96</b>  |
| 8.1 Coding Techniques in General .....                             | 96         |
| 8.2 Coding Goods and Services in FES's .....                       | 98         |
| 8.3 Evaluation of Manual Coding .....                              | 100        |
| 8.4 Automatic Coding .....   | 104        |
| <b>REFERENCES .....</b>  | <b>108</b> |

# **INDICE**

|  |     |
|--|-----|
| <b>INTRODUCCION .....</b>  | 111 |
| <b>CAPITULO 1. ANTECEDENTES .....</b>  | 113 |
| 1.1 Propósitos y Usos de las Encuestas de Gastos Familiares (EGF) .....      | 113 |
| 1.2 La Población Objetivo de la Encuesta .....                               | 114 |
| 1.3 Las Unidades de Población .....  | 116 |
| 1.4 Las Variables de Investigación .....                                     | 117 |
| <b>CAPITULO 2. LA RECOGIDA DE DATOS .....</b>                                | 121 |
| 2.1 Cuestionarios y Diarios .....  | 121 |
| 2.2 Métodos de Recogida de Datos en las Encuestas de Gastos Familiares ..... | 123 |
| 2.3 Ejemplos de Varios Países .....  | 124 |
| <b>CAPITULO 3. EL DIARIO .....</b>   | 128 |
| 3.1 Duración del Período de Recogida de Datos .....                          | 128 |
| 3.2 Variaciones durante el Período de Recogida de Datos .....                | 130 |
| 3.3 Varianzas de Muestreo de Diarios de Diferente Duración .....             | 138 |
| 3.4 Tipos de diario .....  | 143 |
| <b>CAPITULO 4. EL CUESTIONARIO .....</b>                                     | 146 |
| 4.1 El Período de Referencia .....   | 146 |
| 4.2 Efectos del Recuerdo .....   | 147 |
| 4.3 Duración del Período de Referencia y Varianzas del Muestreo .....        | 156 |
| 4.4 Duración Optima del Período de Referencia .....                          | 160 |
| 4.5 Estimación de la Variabilidad .....                                      | 163 |
| <b>CAPITULO 5. MUESTREO .....</b>  | 165 |
| 5.1 Muestreo Bidimensional .....   | 165 |
| 5.2 Marcos y Diseños de Muestreo .....                                       | 166 |

|  |            |
|--|------------|
| 5.3 Muestreo de Familias .....   | 169        |
| 5.4 Muestreo No Proporcional de Familias .....                                     | 171        |
| 5.5 Asignación de la Muestra a un Período de Tiempo .....                          | 177        |
| <b>CAPITULO 6. LA NO RESPUESTA .....</b>   | <b>180</b> |
| 6.1 Efectos sobre las Estimaciones de la Encuesta .....                            | 180        |
| 6.2 Ratios de la No Respuesta .....  | 181        |
| 6.3 Características de la No Respuesta .....                                       | 184        |
| 6.4 Efectos de los Incentivos .....  | 189        |
| <b>CAPITULO 7. PONDERACION .....</b>   | <b>196</b> |
| 7.1 La Ponderación en las Encuestas de Gastos Familiares (EGF) .....               | 196        |
| 7.2 Efectos de la Post-Estratificación .....                                       | 197        |
| 7.3 Arboles de Clasificación de los Subgrupos Homogéneos .....                     | 200        |
| 7.4 Ilustración de los Arboles de Clasificación .....                              | 203        |
| 7.5 Eficacia de los Arboles de Clasificación .....                                 | 208        |
| 7.6 Ponderaciones Externas de las Familias .....                                   | 213        |
| 7.7 Control de la Distribución de Personas con la Ponderación de Familia .....     | 215        |
| <b>CAPITULO 8. CODIFICACION .....</b>  | <b>218</b> |
| 8.1 Técnicas de Codificación en General .....                                      | 218        |
| 8.2 Codificación de Bienes y Servicios en las Encuestas de Gastos Familiares ..... | 221        |
| 8.3 Evaluación de la Codificación Manual .....                                     | 224        |
| 8.4 Codificación Automática .....  | 228        |
| <b>BIBLIOGRAFIA .....</b>  | <b>233</b> |

## **GASTU FAMILIARREI BURUZKO INKESTETAKO ALDERDI METODOLOGIKOAK**

### **LABURPENA**

Gastu Familiarrei buruzko Inkestak (GFI), etxebizitza eta familiaren inguruko inkestaren artetik garrantzitsuenak dira herrialde askotan, eta erabilera asko eta asko dituzte. Inuesta horiei izen desberdina eman dakienee egiten diren herrialdearen arabera, esate baterako Kontsumitzalearen Gastuei buruzko Inkestak, edo Familiaren Aurrekontuari buruzko Inkestak, baina funtsean antzekoak dira. Horren garrantzitsuak izateaz gain, GFIak inesta zalantzazkoenak, zailenak eta horrenbestez egiteko neketsuenak dira, eta gainera horiek egiteko behar den familien laguntza lortzea beste inesta mota batzuk egiteko baino nekosoaagoa gertatzen da. Inestak ezartzen dituen esigentziak direla-eta, erantzun maila altua mantentzea zaila izan da beti. Gainera, familiek trabak jartzen dituzte maiz beren diru-sarrera eta gastuen berri ematerakoan, intimitateari eraso egitea dela erizten bait diote. Hori horrela izanik, GFIetatik ateratako estimazioek akats handiak izan ditzakete eta ondorioz interpretazio okerrak egitea gerta daiteke, egokiak ez diren erabakiak hartzera bultzatzuz.

Horrela bada, bai lehen eta bai orain ere, GFIetako datuen kalitatea ebaluatzenko eta inkestaren fase desberdinakoa akatsen jatorria ezagutzeko izugarritzko ahaleginak egin dira. Ebaluazio azterketa horien helburua da GFIen metodologia hobetzea eta inkestaren osagaien artean baliabide hobeak banatzea.

GFIekin gai metodologiko ugari elkartzen da eta argitalpen hau horretaz ari da, nahiz eta horien arteetik batzuk besterik hautatu ez diren. Liburuki honetan landutako gaiak GFIen testuinguruuan dauden arren, horietako zenbait beste inesta familiar motekin erlazioa izan dezakete.

Lehenengo kapituluau GFIen historiari buruzko zenbait puntu azaltzen dira gai metodologikoak hobeto uler daitezzen. GFIen helburu eta erabilerak deskribatzen dira bertan, inkestaren gai den biztanleria eta ikerketaren aldagaien zerrenda.

Bigarren kapituluak GFIetarako datu-bilketen erabiltzen diren metodologien deskribapen orokorra ematen du eta hirugarren eta laugarren kapitulueta rako sarrera da. Bi kapitulu horietan zehatz-mehatz aztertzen dira diario eta galde sorten erabilerekin lotuta dauden gai metodologikoak.

Bostgarren kapituluua GFIen laginketari buruzko kontuez ari da, laginketaren bi fasesak diren denbora eta familia dimentsioekiko. Seigarren kapituluak GFIen ez-erantzunaren ratio eta ezaugariak aipatzen ditu. Ez-erantzuna eta azpitaldeen arteko differentzialtasuna gutxitzeako moduak ere aztertzen ditu kapitulu horrek.

GFIetako ez-erantzun mailak alborapen larria sor dezakeelako eta GFIek estimazio zehatzak ematen dituztelako funtsezkoa da ponderazio prozedura egokia erabiltzea, bai alborapenak eta bai laginketabarriantza gutxitzeako. Hori da zapzigarren kapituluaren gaia, bertan ponderazio metodoak aztertzen bait dira enfasia post-estratifikazioaren hobekuntzan jarriaz.

Azkenik, zortzigarren kapituluan ondasun eta zerbitzuak diario erregistroan kodetzeko teknikak aipatzen dira. Eskuz egindako kodeketan kode-akatsak ekidin beharrak duen garrantzia eztabaidatzen da eta kodeketa automatikoaren erabilera agertzen.

Erlazionatutako ikerketen emaitza enpirikoak ematen dira landutako gai ia guztientzat, gehienak ere Israelgo esperientzian oinarrituta daudenak nahiz eta beste ikerketetako datuak ere ematen diren.

GFlek gai asko biltzen dituzte, baina argitalpen honen eremutik at gelditzen dira. Gainera, esan bezala, ezin dira hemen gai metodologiko guztiak aztertu horien artean item-en ez-erantzunen inputazioak, inkesta egin deneko aldian gertatutako egokitzapenak eta erreferentziako denboraldian familiaren osakerran gertatutako aldaketak.

# METHODOLOGICAL ISSUES IN FAMILY EXPENDITURE SURVEYS

## INTRODUCTION

Family Expenditure Surveys (FES's) are among the most important household surveys in many countries and they serve a wide range of uses. These surveys may go under different names in various countries, such as Consumer Expenditure Surveys, or Household Budget Surveys, but they are basically of a similar type. As well as being of such importance, FES's are also among the most problematic, complicated and, thus, costly surveys to conduct and also one of the most difficult to get the households' cooperation. Because of the heavy demands which the survey imposes, maintaining a sufficiently high level of response rates has always proved difficult. Furthermore, households are often reluctant to reveal details of their income and of their expenditure regarding this as an invasion of their privacy. Consequently, the estimates derived from FES's may be subject to serious errors and thus may be interpreted incorrectly, so leading to wrong decisions.

Therefore, much effort has in the past been invested and still is invested in evaluating the quality of FES data and in tracing sources of errors in the different phases of the survey. Such evaluation studies are aimed at improving the methodology of FES's, as well as attaining a better allocation of resources among the survey's components.

A wide range of methodological issues are associated with FES's and they are the concern of this publication, although only a selection is covered. Although the topics dealt with in this publication are within the context of FES's, some of them may be relevant to other household surveys.

Some background of FES's is outlined in Chapter 1 for a better understanding of the methodological issues. The purposes and uses of FES's, the survey target population and the scope of the variables under investigation, are described in this chapter.

Chapter 2 gives a general description of data collection methods used in FES's and serves as an introduction to Chapters 3 and 4, where methodological issues associated with the use of diaries and questionnaires are discussed in detail.

Chapter 5 deals with sampling aspects of FES's with respect to the two phases of sampling in the household and time dimensions. Chapter 6 focuses on the rates and the characteristics of the non-response in FES's and on their implications for the reliability of the survey estimates. Ways to reduce non-response and its differentiability between sub-groups, are also included in this chapter.

Since the high rate of non-response in FES's may be a source of serious bias and because FES's provide detailed estimates, a good weighting procedure is essential to reduce both biases and sampling variances. This is the concern of Chapter 7 where weighting methods are discussed, with an emphasis on improving post-stratification.

Finally, in Chapter 8, coding techniques of goods and services of the diary entries are presented. The importance of avoiding coding errors is discussed for manual coding and the use of automatic coding is illustrated.

For most of the issues discussed, empirical results of related studies are presented, the majority being based on the Israeli experience, although data from other studies are also presented.

Many subject-matter issues are involved in FES's, but these are beyond the scope of this publication. Also, as was mentioned above, not all the methodological issues could be covered here, among them, methods of imputation for item non-response and adjustments required for price changes over the survey period and for changes in the household composition over the reference period.

## CHAPTER 1. BACKGROUND

### 1.1 Purposes and Uses of FES's

The primary purpose of FES's is to update the weights for Consumer Price Indices (CPI's). However, FES's have also always aimed to reflect family living standards and to provide statistical data for economic analysis. Although these surveys were originally concerned mainly with the expenditure of households, nowadays, additional information is collected about many characteristics of each household and its members, as well as about sources of income and sometimes about savings. In consequence the survey has become a multi-purpose survey with a unique fund of economic and social data which is widely used by government departments for policy making, by academic researchers to develop and test theories and models, by the market sector for market research, and by others.

More specifically, estimates derived from FES's can be used for many purposes:

- (1) To monitor the general CPI with updated weights of the "basket" of goods and services that goes into the calculation of the CPI.
- (2) To provide updated weights of "baskets" to calculate CPI's for specific sub-groups of the population, such as households of elderly people, households living mainly on welfare, etc.
- (3) To supply estimates required for weighting or allocation of the sample of price collection points. As an example, the survey can supply estimates of the distribution of expenditure on food commodities by type of store (groceries, supermarkets, kiosks, open-markets, etc.) and by the day of the week. Then, the sample for the CPI can be planned to represent this distribution correctly to reflect more accurately the weighted average prices in different purchase points and on different days.
- (4) To study consumption patterns of households by their socio-economic characteristics, e.g., family size and composition, education, income, etc. These can be used for policy decisions in connection with social and economic planning, such as on pensions and welfare programs, wage settlements and various support payments. Furthermore, the survey data can be used for estimating the redistributive effects of direct and indirect taxation and of a wide range of social benefits on the situation of various types of families.
- (5) To analyze nutrition levels of different sub-groups of households by using the estimated quantities of food commodities consumed. Empirical functions are used together with the estimated quantities consumed to obtain estimates of nutrition components. FES's are often the only source for such nationwide information, despite their limitations (because of food wastage, insufficient information on eating out, etc.).
- (6) To analyze demand elasticities for selected goods, i.e., how the demand for certain goods is affected by level of income within sub-groups of the population. For example, demand for essential food, like bread, is expected to be less elastic than expensive meat. Such analysis can often throw light on consumption behavior and can be used for a better understanding of the socio-economic structure of the population. Sometimes, such information can be used for tax policy or for allocation of subsidisation funds between different goods and services.
- (7) To study the housing conditions and the possession of durable goods in relation to income and expenditure, for the whole population, as well as for selected sub-groups.
- (8) To analyse the income distribution of households by detailed sources of income, and their differentiation between certain groups of households.

- (9) To study savings patterns of different groups of the population including negative savings, i.e., indebtedness.
- (10) To study the differences in the cost of living and standard of living between different areas of the country, as well as between different countries.
- (11) To analyse the variation in the standard of living over a period of years and the disparities among households in different socio-economic groups.
- (12) To make business decisions and forecasts of market demands by producers and retailers, as well as for making advertising more efficient.
- (13) To supply data for use in compiling official estimates of consumption expenditure in the system of the National Accounts and Balances.

This list of possible uses of FES's is not exhaustive and the survey estimates can be utilized for more specific purposes. For example, in the Eastern-European countries, the FES now serves as one source of assessing the extent of the recent economic phenomenon of the privatisation of the trade sector (Kordos and Kubiczek, 1991)

## 1.2 The Survey Target Population

FES's have a long history, even in the developing world, and some countries, for example, the U.S.A., France and Poland, report on some kind of FES being carried out for nearly a hundred years. In most countries the first FES's were aimed to cover only some specific sub-groups of the population, such as industrial workers and their families, or workers in large cities. The coverage of the target population has been gradually extended over the years in most countries, and there has been a general trend to cover all sections of the population, including families of employees, self-employed, farmers, unemployed, etc. in both urban and rural areas.

For example, in the U.S.A. the first nationwide expenditure survey was conducted in 1888-91 to study workers' spending patterns. A second survey was conducted in 1901 covering workers' families in order to supply weights for an index of prices of food purchased by workers which was used as a deflator for their incomes. The 1917-19 and the 1934-36 surveys provided weights for the CPI including all kinds of expenditures, but covered only urban wage earners and clerical workers. In 1935-36 the survey included both urban and rural sections of the population, in 1950 the survey covered only urban consumers, but since then, in 1960-61, in 1972-73 and in the ongoing continuous survey since 1979, the surveys have covered the entire civilian non-institutional population (U.S Bureau of Labor Statistics, 1988).

In Hungary (Elteto, 1991), the first FES was taken in the late twenties, covering workers and employee households living in Budapest. Regular FES's began in 1948 covering all workers, employees and peasants and since 1976 the survey population was extended to cover also pensioner households.

In Israel, the first FES in 1950/51 (one of the first surveys conducted immediately after the State was established), and the three subsequent FES's in 1956/57, 1959/60 and in 1963/64, covered only families of salary and wage earners living in localities of 10,000 or more inhabitants. In 1968/69, 1973/74 and in 1979/80 all urban families were covered. The 1985/86 survey included all families in all localities of 2,000 or more inhabitants, and also the 1992/93 ongoing survey. However, the rural population in smaller localities (comprising less than 10% of the entire population, of which about a third live in "Kibbutzim" where families cannot be regarded as separate consuming units) is not yet covered in FES's in Israel.

The goal of covering all the population sectors in the FES's has not yet been achieved everywhere. For example,, the rural population is sometimes excluded from the survey due to operational obstacles (like costs), or due to problems of definition (as with self-produced goods used by the family), but not because of a policy decision.

Since the FES estimates relate to households, none of the FES's cover the institutional population, e.g., patients and inmates of homes for the aged, infirm or needy. FES's usually also do not cover transits in hotels, missions and convents, communes, troop quarters and housing units on military reservations, etc. However, families that are independent economic units, even if sharing housing facilities, have to be covered by FES's. Thus, those living in sponsored housing facilities for the elderly, for students, for worker staff in hospitals, etc., have to be included in the survey. This is also the case with permanent type living quarters in hotels, mobile home parks and other structures used as private places of residence. The inclusion of such groups of the population in the survey may involve coverage problems, as well as some subject-matter issues. As an example, there may be a problem to distinguish between expenditures on rent, electricity, heating, water supply, etc., since they are often included in one overall payment. Nevertheless, the inclusion of such groups is important because of their significant contribution to certain sub-groups of the population, like the elderly.

Occasionally, special FES's are conducted to cover specific sub-groups of the population. For instance, in Israel, supplementary FES's of families of handicapped persons and of families who receive regular welfare support were carried out in 1963/64 and in 1968/69. Special surveys of this kind are performed in order to adapt the inquiry procedures and the set of the investigating variables to the specific sub-groups and use a much larger sample for these groups than would be obtained for them in the sample of the regular FES.

### 1.3 The Population Units

Different terms are used for the population units to which most of the survey estimates refer: "family", "household", "consumer unit", "spending unit", etc., and they may be used interchangeably in the following.

The definition of the population unit is basically the same in most countries, although there are some variations between different surveys. Four criteria are used in defining a group of people as one unit:

- (a) Common provision for most food and most other essentials of living.
- (b) Pooling incomes and sharing expenditures (to some extent).
- (c) Living together in the same dwelling.
- (d) Relationship by blood, marriage, adoption or other legal arrangement.

Criteria (a), (b) and (c) are always used with some variation, whereas criterion (d) is rarely used, because in recent years more and more people live as a normal household, although they do not satisfy this criterion. However, criterion (d) is sometimes used in defining sub-units within the household for analytical purposes. For example, in Canada, besides the "household" which is defined by using the first three criteria, an "Economic Family" is also defined by using all four criteria, and may include such persons as foster children, married children and their spouses, brothers, sisters, parents, grandchildren, in-laws, cousins, etc. Another unit they use is a "Census Family", which refers to a core family, i.e., husband and wife or couple living in common-law or a lone parent, with or without children who have never been married (regardless of age) where adopted children and step-children have the same status as own children (Statistics Canada, 1990 FES Interviewer's Manual).

Criteria (a) and (b) are usually used without any attempt to quantify the extent of sharing incomes and expenditures, and of common provision of food. As for criterion (c), there is some variation in defining the persons who belong to the household, depending on the population coverage, the duration of their stay in the household, etc. For example, in Israel, citizens are excluded if they have been abroad for at least a 3-month period and foreign residents (tourists or temporary residents) are included only if they have been in the country for at least the whole 3-month period preceding the interview (with the exception of foreign diplomats who are always excluded from the survey population). Household members (or the household as a whole) with two or more residential addresses, belong to the address where they spend most nights. However, this is not the rule for married persons who always belong to the household of their spouse, even if for some reason (work, study, but not because of separation) they have a different address. This criterion is not applied to soldiers and children in boarding schools who are always included at their home address. Students are regarded as a household on their own, if they live in university sponsored housing or in some other dwelling. Elderly people living in what is called "sheltered housing" are included in the survey population, if they are considered as financially independent households.

The household composition is usually determined according to the situation on the initial visit to the household. However, changes in the household composition can occur during the survey reference period. In some countries, there is a special treatment for such cases either by collecting information of the changes only, or by collecting further information, for adjustments.

In all surveys, rules are established in order to attain units as close as possible to consuming units, as well as, to avoid biases in the survey estimates due to possible duplicated reporting or, under-reporting of expenses. However, it should be noted that for the majority of the population, there are no problems of definition, whatsoever.

## 1.4 The Investigating Variables

For the primary use of FES's, i.e., providing estimates of weights for the general CPI, only detailed information on expenditures is required. However, usually, much more information is collected in FES's for the many other uses of the survey, and, sometimes, just for improving the accuracy of the data provided by the respondents on expenditures. Thus, data collection in most FES's cover a comprehensive set of investigating variables, with some variation between different countries.

The list of potential investigating variables that is presented in the following is neither exhaustive nor are all the listed variables always included in FES's.

A. Household composition according to characteristics of its members:

- (1) Sex, age, marital status, religion or ethnic group, and country of birth and year of immigration (in countries with many immigrants).
- (2) Relationship to head of household (or other reference person).
- (3) Duration of belonging to the household in the reference period.

B. Socio-economic characteristics of adult members of the household:

- (1) Education characteristics - number of years of study; type of last school attended; highest certificate achieved; present studies; etc.
- (2) Labour force and economic activities - number of weeks and weekly hours of work for employed persons; number of weeks of unemployment but available for work; reasons for not being available for employment (student, housekeeper, etc.).

- (3) Employment characteristics of principle employment (or last employment if the person is not employed at present) - employment status (employee, employer, self-employed, etc.); industry (economic branch); occupation; place of work, etc.

C. Housing conditions and place of residence:

- (1) Characteristics of the dwelling unit - number of rooms (or area measure); housing facilities (number of toilets, kind of heating, etc.). These items are used as an indicator of the standard of living and sometimes also for estimation of the value of the housing owned or rented by the household.
- (2) Possession of consumer durables - vehicles; electrical equipment (refrigerator, cooking oven, etc.); television sets, and others. This information is used not only as an indicator of the standard of living, but also for improving the accuracy of data on expenditure on durable goods, and partly for imputing values of owned durables.
- (3) Characteristics of place of residence - region; type and size of locality; etc.

D. Consumption expenditures in the reference period:

For each expenditure on any good or service (including gifts received or given by the household, items consumed from home production, business stocks and from other receipts in kind), the following details are required:

- (1) Detailed item description - for classifying into one of the predetermined classes of items.
- (2) Value of each expenditure - either the money spent by the household in cash or on credit (and the flow of payments), or the value of the item received or produced by the family with no direct payment.
- (3) The quantity purchased - total quantity, number of units and unit's weight. This variable is optional for specific uses, but also helps to control the quality of the data on expenditures.
- (4) Date of expenditure - usually defined as the day of receiving the good or service and not as the day of actual payment.
- (5) Source of acquisition of items not purchased directly by the household (gifts, home production, taken from business stock, etc.).
- (6) Type of store where the item was purchased - optional for specific additional uses of the survey, but can be used also to improve the quality of the data on expenditures.
- (7) Point of purchase - optional for additional uses of the survey, such as for allocation of the sample of outlets for the CPI.

E. Income from all sources:

- (1) Income from paid work - salaries and wages; special sums received from employer; from self-employment and from other work.
- (2) Income from pensions.
- (3) Income from capital - from housing; vehicles; rents; interests and dividends; etc.

(4) Income from allowances and supports - from National Insurance; welfare; from other households; etc.

(5) Income from all other sources.

F. Non-consumption Expenditures:

(1) Income taxes and other direct taxes - these are sometimes not investigated directly and are imputed according to tax rules.

(2) Pensions and social security contributions - these are sometimes regarded as savings.

(3) Remittances, gifts and transfers to other households (non-cash gifts may be regarded as consumption).

G. Savings of all kinds:

(1) Cash savings.

(2) Saving in real estate.

(3) Business savings.

(4) Non-regular receipts - lump-sum from insurance; sale of household's possessions and property; etc.

(5) Other savings - loans obtained; loan payment received; etc.

There are hardly any other household surveys covering such a wide range of investigating variables. Many subject-matter problems are related to the definitions of the various investigating variables, thus the wording of the questions is of primary importance and the instructions to the interviewers are therefore quite complicated. Although all these issues are beyond the scope of this publication, the methodological issues of FES's, which are the concern of this publication, are also affected by the complexity of the survey due to the wide range of the investigating variables.

## CHAPTER 2. DATA COLLECTION

### 2.1 Questionnaires and Diaries

In most household surveys, the data required are collected from the sampled units by means of either a questionnaire or a diary or, sometimes, by both.

With questionnaires, the required data are collected by means of a set of questions that may refer to information of various types:

- (a) Characteristics that are invariant in time, like sex, year of birth, etc.
- (b) Data related to the time of inquiry, such as last school attended, number of rooms in the dwelling, possession of durable goods, etc.
- (c) Retrospective information on events that happened in a predetermined reference period in the past, e.g., number of visits to doctors in the last month, expenditures on various items in the last three months, time spent abroad in the last year, etc.
- (d) Retrospective information that relates to the last event of inquiry, when there is no predetermined reference period, e.g., information on the last visit to a doctor.

Diaries are used for recording detailed data of specific events that occur in a certain period and may be of two principal types:

- (1) A "recall diary" for recording details of events that occurred in the recent past. For example, in Israel, in the 1985 Travelling Habits Survey a two-day recall diary was used for recording details of every journey of each person. Although the information collected by both recall diaries and retrospective questionnaires is based on recall, they are considered to be different because of the way the required information is collected and because recall diaries must refer to a very short recall period.
- (2) A "current diary" in which the recording is done continuously during the recording period while the event is occurring (or, at least, within a very short time after its occurrence), such as details of each activity on one day in the Time Use Survey, or details of each expenditure in a two-week period in FES's. In practice, even when a current diary is used, some of the recording is done by recall.

The use of questionnaires for collecting retrospective information, or the use of recall diaries, is justified if it can be assumed that the respondents are capable of supplying reliable information on past events. However, the respondent may not remember accurately all the details related to a certain event, or may even forget the occurrence of an event in the reference period. The respondent also may report erroneously on events that occur before the reference period, or place the occurrence of the event wrongly in time, a phenomenon known as "telescoping".

These recall effects depend on the length of the recall period and on the type of event. Thus, details on important events, are usually well-remembered, e.g., moving to a new flat, a trip around the world, or being involved in a serious car-accident. Details of less important events are more often forgotten, e.g., the dates of having eaten away from home in the last year, or the expenditure on frequently purchased items in the last month. Recall effects may also depend on the respondent's characteristics, e.g., old people may have more difficulties in recalling certain events.

Since recall effects can result in serious biases in the estimates, in several surveys certain information has to be collected by means of a current diary. In theory, recall effects should be avoided with the use of a current diary, but, in practice, because some of the recording is done by recall, even with this type of diary some recall effects may be introduced, although less than with recall diaries. On the other hand, data collected by recall are not affected by possible changes in the behavior of the investigated unit.

With current diaries non-response rates are usually higher than with recall diaries. Moreover, the differentiability between non-response rates in different sub-groups of the population is usually greater when current diaries are used. However, the use of recall diaries may cause under-representation of the days preceding the weekend, e.g., Saturdays if interviewers do not work on Sundays, while with current diaries, the distribution of the diaries over the weekdays can be better controlled.

As for cost, a one-day recall diary requires only one visit of the interviewer, whereas for a one-day current diary, two visits are needed, one for placing the diary and one for collecting the completed diary. Mailing back the completed current diaries would save one visit, but this usually results in much lower response rates. It should be noted that when a current diary is used, because the respondent has to be instructed, the interviewer anyway has to make some recording by recall together with the respondent. When several days of recording in the diary are required, say a week or two, the recall diary is more costly than the current diary, because the interviewer must visit the investigated unit more often in order to avoid serious recall effects.

There is no clear preference for either one of these two types of diary. The decision as to which of them to use has to be taken separately in each survey, according to the length of the recording period, to the type of the required information and to the characteristics of the population under investigation. In some surveys, diaries of both types are used. As an example, in the Israeli 1992 Time-Use Survey, both types of diary were used: a one-day recall diary was filled in by the interviewer for the previous day (or, for two days for the weekend), and the respondents were also requested to complete a one-day current diary (or, two-day diary for the weekend) for a predetermined day (usually the day after the interview), and to mail it back.

## 2.2 Data Collection Methods in FES's

For almost all FES's, data collection is based on some combination of both questionnaires and diaries. This is done in order to obtain the very detailed information required for the survey with the highest possible degree of accuracy. Reliable information on all frequently purchased items and on all small expenditures, even if not frequently purchased cannot be supplied by the respondent by recall, unless the reference period is very short. Since a too short diary period will lead to high sampling variances, a current diary is used for recording of all the detailed expenditures (or only for certain groups of items, such as food, or clothing). However, it is not possible to impose too much of a burden on the respondents by asking them to keep a diary for a long period, because less families will agree to participate in the survey, and more will drop-out before the end of the recording period. Furthermore, the quality of the recording in the diaries deteriorates through the recording period.

Therefore, with a reasonable length of the recording period, infrequent expenditures will be recorded in the diary very rarely and thus no reliable estimates for them can be obtained because of a too-high sampling variance. Thus, retrospective questions in a questionnaire are used to collect information on infrequent expenditures. This is done on the assumption that recall effects are not too serious, providing the reference period is not too long. It should be noted that some of the information collected by the questionnaires may be based on documents, such as bills, salary forms, etc. Although no recall effects are involved when the respondent refers to such documents, he may not still have them, or may not be willing to show them while being interviewed.

Questionnaires are used as well for collecting data on the required characteristics of the household and its members and some information, such as consuming habits. In some surveys the questionnaires or the diaries or both are used for each adult member of households, and in other surveys they are used for the household as a whole.

In most FES's, the questionnaires are completed by an interviewer, whereas the respondents are requested to complete the diaries by themselves. Diaries are rarely planned to be filled in by interviewers (except in some developing countries where a high rate of illiteracy still exists). Interviewers complete the diary only when there is no one among the household members capable of doing so, e.g., households of old people.

Sometimes, a comprehensive interview is held preceding the placement of the diary, and sometimes only a short interview is conducted to start with and the main interview is held after collecting the completed diary.

In some countries, one survey is conducted where each sampled unit is requested to keep a diary as well as to be interviewed about all the required information. In other countries, separate surveys are carried out on separate samples, where each sample concentrates on one form of data collection, interview or diary. Separate diary surveys are sometimes aimed at all expenditures (or at all monetary transactions), and sometimes at specific major groups of expenditures, usually on food, but also on other groups, such as clothing and footwear.

## 2.3 Examples from Various Countries

- (1) In Israel, the recent 1986/87 and the ongoing 1992/93 FES's have been based on a two-week diary for recording all expenditures, whereas in previous surveys, monthly diaries were used. An introductory interview is made for collecting data on demographic and socio-economic characteristics of all household members, as well as on consuming habits of the household. At the end of the recording period, a comprehensive questionnaire is filled in by the interviewer relating to durable goods and other large expenditures, as well as on income and savings. The reference period is the last three months for most of the items, and, for very infrequent expenditures, 12 months. In previous surveys, the reference period was the last 12 months, except for some regular expenditures such as telephone, water supply, etc., where only the last payment was requested.
- (2) In the United Kingdom (Matheson, 1991), a personal two-week diary is used for recording all expenditures, for each household member aged 16 years and over. Two questionnaires are used, the first one includes questions about household composition, housing expenditures, etc., as well as refunds and rebates relating to any of these items. This questionnaire relates to the household as a whole, whereas the second questionnaire, the income schedule, asks each individual about income from all sources. The questionnaires are completed by the interviewer, before the placement of the diaries.
- (3) In Australia (Harrison, 1991), in the 1988/89 FES, a personal diary of all expenditures over a period of two weeks was used for each member of the household aged 15 years and over. In addition, two questionnaires were used, one for expenditures and one for income. Both were completed in the initial interview before placing the personal diaries (except for information relating to adults absent on the initial visit). The first part of the expenditure questionnaire contained data about the household characteristics and only one adult in each household was asked to provide this information. The second part collected information on certain types of expenditures by all respondents in the households. The questions were based on recall with a reference period of between 3 months to 2 years, depending on the type of item. A separate income questionnaire was completed for each adult to obtain all sources of income, tax paid etc, as well as some socio-economic information, such as employment status, occupation, etc.

- (4) In Denmark (Moller, 1991), a two-week diary is used for recording every purchase of food, and a 4-week diary is used for all other purchases. A detailed description of each purchase, the expenditure and the quantity, have to be recorded in the diary by the respondent. Some supplementary information for special goods is also included, for example, sex and age (for a better classification of clothes) and whether the goods were bought abroad. After a year of conducting the diary survey, at the beginning of the following year, an extensive interview is carried out covering regular payments (rent, insurance, etc.), expenditures of more expensive goods, income by source, etc., as well as socio-economic characteristics of the household and its members. An additional source of data are administrative registers, such as the tax-register of personal incomes and wealth. These sources of data are easy to obtain in Denmark via a person's registration number and this thus lessens the heavy burden both on the respondents and the interviewers.
- (5) In France (Glaude, 1982), in the 1978/79 FES, a personal 10-day diary was used for each household member aged 15 years and over. Three questionnaires were used, in three separate interviews. One questionnaire on family composition, accommodation, cars, list of durables etc., was completed just before the recording period. The second one was completed during the recording period, including questions about buying habits, consumption from own products and business and some other information. The third questionnaire completed immediately after the recording period related to income and expenditures in the last 12 months, and to clothing and footwear and heating in the previous two months.
- (6) In Hungary (Elteto, 1991), until 1976, households participating in the FES, recorded their incomes and expenditures during the whole year, many of them even for a period of several years. Since 1976, the amount of diary recording has been significantly reduced and households are asked to record their incomes and daily expenditures and quantities in a diary during a 2-month period. Additional diaries are given to adult household members who want to keep their diaries separately. The interviewer visits the household several times during the two months to give assistance if any problem arises and to check the diary entries. At the end of the year, the households are visited again by the interviewers and questioned about various sources of annual income (e.g., wages from second jobs, personal income tax, etc.), yearly expenditure on some items of considerable value (e.g., household durables, vehicles, holiday abroad, etc.). At the same time, a stock of durable goods is also taken.
- (7) In Canada (1990 FES' Manual), the 1990 FES' program includes two separate surveys. A diary survey is conducted throughout a calendar year and provides details of expenditure and quantity only of food commodities. Households are asked to complete a detailed daily diary for two weeks. Before placing the diary, a short interview is conducted including questions on certain socio-economic characteristics of each household member, on characteristics of the dwelling occupied by the household, on personal income in the past 12 months, on food expenditure away from home and on household spending habits. In February and March of the following year, an annual recall survey is administered. This survey provides details on all other purchases of consumer goods and services for the previous calendar year, and includes also the yearly expenditure on food (as a single group) as well as on personal income, taxes, insurance premiums, gifts, etc., and on changes in the household financial position in the previous year. In addition, information is collected by an interviewer about household composition.
- (8) In the U.S.A. (Jacobs, Jacobs and Dippo, 1989), an annual recall procedure was used in the 1960/61 FES, but this data collection method was dropped and a new methodology was instituted for subsequent surveys. The 1972/73 FES and the current survey initiated in the late 1979, consist of two separate surveys, each with a different collection technique and sample: the Diary Survey and the Interview Survey which is a rotating panel survey.

In the Diary Survey each family is requested to keep a diary of all its expenditures for two consecutive 1-week periods. On the interviewer's first visit to the family, data on the housing unit and demographic data for each household member are collected before the diary is placed. On completion of the two weekly diaries, the interviewer collects data on labour force characteristics and on income on a personal basis.

In the Interview Survey each investigated unit is requested to be interviewed in five calendar quarters. In the initial interview, data are collected for only a one-month recall period and they are used only for bounding purposes. This first interview is also used to collect demographic characteristics and to provide an inventory of major durable goods. In the other four interviews all information is collected by the use of uniform questionnaires with a reference period of three months. Some details on income are collected only in the second and fifth interviews.

Although the few examples outlined here do not reflect all the data collection methods in use, it seems that data collection in almost all FES's is based on diary recording with one or more supplementary interviews. The main differences between countries are with respect to:

- The duration of the diary recording period.
- The length of the reference period for the retrospective questionnaires and the number of interviews made to cover the period.
- The timing of the interview(s) and of the recording period.
- The use of one fixed reference period for the retrospective questionnaire for the whole sample versus a moving reference period.
- The data requested to be recorded in the diaries and the content of the questionnaire(s).
- The collection of data from each household member or from the household as a whole, or from both.
- The use of one comprehensive survey versus two (or more) separate surveys for different major groups of expenditure.

## CHAPTER 3. THE DIARY

Current diaries are used in FES's for collecting part of the required information because it is believed that more accurate data for this information can be provided by the respondents by a diary than by a retrospective interview. This does not mean, however, that data collected by diaries are free of various types of errors and their quality depends on the type and the content of the diary and, of course, on the length of the diary period. Some methodological issues of using diaries in FES's are discussed in the following sections of this chapter and results of several related studies are presented.

### 3.1 Length of the Recording Period

The length of the recording period can affect the reliability of the survey estimates with respect to the following:

- (a) Recording errors - The longer the recording period is, the more likely are biases due to fatigue in recording. However, telescoping effects in diaries, i.e., inclusion of items bought prior to the diary period in the first day or days of recording, may result in more bias in the survey estimates, the shorter the recording period is.
- (b) Changing spending behavior - Changes in household behavior because of the recording of expenditures are to be expected after being involved in the survey for at least a few days. Such changes may result in biases that increase with the length of the recording period.
- (c) Sampling errors - For a given number of households in the sample, sampling errors decrease with the length of the recording period. However, this is not always true for a given investigation cost. The cost per household increases with the length of the recording period, although not necessarily at the same rate, since some fixed overhead cost is invested in each sampled household, not dependent on the length of the recording period.
- (d) Non-response - The longer the recording period per household is, the more difficult it is to persuade the households to participate in the survey, thus, a higher rate of non-response is expected. Furthermore, during a long recording period, the problem of participants dropping out is more serious. When a very long period is requested, more households may move and thus more units are "lost" during the survey.
- (e) Changes in household composition - The longer the recording period is, the more changes that are expected. Thus, additional information, not always easy to get, is required for adjustment of the household data.
- (f) Estimates of the distribution and of parameters of variability of the population - With a short length of the recording period, such estimates by expenditure are limited, or even impossible to obtain, when high variability in expenditure exists between short periods within the families.

One of the first extensive studies on the effects of using different diary methods for FES's was conducted in Israel as part of the 1956/57 FES. The sample of households was divided at random into three approximately equal sub-samples. All households in the three sub-samples were requested to record non-food expenditures for a whole month, but, households in sub-sample (A) were requested to record food expenditures for the whole month, in sub-sample (B) only for the first half of the month and in sub-sample (C) only for the second half of the month.

The main results, based on a total sample of 6,600 households, are presented in Table 3.1.

Table 3.1  
Average Expenditures of Sub-samples (B) & (C) in Comparison to Sub-sample (A)  
Israeli 1956/57 FES

| Expenditure group | Sub-sample |       |       |         |
|-------------------|------------|-------|-------|---------|
|                   | (A)        | (B)   | (C)   | (B)+(C) |
| TOTAL EXPENDITURE | 1.000      | 1.099 | 1.053 | 1.076   |
| Food              | 1.000      | 1.111 | 0.982 | 1.046   |
| Non-food          | 1.000      | 1.091 | 1.103 | 1.097   |

Source: FES's, Israel C.B.S., Special Series No. 148, Jerusalem (1963).

For Food, the comparison between (B)+(C) and (A) relates to the effect of the length of the recording period, showing an excess of 4.6% of the half-month recording over the full month. The comparison between (B) and (C) for Food, may express real differences between expenditures in the two halves of the month due to the date of salary payment which is usually on the first of the month.

As for Non-food, the households in all the three sub-samples recorded their expenditures for precisely the same calendar period of a month. The different results can only be the consequence of the different requirements of recording food expenditures. The results show that the average non-food expenditure is high (+10.3%) for (C) and somewhat less (+9.1%) for (B) when compared to (A). These results indicate that reporting is less when a heavier burden is imposed on the respondents.

Among other results that were obtained from this study (not presented here), a comparison of the average expenditure on Food in the second half of the month between (A) and (C) showed that in (C) higher averages were obtained: 4% for Total Food, and for some detailed food commodities the differences were much higher, e.g., +14% for Meat and +23% for Oil. This could be because of fatigue effects in recording. A comparison of average of Total Food expenditure from only the first half month of (A) and (B), showed that in (B) a higher average was obtained (+11%). This finding was not easy to explain.

The results of this pioneer study indicate how serious can be the effects of different diary methods on the estimates, and that these effects are not the same for different groups of items. Thus, weights for the CPI could be even more seriously biased.

### 3.2 Variation Over the Recording period

Variation in average expenditure levels as recorded in the diaries on different days of the recording period may be due to real differences in actual expenditures, or the result of recording effects, or the interaction between them.

Real differences may be due to effects of:

- (1) The day of the week.
- (2) Holidays and their eves (similar to those of weekends).
- (3) The period preceding holiday eves, when expenditure on certain commodities is higher than usual.

- (4) The day on which salaries are paid (e.g., the first of the month, where salaries are paid on a monthly basis, or Thursday, where salaries are paid every week or fortnight).
- (5) Days on which purchasing behavior can be affected by various other factors, such as first days of seasonal "sales", or periods of changing price expectations.

Recording effects may be caused by:

- (1) Fatigue effect or reduced interest in recording, which grows during the recording period resulting in a fall in diary entries, the longer the period is.
- (2) "Telescoping" effects, i.e., inclusion of items bought prior to the diary period on the first day or days of recording.
- (3) Failure to record each expenditure on the actual day of purchase, e.g., the respondents forget to start a new page when the diary is structured with separate pages for each day, or forget to change the date when this is required to distinguish between days.
- (4) Recall effects, i.e., when respondents delay making entries in the diary and then attempt to reconstruct purchases. Recall effects may also occur when interviewers on their visits to the households reconstruct purchases either while checking the recordings of the respondents, or when they have to fill in the diaries by themselves.
- (5) Temporary changes in actual spending behavior due to involvement in the survey. Households may over-purchase in the first days of recording to impress the interviewer and thus under-purchase in the later days having already stocked up. Respondents may sometimes make purchases within the recording period that they would have normally delayed in order to show "typical" expenditure and thus to be "representative". Keeping records may also disturb expenditure habits when the diary records in the beginning of the diary period are a mirror of respondent behavior, and if the image is not flattering, they will refrain from making certain purchases. It is almost inevitable that when people know they are being observed, they do not behave completely normally, although they would not admit it.
- (6) The timing of the initial visit of the interviewer to the household for placing the diary and further visits made during the recording period may influence recording levels, i.e., higher levels of reporting are motivated by the interviewer's visits, with decreased effects on days after the visit.

The recording effects may result in biases in the survey estimates, but these may cancel out, e.g., "telescoping" may cause over-estimation while fatigue may result in under-estimation. It is usually believed that the recordings in the diaries tend to under-estimate the true values, although this is not proven.

The phenomenon of variations in expenditure levels within the diary recording period in FES's was also found in many diary surveys. This topic is discussed by Prais (1958), Kemsley and Nicholson (1960), Kemsley (1961), Turner (1961), Sudman and Ferber (1971), McWhinney and Champion (1974), Pearl (1979), Nevraumont (1991) and others. Variation was found in either the average daily number of entries, or the daily expenditure per entry, or in both. The extent of the variations is not the same in different surveys, neither for different groups of expenditure in the same survey.

A study on variation within the diary period was carried out also for the Israeli 1975/76 FES (Kantorowitz, 1981). This study was based on a sample of 2,270 households that were asked to record all their expenditures in a diary for a whole month, starting in most cases on the first day of the month. The initial visit of the interviewer to each household was made as close as possible before the starting day.

A simple comparison of the average expenditure by the days of the diary would not separate real differences from recording effects. The ratios of average daily expenditure to the overall daily average were computed for each weekday, as shown in Table 3.2 and these real effects of the days of the week were found to be substantial. Thus, these effects, as well as the effects of holidays and their eves (treating them as Saturdays and Fridays, respectively), were eliminated. However, the other possible real effects mentioned above could not be eliminated.

Table 3.2  
Ratio of Average Weekday Expenditure to Overall Daily Average  
Israeli 1975/76 FES

| Selected group of items         | Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday* |
|---------------------------------|--------|--------|---------|-----------|----------|--------|-----------|
| TOTAL DIARY EXPENDITURES        | 1.00   | 1.07   | 1.10    | 1.25      | 1.28     | 1.00   | 0.21      |
| Food (excl. Vegetables & Fruit) | 0.80   | 1.00   | 1.10    | 1.32      | 1.43     | 1.19   | 0.16      |
| Bread                           | 0.99   | 0.98   | 1.00    | 0.95      | 1.08     | 1.19   | 0.09      |
| Milk                            | 1.09   | 1.05   | 1.08    | 1.02      | 1.06     | 1.66   | 0.04      |
| Meat & Poultry                  | 0.49   | 1.03   | 1.22    | 1.82      | 1.82     | 0.52   | 0.10      |
| Vegetables & Fruit              | 0.67   | 1.04   | 1.07    | 1.53      | 1.57     | 1.03   | 0.09      |
| Clothing & Footwear             | 1.31   | 1.20   | 1.18    | 1.23      | 1.13     | 0.82   | 0.13      |
| Transport & Communication       | 1.20   | 1.06   | 1.08    | 1.09      | 1.01     | 1.12   | 0.44      |

\*) Saturday is the day off work in Israel.

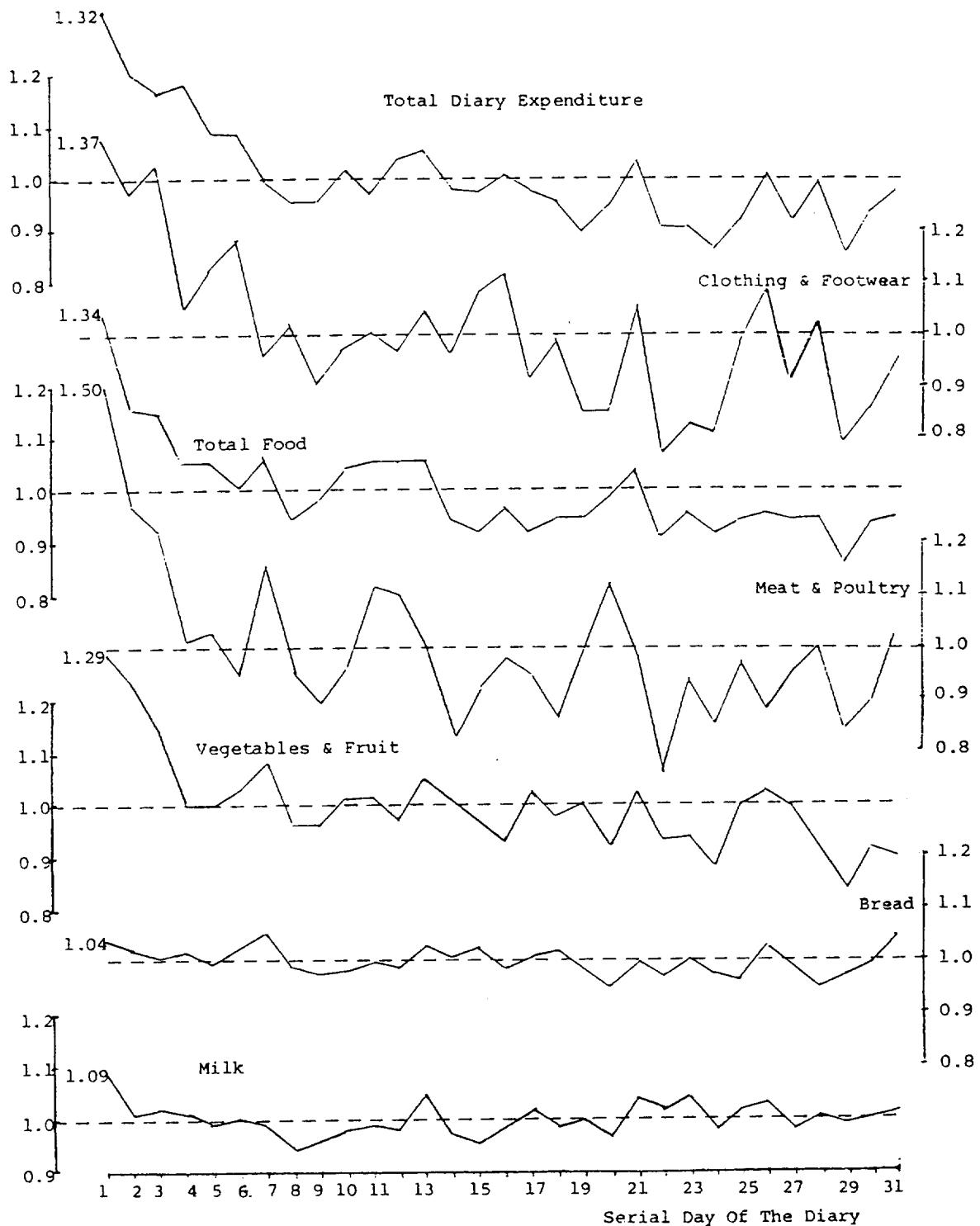
Source: Kantorowitz (1981)

The difference between the serial days of the diary (which almost always coincide with the calendar days of the month), were expressed through the ratios of the average expenditure for each day, relative to the overall daily average (after adjusting for weekdays and holidays). The ratios are presented in Figure 3.1 for selected groups of items: bought frequently (Milk, Bread); less frequently (Vegetables & Fruit, Meat & Poultry); comparatively infrequent purchases (Clothing & Footwear), and also for Total Food and Total Diary Expenditure.

The outstanding finding was the substantially high ratio of expenditure on the first day of recording for most items: 1.50 for Meat & Poultry, 1.37 for Clothing & Footwear; 1.34 for Total Food; 1.32 for Total Diary Expenditure; 1.29 for Vegetables & Fruit, but less for the frequently purchased items, Bread and Milk, where these ratios were 1.04 and 1.09, respectively. Moreover, for most items other than those frequently purchased, the first few days of the recording month also had relatively high levels of expenditure, but not as high as those of the first day. It should be noted that the differences between the average daily expenditures without the elimination of real effects for weekdays and holidays were usually even greater, but followed a similar pattern. For example, the first day ratios for Meat & Poultry were 1.50 and 1.74 with and without this elimination, respectively.

The first day effect, especially for infrequent purchases, could be partially explained in this survey by the real effect of salaries often paid on the first day of the month. However, similar first day effects were also found in other surveys, where real effects of this kind do not exist. Therefore, it is more likely that these results are due to telescoping. Telescoping could have a substantial recording effect in the Israeli survey, as in similar ones, because the initial visit of the interviewer and the diary placement is made a day or two before the first diary day. Thus, the respondents could mistakenly understand that they have to start their recording immediately. As a result of these findings, since the 1986/87 FES in Israel, the first day is recorded for boundary purposes, but not used for the survey estimates.

Figure 3.1  
 Ratio of Average Expenditure(\*) for Calendar Day of the Month Relative to Overall Daily Average  
 Israeli 1975/76 FES



(\*) After adjusting for weekdays and holidays.

The differences between weeks were measured by the ratios of the average expenditure for each of the four weeks in relation to the overall weekly average (the last 2-3 days of the month were ignored). These ratios were computed after elimination of most of the holiday effects by exclusion of data for two months (October and April) in which most of the important holidays occur in Israel. This slightly reduced the differences between weeks. Table 3.3 presents the differences between the four weeks starting either with the first day of the month, and with the second day.

Table 3.3  
Average Expenditure for Each Week Relative to Overall Weekly Average  
Israeli 1975/76 FES

| Week  | Selected groups of expenditure |            |                    |                |       |      |                     |
|---|--------------------------------|------------|--------------------|----------------|-------|------|---------------------|
|   | Total diary                    | Total Food | Vegetables & Fruit | Meat & Poultry | Bread | Milk | Clothing & Footwear |
| Four weeks starting with the 1st day of the month |                                |            |                    |                |       |      |                     |
| Week 1  | 1.14                           | 1.10       | 1.10               | 1.12           | 1.01  | 1.01 | 1.21                |
| Week 2  | 0.98                           | 0.99       | 0.98               | 0.99           | 1.00  | 0.99 | 0.94                |
| Week 3  | 0.95                           | 0.96       | 0.97               | 0.97           | 0.99  | 0.99 | 0.91                |
| Week 4  | 0.93                           | 0.95       | 0.95               | 0.92           | 1.00  | 1.01 | 0.94                |
| Four weeks starting with the 2nd day of the month |                                |            |                    |                |       |      |                     |
| Week 1  | 1.09                           | 1.07       | 1.07               | 1.08           | 1.01  | 0.99 | 1.17                |
| Week 2  | 0.98                           | 0.99       | 1.00               | 0.99           | 0.99  | 0.99 | 0.94                |
| Week 3  | 0.98                           | 0.98       | 0.99               | 0.98           | 1.00  | 1.01 | 0.94                |
| Week 4  | 0.95                           | 0.96       | 0.94               | 0.95           | 1.00  | 1.01 | 0.95                |

Source: Kantorowitz (1981).

No significant weekly differences could be traced for Milk and Bread. For other expenditure groups, the first week averages were significantly higher (1.10 and over), while the 4th week was significantly lower (0.92-0.95). For Clothing & Footwear the first week effect was relatively high (1.21), part of this was probably due to the real effect of the beginning of the month. The results show that even after excluding the first day expenditures, the first week still had higher levels of expenditure, though not as high as when the first day was included.

The differences between weeks might be the result of a possible interaction between households' level of expenditure and the pattern of recording their expenditure. Therefore, ratios for each week were also computed using the weekly percentage out of the total expenditure for each household, instead of their absolute weekly expenditures. Most of these ratios were either the same or differed only slightly (about 1%) from those in Table 3.3. Clothing & Footwear showed somewhat larger differences indicating the existence of a slight interaction.

It was also of interest to see whether most of the households had a similar pattern of diary recording, or whether the results obtained characterized a minority of the respondents with extreme variation in their recordings over the diary period. For this purpose, within each household the weeks were ranked by level of expenditure, i.e., the week with the largest expenditure was assigned rank 1, the second largest received rank 2, and so on, altogether 24 possible combinations (or more if ties occur). Households were grouped according to which week was ranked 1, 2, etc., and according to the ranks assigned to the 1st week, 2nd week, etc. Table 3.4 presents the proportions of these groups in the population and the average expenditures for each week in relation to the weekly group average for Total Diary Expenditure.

Table 3.4  
 Average Weekly Expenditure Relative to Overall Weekly Average  
 of Households with Different Recording Patterns  
 Israeli 1975/76 FES

| Recording patterns  |               |     |     | Percent<br>of<br>households | Average weekly expenditure relative to<br>overall weekly average |          |             |             |             |             |
|---------------------|---------------|-----|-----|-----------------------------|--|----------|-------------|-------------|-------------|-------------|
| Expenditure<br>rank | Week of month |     |     |                             | All weeks  | 1st week | 2nd week    | 3rd week    | 4th week    |             |
|                     | 1st           | 2nd | 3rd | 4th                         |  |          |             |             |             |             |
| ALL HOUSEHOLDS      |               |     |     | 100.0                       | 1.00   | 1.14     | 0.98        | 0.95        | 0.93        |             |
| 1                   | 1             | -   | -   | -                           | 37.3   | 1.00     | <u>1.56</u> | 0.85        | 0.81        | 0.78        |
|                     | -             | 1   | -   | -                           | 22.5   | 1.00     | <u>0.89</u> | <u>1.48</u> | 0.83        | 0.80        |
|                     | -             | -   | 1   | -                           | 20.2   | 1.00     | <u>0.88</u> | 0.82        | <u>1.47</u> | 0.83        |
|                     | -             | -   | -   | 1                           | 20.0   | 1.00     | <u>0.89</u> | 0.83        | 0.82        | <u>1.46</u> |
| 2                   | 2             | -   | -   | -                           | 25.7   | 1.00     | <u>1.07</u> | 1.01        | 0.96        | 0.96        |
|                     | -             | 2   | -   | -                           | 26.6   | 1.00     | 1.22        | <u>1.04</u> | 0.87        | 0.87        |
|                     | -             | -   | 2   | -                           | 24.2   | 1.00     | 1.17        | 0.94        | <u>1.04</u> | 0.85        |
|                     | -             | -   | -   | 2                           | 23.5   | 1.00     | 1.08        | 0.94        | 0.93        | <u>1.05</u> |
| 3                   | 3             | -   | -   | -                           | 19.5   | 1.00     | <u>0.85</u> | 1.09        | 1.04        | 1.02        |
|                     | -             | 3   | -   | -                           | 26.0   | 1.00     | 1.19        | <u>0.82</u> | 1.02        | 0.97        |
|                     | -             | -   | 3   | -                           | 28.5   | 1.00     | 1.21        | <u>1.03</u> | <u>0.81</u> | 0.95        |
|                     | -             | -   | -   | 3                           | 26.0   | 1.00     | 1.22        | 1.01        | 0.96        | <u>0.81</u> |
| 4                   | 4             | -   | -   | -                           | 17.5   | 1.00     | <u>0.65</u> | 1.13        | 1.12        | 1.10        |
|                     | -             | 4   | -   | -                           | 24.9   | 1.00     | 1.22        | <u>0.63</u> | 1.07        | 1.08        |
|                     | -             | -   | 4   | -                           | 27.1   | 1.00     | 1.23        | <u>1.08</u> | <u>0.62</u> | 1.07        |
|                     | -             | -   | -   | 4                           | 30.5   | 1.00     | 1.27        | 1.09        | 1.04        | <u>0.60</u> |

Source: Kantorowitz (1981).

These findings show that 37% of all households recorded their highest expenditure in the first week, while 20% recorded their highest outlay in the fourth week. This distribution of households differed significantly from the expected uniform distribution. Although significantly more households recorded their highest expenditure in the first week, they did not constitute a majority.

For the 37% of the households whose expenditure was highest in the first week, the level in the highest week was 56% higher than their overall weekly average. Expenditures of households which were highest in the second, third and fourth week, exceeded their weekly average by only 48%, 47% and 46% respectively. Moreover, when grouping households by the week assigned rank 2 (and similarly ranks 3 and 4), the first week had a relatively somewhat higher level of expenditure: 1.07 against 1.04-1.05 for rank 2, 0.85 against 0.82-0.81 for rank 3, and 0.65 against 0.63-0.60 for rank 4. Thus, the phenomenon of the highest levels of expenditure in the first week was partly due to more households having their highest expenditure in the first week (37% exceeded the 25% expected), but was also due to higher relative expenditure in the first week for all weeks having the same rank.

In the Israeli survey, the days on which the interviewers visited the households during the diary period were not fixed but depended on each household's ability to keep the diary, on the average every 3 or 4 days. Thus, it is unlikely that the timing of the interviewers' visits affected the overall daily variation, although it could have affected each household separately.

However, in other surveys with fixed days of calling back, the effect of callbacks was substantial. For example, the results of the U.S.A. 1972-73 FES (Pearl, 1979), where most households were visited on the 8th day of recording, a significant increase in the average expenditure for that day was observed, as shown in Table 3.5.

Table 3.5  
Relative Expenditure(\*) for Each 14-Day Diary  
U.S.A. 1972/73 Survey

| Expenditure | All<br>days | Recording day |     |     |     |     |     |     |             |     |     |     |     |     |     |
|-------------|-------------|---------------|-----|-----|-----|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|
|             |             | First week    |     |     |     |     |     |     | Second week |     |     |     |     |     |     |
|             |             | 1st           | 2nd | 3rd | 4th | 5th | 6th | 7th | 1st         | 2nd | 3rd | 4th | 5th | 6th | 7th |
| Food        | 100         | 148           | 106 | 102 | 94  | 101 | 92  | 87  | 116         | 96  | 91  | 98  | 91  | 87  | 90  |
| Non-food    | 100         | 139           | 108 | 106 | 99  | 98  | 93  | 92  | 108         | 98  | 95  | 91  | 90  | 90  | 94  |

(\*) Adjusted for imbalances in diary placement over weekdays.

Source: Pearl (1979).

Even if the significant differences between the different periods of the diary that were found in various studies are all due to recording effects (although some of them could be caused by real effects that were not completely eliminated), it was not possible to say if such differences introduced biases in the survey estimates and in what direction. Therefore, the conclusions to be drawn from these findings are far from clear. Even if the most likely explanation of telescoping on the first day and some drop-out in reporting due to fatigue later on, are acceptable, no simple solution can be concluded. One approach is not to use the first day recordings for estimation, or to collect "bounding" information by recall for some previous days not for use in estimation. However, there is always a risk in increasing the overall bias when trying to reduce only one source of bias. Thus, maybe a better approach would be to retain the high level days on the justification that any exaggerations arising from that source merely offset omissions resulting on later days. Further research is needed to reach more conclusive results, as to the existence of such biases in the survey estimates and to determine changes needed in the survey design in order to overcome them.

### 3.3 Sampling Variances of Different Diary Lengths

Although sampling errors constitute only one part of the total error of the survey estimates, their role may increase, the more detailed are the estimates and the smaller is the sample size. When differences between groups or between two consecutive surveys are studied, sampling errors may become dominant, since biases of similar magnitude might cancel each other out. Thus, in determining the most efficient length of the recording period, the effects on the sampling variance versus the cost also have to be considered.

For this purpose, let us examine the efficiency of different lengths of diary periods, each regarded as a continuous recording period. Without loss of generality, let us assume for simplicity, that the sample is a simple random sample and that it is evenly distributed between non-overlapping diary periods.

Let us denote by  $p$  the diary period of sub-sample  $p$  ( $p=1,2,\dots,T$ ), where  $T$  is the number of the different diary periods of the same length over the survey period, say a year. Thus, for example,  $T=12$  for a monthly diary,  $T=24$  for a half-month diary,  $T=26$  for a two-week diary, etc.

Let  $n(T)$  denote the total size of the sample, and  $n_p(T) = \bar{n}(T)/T$  the size of each sub-sample, for a diary period corresponding to  $T$ .

Let  $X_{ip}$  be the value of the investigating variable (e.g., expenditure) of unit  $i$  in period  $p$ .

The simple estimate of the average per unit in period  $p$  corresponding to  $T$  is

$$\bar{x}_p(T) = \frac{1}{n_p(T)} \sum_i X_{ip}$$

and the simple estimate of the average per unit in the entire period is

$$\bar{x}(T) = \frac{1}{T} \sum_p \bar{x}_p(T).$$

The variance of  $\bar{x}(T)$  under the above assumptions is

$$\text{Var}[\bar{x}(T)] = \frac{1}{T} \sum_p \text{Var}[\bar{x}_p(T)] = \frac{1}{T} \sum_p \sigma_p^2(T), \text{ when } 1 - n_p(T)/N \approx 1.$$

Let us denote by  $c(T)$  the investigating cost per unit of the sample for a diary period that corresponds to  $T$ . Thus,  $n(T) = C/c(T)$  for a fixed total investigating cost,  $C$ . Therefore, the optimal length of the diary for a fixed total cost is the one for which

$$\text{Var}[\bar{x}(T)] = T c(T) \sum_p \sigma_p^2(T) \text{ is minimal.}$$

The efficiency of the length of a diary period based on  $T^*$  in comparison to one based on  $T$  is

$$E(T^*/T) = \frac{\text{Var}[\bar{x}(T^*)]}{\text{Var}[\bar{x}(T)]} = \frac{T^*}{T} \left[ \frac{c(T)/c(T^*)}{\sum_p \sigma_p^2(T) / \sum_p \sigma_p^2(T^*)} \right].$$

As an example, let us consider two lengths of a diary period:

- (1) A full-month diary, where  $T=12$  and  $p$  denotes a calendar month.
- (2) A half-month diary, where  $T^*=24$  and  $p(1)$  and  $p(2)$  denote the first and the second halves of each month  $p$ , respectively.

In this case  $E(T^*/T) = \frac{1}{2} [c(T)/c(T^*)] [1+\delta]$ , where,

$$1+\delta = \frac{\sum_p \{\sigma_{p(1)}^2 + \sigma_{p(2)}^2 + 2\text{Cov}[p(1), p(2)]\}}{\sum_p \{\sigma_{p(1)}^2 + \sigma_{p(2)}^2\}}$$

and  $\delta$  is approximately the correlation between expenditures of the two halves of all months (on the assumption that  $\sigma_{p(1)}^2 = \sigma_{p(2)}^2 = \sigma^2$  for all  $p$ ).

Although the cost per unit for a half-month diary is lower than for a full-month diary, it is more than half. On the other hand, the variance between months when using a full-month diary will be higher than when using two half-month diaries for two independent sub-samples, only if the covariance between the two half-months is positive. This is expected to be the case for frequent expenditures, such as often-purchased food commodities. However, this may not be the case with infrequent expenditures, where a negative covariance may be expected.

An analysis of the relative efficiency of two such different diary periods was made in Israel based on a sample of 2,250 households in the 1975/76 FES (Kantorowitz, 1981). A half-month diary was compared to the full-month diary used in the survey. Although it would have been better to have compared a two-week diary (since it is for a number of reasons preferable to one of a half-month), the available data did not allow simulation required for estimation of sampling variances.

The sampling variances for the two alternatives were estimated under the following assumptions: (1) sampling design effects are similar for different groups of items for each method; (2) the net sample is evenly distributed over all diary periods; (3) possible changes in recording in the diaries by the respondents between the two halves of the month have no effect on the estimated variances; (4) the rates and characteristics of non-response are the same for the two alternatives; and (5) there are no other differences in the inquiry method.

A detailed cost analysis of the two methods was carried out, taking into account many components, such as inquiry cost in the various stages, travel and per diem costs, editing, coding, keying, etc. Some of the cost components for the half-month period could only be assessed, because it had never been actually used. From this cost-function analysis it was estimated that the ratio of the cost per respondent for a full-month diary to that of a half-month was 1.45 to 1 (including the interviewing costs).

The efficiency of the half-month diary relative to the full-month diary for estimates of average expenditure per household was estimated as the inverse ratio of the respective sampling variances for samples of the same cost. This was done for major groups of items and for sub-groups of items, for the entire population as well as for some main sub-groups of the population.

The half-month method would be preferable whenever the efficiency is greater than 1. The efficiencies calculated by using only the diary data,  $E(k)$ , for each group  $k$  are relevant to expenditure estimates derived only from the diary data. For the other expenditure groups, an overall efficiency which takes into account the contribution of the data of the questionnaires as well, has to be considered. For estimates that are based on both the diary and the questionnaire, the overall efficiency,  $\epsilon(k)$ , was also calculated, by taking into account the percentage of the estimate that is based on the diary,  $\pi(k)$ . It should be noted that this was done for the recall period that was used in that survey, i.e., 12 months for most of the data collected by the questionnaire. Of course, a different recall period may result in different results. The results of this analysis for major groups of items for the entire population are shown in Table 3.6.

TABLE 3.6  
Efficiency of Half-Month Diary Relative to Full-Month Diary  
Israeli 1975/76 FES

| Major expenditure group (k)         | Percentage of estimate based on diary $\pi(k)$ | Correlation between two half-months $\delta(k)$ | Efficiency of diary estimate $E(k)$ | Efficiency of total estimate $\epsilon(k)$ |
|-------------------------------------|--|---|-------------------------------------|--|
| TOTAL EXPENDITURE                   | 48.5   | 0.49  | 1.18                                | 1.36                                       |
| Food (excluding Vegetables & Fruit) | 100.0  | 0.56  | 1.26                                | 1.26                                       |
| Vegetables & Fruit                  | 100.0  | 0.55  | 1.25                                | 1.25                                       |
| Clothing & Footwear                 | 100.0  | 0.17  | 0.86                                | 0.86                                       |
| Miscellaneous Goods & Services      | 50.3   | 0.32  | 1.00                                | 1.28                                       |
| Household maintenance               | 41.5   | 0.25  | 0.93                                | 1.12                                       |
| Transport & Communication           | 40.7   | 0.28  | 0.96                                | 1.31                                       |
| Education, Culture & Entertainment  | 34.3   | 0.31  | 0.98                                | 1.35                                       |
| Furniture & Household equipment     | 20.7   | 0.09  | 0.80                                | 1.35                                       |
| Health                              | 10.2   | 0.25  | 0.93                                | 1.39                                       |
| Housing                             | 0.0  | -   | -                                   | 1.45                                       |

A positive correlation,  $\delta(k)$ , between the expenditures of the diary in the two half-months was found for Total Expenditure, for all the major groups and for almost all the sub-groups. However, for many sub-groups the correlation was quite close to zero and it is to be expected that for more detailed groups of expenditure the correlation may even be negative.

The efficiency  $E(k)$ , if only the diary data were used for the survey estimates (which is the case for Food, Vegetables & Fruit, Clothing & Footwear, as well as for some sub-groups), was found to be greater than 1 for the Total expenditure (1.18), but only for two major groups: Food (1.26) and Vegetables & Fruit (1.25). Except for most of the sub-groups within these two major groups, only for a few other sub-groups  $E(k)>1$ , such as Cigarettes & Smoking ( $E(k)=1.31$ ), or for Public Transport ( $E(k)=1.09$ ).

As mentioned above, the ratio of the cost per respondent in the full-month diary to that in the half-month was 1.45. Thus, reducing the diary period from a full-month to half a month, the sample size can be increased by 1.45 for a given cost. This means that for estimates of expenditures that are based only on the questionnaires,  $\epsilon(k)=1.45$ . This is of considerable importance, since most estimates of this kind are usually subject to higher sampling variance than those based only on the diary.

The overall efficiency for Total Expenditure was  $\epsilon(k)=1.36$  and for all major groups  $\epsilon(k)>1$ , with the exception of Clothing & Footwear where  $\epsilon(k)=0.86$ . The efficiency for sub-groups of expenditure within the major groups (not presented here) was usually lower, except for 10 sub-groups that were based almost entirely on the questionnaire with  $\epsilon(k)=1.45$ . For example, whereas for Food or for Vegetables & Fruit  $\epsilon(k)=1.26$ , for the 18 sub-groups within these two groups  $0.91\leq\epsilon(k)\leq1.28$ ; for Transport & Communications  $\epsilon(k)=1.31$ , while for its sub-groups  $1.07\leq\epsilon(k)\leq1.41$ .

Altogether, out of all 62 sub-groups, for 10 sub-groups  $\epsilon(k)=1.45$ , for 10 sub-groups  $1.30\leq\epsilon(k)<1.45$ , for 15 sub-groups  $1.10\leq\epsilon(k)<1.30$ , for 7 sub-groups  $1.00\leq\epsilon(k)<1.10$ , and for 22 sub-groups  $\epsilon(k)<1.00$  (10 of them of Clothing & Footwear).

The estimates for the major group of Clothing & Footwear and its sub-groups have a priori relatively high sampling variances, thus reducing the diary period will only make them even less efficient. The data on expenditures on Clothing & Footwear must, therefore, be collected for a longer reference period, either by using a separate diary (or even a separate sample), or by recall.

An overall measure of the efficiency of the two diary periods, for all the expenditure groups may be calculated. Three possible criteria may be used:

- (1)  $\phi^2_1 = \sum_k \sigma^2(k)$ , the sum of all variances, which will depend on the absolute amounts spent for each group of items  $k$ , thus giving more importance to groups with higher variances and probably of higher expenditures;
- (2)  $\phi^2_2 = \sum_k \sigma^2(k)/x^2(k)$ , the sum of all relative-variances, which would attribute equal importance to each group irrespective of its size;
- (3)  $\phi^2_3 = \sum_k \sigma^2(k)/x(k)$ , which is a compromise between the above two criteria.

Based on each of these three functions for each of the two diary periods, an overall average efficiency  $\epsilon^*$  was calculated and the following results were obtained:

For the 10 major groups:  $\epsilon^*_1=1.27$   $\epsilon^*_2=1.24$   $\epsilon^*_3=1.22$   
 For the 62 sub-groups:  $\epsilon^*_1=1.27$   $\epsilon^*_2=1.15$   $\epsilon^*_3=1.13$

Despite the inefficiency of reducing the diary period for Clothing & Footwear, overall efficiency was achieved, whichever of the above three criteria was used. Thus, based on considerations of sampling variances, the half-month diary was found to be more efficient than the full-month diary.

The conclusion that the diary period should be reduced was supported by the fear of biases expected from various sources with a too-long diary period. However, the problem with Clothing & Footwear will then become even more serious.

It should be noted that different spending behavior of families may result in a completely different picture. For instance, whereas in the recent past, most fresh food like bread, milk and dairy products, meat etc., were purchased very often (even daily), nowadays, because of the use of refrigerators and freezers, these food commodities are not purchased so frequently. Also, spending patterns may not be similar in different countries, thus the optimal length of the diary period may also vary between countries. However, it is interesting to point out that a two-week diary was recommended in many countries on the basis of similar studies.

### **3.4 Types of Diaries**

Two main types of diaries can be used for FES's:

- (a) An open-type diary;
- (b) A structured-type diary.

This distinction relates primarily to the order of recording the various expenditures in the diary and not to the way of recording the details of each expenditure, since this is always structured, i.e., the required information for each expenditure is recorded in pre-specified columns of the diary (item description, quantity, payments, etc.). In the open-type diary, the purchases are entered in the order in which they were made. Usually, a list of products and services (not exhaustive) is somehow attached to the diary, e.g., printed on the front page of the diary. Structured-type diaries may be structured in various levels and according to various classifications. Of these, a product-diary is the most widely used, although also an outlet-diary may be considered. Open-type diaries may use separate pages for each day, but daily pages could be inefficient the more specifically structured the diary is.

The recordings in outlet-diaries are made according to the type of the store where the purchase was made, e.g., grocery, gas station, drug store, restaurant, etc. Thus, the diary is structured so that in each page of the diary, or parts of pages, pre-specified lines are used for the recording of the different items for each type of outlet.

A product-diary is structured in a similar way, but the diary is arranged by product groups, such as dairy products, meat, fish, fresh fruit, bread, clothes, household supplies, etc. Different levels of structure may be used, from only a few wide groups to many detailed ones. For each product group, blank lines for recording the purchases are provided under each of the section headings. The most structured diary is the closed-type, in which for each product group each line of the diary is used for one specific item and the item's description is pre-printed. Since the diary cannot include as many lines as the number of all products, additional blank lines are left for each of the product groups.

The structured diary can help the respondent to make sure that no expenditure (or receipts) are forgotten and may help to better distinguish between commodities. Also, with such a diary, extra information on certain special items, such as the age and sex of clothing items for a better classification, can be more easily structured into the diary.

When a product-type diary is used, coding can be simplified by using pre-printed codes. On the other hand, a large number of headings can confuse the respondent and the searching time for the appropriate line may be longer. Errors by the family in the entry line can lead to incorrect coding. Furthermore, respondents may be influenced by the listed headings and not report on expenditures not specified in the diary. The outlet-type diary is also subject to classifying errors, particularly for stores that sell groceries, drugs and clothing under one roof.

Preliminary experiments with different types of diaries were conducted in the United Kingdom in the early fifties (Kemsley and Nicholson, 1960). The results of these experiments did not show significant differences either in the response rate or in the reported average expenditure. Therefore, since less work is involved in coding if the diary entries are grouped together under broad headings of expenditure, it was recommended to structured diaries rather than open-type ones.

A pilot study conducted for the U.S. Bureau of Labour Statistics (Sudman and Ferber, 1971) included an attempt to explore the effectiveness of three types of diaries of the same content:

- (a) An Open-type diary;
- (b) An Outlet-type diary;
- (c) A Product groups-type diary.

A sample of 525 eligible households were split between the three types, half of them for the outlet-type, and a quarter for each of the other two types. Each of these sub-samples was further divided into four sub-samples, of approximately the same size, and households in each of them were asked to keep a diary for 1 week, 2 weeks, 3 weeks or 4 weeks. The main results of this study are presented in Table 3.7.

Table 3.7  
Effect of Diary Type on Extent of Co-operation  
Percentages(\*)

| Extent of co-operation  | Total | Type of diary |        |         |
|-------------------------|-------|---------------|--------|---------|
|                         |       | Open          | Outlet | Product |
| Kept at least 1 diary   | 87    | 79            | 81     | 84      |
| Kept at least 2 diaries | 81    | 55            | 69     | 76      |
| Kept at least 3 diaries | 66    | 46            | 56     | 68      |
| Kept all the 4 diaries  | 43    | 24            | 47     | 55      |

(\*) Percentages are out of all households asked to keep a diary for at least that length of time.  
Source: Sudman and Ferber (1971).

Although the sample was relatively small, the results indicated that by using an open-type diary less co-operation was obtained and it seems that record keepers find their task easier when the diary is somewhat structured, with a preference for the product-type, rather than the outlet-type. As a result of this study, a structured product-type diary with a two-week recording period was used in the U.S.A. since the 1972-73 FES.

Following are examples of the types of diaries in use in other countries. In the United Kingdom, a double page for each day is divided into 10 major categories and within each category the respondent himself has to describe what was purchased. A similar structured product-diary is used for the Canadian Food Expenditure Survey. In Italy, a double page for each day is divided into 4 categories with 62 lines, of which 45 contain pre-printed expenditure headings where no further description of the purchased goods can be added. In Greece, an open-type chronological diary is used for the family as a whole, whereas a daily personal diary (also used) is divided into 9 major groups of small expenditures (restaurants, cafes, alcohol, tobacco, theatre, etc.), with a different number of lines beneath each category. In Israel, an open-type diary is used (with a list of products printed on the back of the diary), with separate pages for each day. Each page is divided into two parts: the first for all expenditures actually purchased by the family and the second for all goods and services received in kind.

## CHAPTER 4. THE QUESTIONNAIRE

As was mentioned in Chapter 2, the estimates of FES's are almost always based not only on the diary data, but also on supplementary data that are collected by using one or more retrospective questionnaires, that relate to a given reference period.

### 4.1 The Reference Period

In some surveys, the reference period is the same calendar period for all the sampled households, e.g., all households are interviewed within a short time after the end of the yearly diary survey and the reference period is the preceding calendar year (as in Sweden, Canada and other countries). Another survey design is based on "moving" reference periods, where for each sub-sample a different reference period is used during the survey year, say, the last three months or the last twelve months.

The length of the reference period of the questionnaire (not necessarily the same for all the different groups of items), like that of the diary, may affect the reliability of the survey estimates. Sampling variances will increase for a fixed net sample size, the shorter the reference period is and the higher is the variability between partial periods of the same households. Since for most of the questionnaire's items this variability is higher than that of the diary's items, the length of the questionnaire's reference period has a greater effect on sampling variances. This is also the case with the reliability of estimates of distributions and parameters of variability of the population. The shorter is the reference period, the higher probably will be the estimated variability and the characterization of households by their annual level of income and expenditure will be prone to greater discrepancies. On the other hand, whereas with a shorter reference period of the diary the gain in the investigation cost is substantial, with the questionnaire it is marginal. Thus, the sample size may be only slightly increased with a shorter reference period, for the same overall cost.

With a longer reference period, an increase of the non-response rate is expected due to greater burden that is imposed on the respondents, but not as much as with a longer diary recording period.

As with all data collected by retrospective questionnaires, recall effects may result in biases in the survey estimates and the rate of item non-response may increase due to recall problems, requiring more imputations. Recall effects, either because poor information may be supplied by the respondents, or because they forgot to report some information, or report wrongly events that did not occur in the reference period (i.e., telescoping), depend primarily on the length of the recall period.

Of all these factors affected by the length of the recall period, most recall effects (and, to some extent, considerations of non-response and cost) favour a reduction of the length of the period, while the other factors (and, sometimes, certain effects due to telescoping) favour the opposite.

Since the reference period need not be the same as the recall period, i.e., may be divided into a few recall periods that cover the entire reference period, the solution is to reduce the recall period, but to keep a longer reference period, say a year, by conducting several interviews during the survey period. Such a survey design is used in the U.S.A. current FES, where four consecutive 3-month recall period interviews are conducted for each household. However, with this solution the cost increases significantly with the number of interviews per household and the non-response rate is also expected to increase. Furthermore, with such a survey design, changes in the respondents' purchasing behavior may occur and thus introduce bias into the survey estimates.

In order to decide what is the optimal length of the recall period and, to study the efficiency (cost versus the effects on the survey estimates), of conducting several interviews instead of one, the evaluation of the mutual effect of all the above-mentioned factors, is required. This is not always simple, because not all the relevant data for such evaluation are available, or cannot be obtained at a reasonable cost. Several evaluation studies that have been conducted are presented in the following sections.

## 4.2 Recall Effects

Recall effects depend on a number of factors, among them the type of the investigated variables, the respondent characteristics and the questionnaire wording, but the most significant is the length of the recall period. The longer the recall period, the less reliable is the respondent's memory and the less accurate are likely to be the survey results.

Two major types of recall errors can be distinguished:

- a) Memory effects - Under-reporting of events that the respondent has forgotten which leads to downward bias in the survey estimates. In addition, if the respondent remembers the occurrence of the events, he may not provide accurate details, which may lead to biases in either direction.
- b) Telescoping effects - When the respondent remembers the events, but places them incorrectly on the time-axis. These may be of two types, a forward-telescoping when events are reported later than they actually occur, or, a backward-telescoping when events are reported as if they occurred earlier. Telescoping will primarily affect the accuracy of the survey estimates when events outside the recall period are mistakenly placed inside, or vice versa, but also when within the recall period, events are wrongly placed. Telescoping effects may lead either to over-reporting, as well as under-reporting, depending on the direction of the telescoping.

Telescoping is likely to be dominant when a short recall period is used especially for infrequent events of considerable importance. Memory effects are likely to be more significant when a longer recall period is used especially for frequent and less important events.

Therefore, in order to reduce both types of recall effects: (1) the recall period should be as short as possible; (2) an additional boundary interview should be conducted preceding the reference period; and (3) the recall period should terminate at the time of the interview. The boundary data are not to be used for the survey estimates.

Such a solution (adopted in the U.S. survey) is, however, not so simple to adopt in practice, because it is too costly because of the additional interview required for each household. Furthermore, a shorter reference period may lead to a significant increase in sampling variances, unless either the sample or the number of interviews per household are increased, i.e., the survey becomes more costly. The solution for designing a survey, for limited sources, lies in obtaining estimates of recall effects for various lengths of recall periods and at the same time assessing the effects on the other factors mentioned earlier.

A study on recall effects is at present being conducted in Israel on the data collected by the interview in the 1986/87 FES (based on a sample of 5,000 households). Some preliminary results are presented in the following. In this survey, a 3-month recall period was used for the majority of expenditures, and a 12-month period for a few infrequent expenditures. No boundary interview was held prior to the recall period. However, each sampled household was asked to report on expenditures since the beginning of the previous three (or twelve) calendar months until the time of the interview. Data referring to the time after the end of the previous month were collected but were used only as a boundary information.

The 3-month "moving" period related to 15 calendar months from March 1986 to May 1987 (although not all months are represented equally) as illustrated in Figure 4.1. For each of the 11 calendar months, May 86 - March 87, data are available for three recall months: (1) the most "recent" month; (2) the "middle" month; and (3) the most "distant" month. For each calendar month, the data of the three recall months relate to different sub-samples of households. The data on the four calendar months at the two extremities were excluded from the study in order to avoid possible effects of the timing of purchases.

Figure 4.1  
Recall Months by Interview Month for the 3-Month Recall Periods

| Interview month | Recall month |      |      |      |      |      |      |       |       |       |      |      |      |      |      |
|-----------------|--------------|------|------|------|------|------|------|-------|-------|-------|------|------|------|------|------|
|                 | 3/86         | 4/86 | 5/86 | 6/86 | 7/86 | 8/86 | 9/86 | 10/86 | 11/86 | 12/86 | 1/87 | 2/87 | 3/87 | 4/87 | 5/87 |
| 6/86(*)         | [3]          | [2]  | 1    |      |      |      |      |       |       |       |      |      |      |      |      |
| 7/86            |              | [3]  | 2    | 1    |      |      |      |       |       |       |      |      |      |      |      |
| 8/86            |              |      | 3    | 2    | 1    |      |      |       |       |       |      |      |      |      |      |
| 9/86            |              |      |      | 3    | 2    | 1    |      |       |       |       |      |      |      |      |      |
| 10/86           |              |      |      |      | 3    | 2    | 1    |       |       |       |      |      |      |      |      |
| 11/86           |              |      |      |      |      | 3    | 2    | 1     |       |       |      |      |      |      |      |
| 12/86           |              |      |      |      |      |      | 3    | 2     | 1     |       |      |      |      |      |      |
| 1/87            |              |      |      |      |      |      |      | 3     | 2     | 1     |      |      |      |      |      |
| 2/87            |              |      |      |      |      |      |      |       | 3     | 2     | 1    |      |      |      |      |
| 3/87            |              |      |      |      |      |      |      |       |       | 3     | 2    | 1    |      |      |      |
| 4/87            |              |      |      |      |      |      |      |       |       |       | 3    | 2    | 1    |      |      |
| 5/87            |              |      |      |      |      |      |      |       |       |       |      | 3    | 2    | [1]  |      |
| 6/87(*)         |              |      |      |      |      |      |      |       |       |       |      |      | 3    | [2]  | [1]  |

(\*) The size of the sample of months at the two extremities was about half that of the other months.

For the 3-month analysis of recall effects let us denote by  $y_{im}(r)$  the reported expenditure on a certain item by household  $i$  ( $i=1,2,\dots,n$ ) in calendar month  $m$  ( $m=1,2,\dots,11$ ) which is the  $r$ -th recall month of that household ( $r=1,2,3$  for the "recent", "middle" and "distant" months, respectively).

Let  $x_{im}(r) = d_{im}(r) y_{im}(r)$ , where

$$d_{im}(r) = \begin{cases} 1 & \text{when } y_{im}(r) > 0 \\ 0 & \text{when } y_{im}(r) = 0 \end{cases}$$

For the sake of simplicity, let us assume that a simple random sample of size  $n$  participates in each month and that simple estimates are used, although this was not actually the case with the analysis of the data.

The estimates of the average expenditure per household, the proportion of households who reported expenditure on the item and the average expenditure per reporting household, in calendar month  $m$  for recall month  $r$ , are:

$$\bar{x}_m(r) = \sum x_{im}(r)/n, \quad p_m(r) = \sum d_{im}(r)/n \quad \text{and} \quad \bar{y}_m(r) = \sum y_{im}(r)/n = \bar{x}_m(r)p_m(r).$$

Then the Recall Effects (RE) over all the 11 calendar months of the "middle" and the "distant" recall months, in comparison to the "recent" month, can be estimated as follows:

$RE(2) = \bar{z}_m(2) / \bar{z}_m(1)$  for the "middle" recall month, and

$RE(3) = \bar{z}_m(3) / \bar{z}_m(1)$  for the "distant" recall month,

where  $\bar{z}_m(r)$  may be either  $\bar{x}_m(r)$ , or  $\bar{y}_m(r)$ , or  $\bar{p}_m(r)$ .

The variances of the estimated recall effects have to be estimated as well, taking into account the study model, so that a Z-test can be applied.

Some results of the Israeli study for recall effects by month of recall, for the 3-month period, are shown in Table 4.1. The analysis of recall effects was made only for selected items of the questionnaire whose estimates were subject to relatively low variance.

Table 4.1  
Recall Effects by Month of Recall - 3 Month period  
Israeli 1986/87 FES

| Expenditure group                  | Monthly average reporting percent | Monthly recall effect relative to recent month |         |                        |         |        |         |
|------------------------------------|-----------------------------------|--|---------|------------------------|---------|--------|---------|
|                                    |                                   | of Percent of HH with non-zero expenditure     |         | of Average expenditure |         |        |         |
|                                    |                                   | Middle   | Distant | Middle                 | Distant | Middle | Distant |
| All items included in the study    | 31.7                              | 0.92**   | 0.78**  | 1.07                   | 1.30**  | 0.98   | 1.00    |
| Jewellery                          | 2.7                               | 0.61**   | 0.59**  | 1.57**                 | 1.57**  | 0.90   | 0.90    |
| Watches                            | 2.1                               | 0.74**   | 0.59**  | 1.04                   | 1.20    | 0.78*  | 0.77    |
| Quilts & Blankets                  | 1.1                               | 0.77*  | 0.69*   | 1.66*                  | 1.23    | 1.03   | 0.88    |
| Bedding & Towels                   | 2.3                               | 0.86   | 0.70**  | 1.13                   | 1.23    | 0.98   | 0.79    |
| Dwelling repairs & renovation      | 3.3                               | 0.88*  | 0.54**  | 1.27                   | 1.00    | 0.99   | 0.71*   |
| Household help                     | 4.2                               | 0.97   | 0.69**  | 1.04                   | 0.71**  | 1.02   | 0.66**  |
| Furniture                          | 5.8                               | 0.99   | 1.01    | 0.92                   | 1.03    | 0.94   | 1.07    |
| Electrical equipment               | 8.1                               | 1.00   | 0.80**  | 0.98                   | 1.08    | 1.00   | 0.91    |
| Durables for Culture&Entertainment | 2.8                               | 0.92   | 0.82*   | 1.09                   | 1.29*   | 0.95   | 1.01    |
| Pictures & Frames                  | 0.8                               | 0.68*  | 0.56**  | 1.13                   | 1.76*   | 0.74   | 0.92    |
| Dental treatment                   | 8.8                               | 0.88**   | 0.50**  | 1.01                   | 0.79*   | 0.88   | 0.56**  |
| Private physician's fees           | 4.9                               | 0.99   | 0.52**  | 1.29**                 | 1.45    | 1.28*  | 0.95    |
| Optical rehabilitation aids        | 3.7                               | 0.80**   | 0.62**  | 1.27**                 | 0.93    | 0.98   | 0.80*   |
| Vehicle licence                    | 2.7                               | 0.77**   | 0.63**  | 0.83*                  | 0.66**  | 0.75*  | 0.52**  |
| Vehicle upkeep & repairs           | 6.6                               | 0.91   | 0.87*   | 1.06                   | 1.17    | 0.97   | 1.02    |

Note: recall effects significantly different from 1 are denoted: \*\* for  $\alpha \leq 1\%$  and \* for  $1\% < \alpha \leq 5\%$ .

For all the items that were included in the study, 35.2% of all households reported on at least one item in the recent month, 32.4% in the middle month and only 27.5% in the distant month, i.e., recall effects of 0.92 and 0.78 respectively. Declining rates (not of the same degree) of the percentage of households with non-zero expenditure were observed for almost all items, although not all of them were statistically significant. For some items the recall effect of the middle month is moderate, but is substantial in the distant month, e.g., for Dental treatment  $RE(2)=0.88$  and  $RE(3)=0.50$ , for Dwelling repairs & renovation  $RE(2)=0.88$  and  $RE(3)=0.54$ . For other items the rates of reporting decline more gradually, e.g., for Bedding & Towels  $RE(2)=0.86$  and  $RE(3)=0.70$ . For some items the main fall in reporting rates are observed in the middle months, e.g., for Jewellery  $RE(2)=0.61$  and  $RE(3)=0.59$ . For some items, such as Furniture, no recall effects were found.

The results for the average expenditure per reporting household indicate for many items recall effects in the opposite direction. For total reported expenditures of all the items included in the study, the average in the distant month is significantly higher than in the recent month RE(3)=1.30. For Jewellery, the averages of the middle and the distant month are about 1.6 higher than the average of the recent month. For Private physician's fees, the averages of the middle and the distant months are higher than of the recent month by 1.3 and 1.4, respectively. An opposite result is that of payments for Vehicle licences, where the averages are lower in the middle month, 0.83, and in the distant month, 0.66, as compared with the recent month.

Whereas, most of the recall effects for the rate of reporting were statistically significant, especially for the distant month, this is not the case with recall effects for the average expenditure per reporting household. This is because of the high variability between expenditures either on the same type of item, or on different items belonging to the same expenditure group. Consequently, the estimates of the overall recall effect relating to the average expenditure per household also suffer from high variances. Only for several items is there evidence that the highest average is in the recent month and the lowest in the distant one. For most other items, the recall effects are not significantly different from 1, although some recall effect is indicated.

It should be noted that when estimates of recall effects for the average expenditures per reporting household are calculated, differentiability of price changes between different items over the recall period must be taken into account. Although all expenditures in the entire survey period were adjusted according to changes in the general CPI so that they should all relate to the same price level, they were not adjusted for price changes of each item separately. Without adjusting the data by the specific CPI for each item, the measured recall effects do not eliminate the differentiability between-item price changes. However, for a short period of three months and the degree of inflation in the survey period, this was of minor importance.

A 12-month recall period was used for some items, as was mentioned above. The recall periods of the different sub-samples referred to 24 different calendar months (June 1985 to May 1987), as illustrated in Figure 4.2.

Figure 4.2  
Recall Months by Sub-Samples for 12-Months Recall Periods

| Interview month | Recall month |         |         |         |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
|-----------------|--------------|---------|---------|---------|----|----|----|---|---|---|---|---|------|---|---|---|----|----|----|---|---|---|---|---|------|---|---|---|----|----|----|--|--|--|--|--|
|                 | 1985         |         |         |         |    |    |    |   |   |   |   |   | 1986 |   |   |   |    |    |    |   |   |   |   |   | 1987 |   |   |   |    |    |    |  |  |  |  |  |
|                 | 6            | 7       | 8       | 9       | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6    | 7 | 8 | 9 | 10 | 11 | 12 | 1 | 2 | 3 | 4 | 5 | 6    | 7 | 8 | 9 | 10 | 11 | 12 |  |  |  |  |  |
| 6/86*)          | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 7/86            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 8/86            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 9/86            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 10/86           | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 11/86           | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 12/86           | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 1/87            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 2/87            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 3/87            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 4/87            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 5/87            | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |
| 6/87*)          | 12-11-10     | 9--8--7 | 6--5--4 | 3--2--1 |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |   |   |   |   |   |      |   |   |   |    |    |    |  |  |  |  |  |

\*) The sample size of these two months was about half that of the others.

Only two calendar months (May and June 1986) have 12 different recall months, two calendar months have 11 recall months, etc. Thus, a simple elimination of possible seasonality as was done with the 3-month recall period, is not feasible for the 12-months recall period. Even if recall months are grouped into quarters, only four quarters have all four different recall quarters. Thus, a more sophisticated model (not yet adopted) has to be applied to analyse recall effects by using most of the available data.

Nevertheless, a preliminary comparison of the data between the four recall quarters, averaged over all sub-samples, may provide some indication on the existence of recall effects, although the results have to be interpreted with caution since seasonality effects are not eliminated. For the average expenditure, the comparisons were calculated with and without adjusting for the differentiability in the price changes mentioned above. The results were affected by this adjustment, but not all of them to the same degree. Table 4.2 show the results for two examples, after this adjustment.

Table 4.2  
Recall Effects by Recall Quarter - 12-Months Period  
Israeli 1986/87 FES

| Expenditure group       | Monthly average reporting percent | Quarterly recall effect relative to recent quarter |        |        |                           |        |        |        |       |        |
|-------------------------|-----------------------------------|--|--------|--------|---------------------------|--------|--------|--------|-------|--------|
|                         |                                   | of Percent of households with non-zero expenditure |        |        | of Average expenditure(*) |        |        |        |       |        |
|                         |                                   |  |        |        | per reporting HH          |        |        | per HH |       |        |
|                         |                                   | Second   | Third  | Fourth | Second                    | Third  | Fourth | Second | Third |        |
| Total                   | 4.6                               | 0.80**   | 0.65** | 0.47** | 1.12                      | 1.45** | 1.41** | 0.89   | 0.95  | 0.61** |
| Recreation & Excursions | 4.4                               | 0.80**   | 0.66** | 0.47** | 1.13*                     | 1.39** | 1.23** | 0.89   | 0.94  | 0.53** |
| Trips abroad            | 1.1                               | 0.86*  | 0.84*  | 0.62** | 1.05                      | 1.14*  | 1.18** | 0.90   | 0.96  | 0.73   |

(\*) After adjusting for the differentiability in the item price changes.

Note: recall effects significantly different from 1 are denoted: \*\* for  $\alpha \leq 1\%$  and \* for  $1\% < \alpha \leq 5\%$ .

Substantial (and statistically significant) recall effects for the percentage of households reporting on the related expenditures, were observed. Total expenditure on Trips abroad and Recreation & Excursions was reported by 6.3% of the households in the recent quarter, 5.0% in the second one, 4.1% in the third quarter and only 3.0% in the most distant quarter. Recall effects for Trips abroad are smaller than for Recreation & Excursions, e.g., for the most distant quarter, the recall effects are 0.62 and 0.47, respectively. Part of the relatively high reporting rate on these items in the recent quarter may be explained by biases due to delays in investigating households that were absent at the time that was initially designated for their investigation.

The recall effects of the average expenditure per reporting household for the aggregate of the two expenditure groups have a pattern similar to that found with many of the items with a 3-month recall period, i.e., the more distant the quarter, the higher is the average: 1.12, 1.45 and 1.41 for the second, third and fourth quarters, respectively. This is almost the case also for each group separately: for Trips abroad the effects are 1.05, 1.14 and 1.18 and for Recreation & Excursions, 1.13, 1.39 and 1.23, for the three recall periods. The overall recall effect on the average per household still show decreasing averages with more distant recall periods.

The results of this study even though they do not measure "pure" recall effects nonetheless provide evidence of the existence of severe recall effects in the data, for both lengths of the recall period. Therefore, further analysis is required for a better understanding of the nature of recall effects in FES's in order to try to find ways of reducing them in future surveys.

Recall effects were also analysed in the 1984 U.S. Consumer Expenditure Interview Survey, by type of respondent, e.g., family type, age of respondent, the use of records etc. (Silberstein, 1989). The U.S. survey is conducted under a different design from the Israeli one. The primary difference that is relevant to recall effects is the use of a bounding interview prior to the first 3-month recall period, and, for each of the other three consecutive 3-month recall periods the previous interview serves as a bounding interview. The reporting period includes the portion of the current month up to the interview day and expenses reported for that month are available to the interviewer in the following wave to minimize duplications. This study included units that participated in the third quarter of 1984 covering expenses made between April and August and the analysis related to the two expenditure categories, Apparel and House furnishings. Recall effects were estimated by comparisons between the three recall months of each unit, ignoring possible seasonal shopping effects. Only cases that reported at least one expense in the categories under analysis were included and this could be a further cause of bias in the estimated recall effects, however, with reduced variances.

The percentage of cases that reported some expenses are shown in Table 4.3 by recall month and the results show recall effects of a similar magnitude for both Apparel and House furnishings, about 15% less reporting in the middle month and about 25% less in the distant month, as compared to the recent month.

Table 4.3  
Percent of Consumer Units with Reported Expenses by Recall Month  
U.S. 1984 Consumer Expenditure Interview Survey

| Expenditure category | Reporting Percent by recall month |        |         | Reporting percent Relative to recent month |        |         |
|----------------------|-----------------------------------|--------|---------|--|--------|---------|
|                      | Recent                            | Middle | Distant | Recent                                     | Middle | Distant |
| Apparel              | 83                                | 69     | 61      | 1.00                                       | 0.83   | 0.73    |
| House furnishings    | 71                                | 60     | 55      | 1.00                                       | 0.85   | 0.77    |

Source: Silberstein (1989).

For further analysis of the data, each unit was assigned to one of three groups of recall effects according to the value of the expenses reported for the recent recall month as a percentage of its total expense reported for the three recall months. Three levels of the individual recall effect were defined: "None" with 0% to 35%, "Moderate" with 35% to 75%, and "Great" with 75% to 100%. The distributions and the mean expenditures by these three levels are given in Table 4.4.

Table 4.4  
Mean Expenditures by Levels of Individual Recall Effect  
U.S. 1984 Consumer Expenditure Interview Survey

| Recall effect level | Apparel          |                          | House furnishings |                          |
|---------------------|------------------|--------------------------|-------------------|--------------------------|
|                     | Percent of units | Monthly mean expenditure | Percent of units  | Monthly mean expenditure |
| TOTAL               | 100              | 102                      | 100               | 141                      |
| None                | 44               | 106                      | 53                | 146                      |
| Moderate            | 29               | 125                      | 18                | 163                      |
| Great(*)            | 27               | 71                       | 29                | 119                      |

\*) About half reported all expenditure in the recent month.

Source: Silberstein (1989).

The average expense on furnishings is about 1.4 time of that on apparel, and this, as expected with items of larger expenses, may lead to smaller recall effects: 53% with no recall effects ("None") against 44%. The highest average reported expenses for the "Moderate" level of recall effect in both categories is quite surprising.

On the basis of the three categorical levels of recall effect for each unit, log-linear models were fitted to measure the type of interaction of each independent variable, i.e. the respondent characteristics, with the dependent variable, i.e., the recall effect level, after eliminating interactions between the independent variables. No more than three independent variables were included in any model due to sample size limitations (for further details, see Silberstein, 1989).

The results of this study (for apparel) show that recall effects are affected by the characteristics of the respondents and the interview performance. Some specific examples of these findings are as follows. Significant relationships with recall effects were found for age and education. Respondents with low education are the most likely to have great recall effects. Respondents of ages 25-44 show moderate effects, whereas older respondents (65 and over) exhibit either no effects or extreme effects. However, no recall effects are evident for older respondents when education is controlled.

It is reasonable to argue that size of family will influence recall effects, and the results indeed show that no effects were found for a family of 1 or 2 members, moderate effects were found with 3-4 members and moderate to great effects were observed for larger families. Recall effects appear to differ by family income, where families with lower incomes show greater recall effects.

#### 4.3 Length of the Reference Period and Sampling Variances

Many of the variables that are investigated for each sampled household by the FES questionnaire vary in time. Regular expenditures, such as payments for electricity, water, rent, etc., and probably income, vary less than infrequent large expenditures, such as on durable goods and recreation. For a given survey period, each sampled household may be regarded as a cluster of sub-periods, e.g., months or quarters. The reference period of the questionnaire for each sampled household may be either the entire survey period, usually 12-months, or a sample of one or more sub-periods of at least one month, and need not be of the same length for different items.

Different lengths of the reference period for a fixed sample size of households have different implications on the sampling variance, and these depend on the intra-cluster correlation between expenditures in the sub-periods within the yearly period.

Let us denote by  $X_{ij}$  the value of a certain investigating variable in household  $i$  ( $i=1,2,\dots,N$ ) in month  $j$  ( $j=1,2,\dots,M$ ).

Let  $S^2_b$  be the variance of the monthly average values between the  $N$  households, and  $S^2_w$  the variance between the monthly values within the households:

$$S^2_b = \sum (\bar{X}_i - \bar{X})^2 / (N-1) \quad \text{and} \quad S^2_w = \sum \sum (X_{ij} - \bar{X}_i)^2 / N(M-1)$$

where  $\bar{X}_i = \sum X_{ij}/M$  is the monthly average for household  $i$ , and  $\bar{X} = \sum \sum X_{ij}/MN$  is the overall monthly average per household.

The overall variance between households and months is

$$S^2 = \sum \sum (X_{ij} - \bar{X})^2 / (NM-1) \approx (1/M)S^2_b + (1-1/M)S^2_w.$$

Then, for a simple random sample of  $n$  households and  $(N-n)/N \approx 1$ , with a reference period of all  $M$  months, (one-stage sampling) the sampling variance is

$$\text{Var}_M(\bar{X}) \approx S_b^2/n.$$

When for each sampled household, a sample of  $m < M$  months are selected at random for the reference period (two-stage sampling), the sampling variance is

$$\text{Var}_m(\bar{X}) \approx S_b^2/n + (1/m - 1/M) S_w^2/n.$$

Thus, for a fixed sample size of households, with two-stage sampling, the sampling variance will be higher, the higher is the variability between months within the households and the smaller is  $m$ .

The relative increase in the sampling variance by reducing the reference period from  $M$  to  $m$  will be:

$$\alpha = \frac{\text{Var}_m(\bar{X})}{\text{Var}_M(\bar{X})} \approx 1 + (1/m - 1/M) S_t^2/S_b^2.$$

However, in FES's either a reference period of  $M=12$  months, or one sub-period  $t$  comprising  $m$  consecutive months out of  $T=M/m$  sub-periods, is used, e.g., 1 quarter out of 4, or 1 period of two months out of 6. In this case:

$$\alpha = 1 + (1-1/T)S_w^2/S_b^2, \text{ where } S_w^2 \text{ is the within-variance between the } T \text{ sub-periods.}$$

One extreme is when expenditures are evenly distributed over the survey sub-periods. Then,  $S_w^2=0$  and thus  $\alpha=1$ , e.g., the length of the reference period can be reduced with no increase in the sampling variance. However, this is rarely the case with the investigating variables of the FES's questionnaire.

At the other extreme are items that are either purchased by each household only once, or are not purchased at all during a period of one year, such as most of the expensive durable goods. It can be shown that relative increase in the variance due to reducing the length of the reference period for such variables under the assumption of equal expenditure per item, will be  $\alpha=1+(T-1)/(1-\pi)$ , where  $\pi>0$  is the probability of a household purchasing the item in the survey year. For example, the increase in sampling variance from reducing the reference period from 12 months to 3 months for variables of this type  $\alpha=1+3/(1-\pi)$ . Thus, as could be expected, the variance for items with this purchasing pattern for a 3-month period would be at least four times of that of a 12-month period, and the relative increase in the variance increases with  $\pi$ .

In practice, however, the investigating variables of the FES questionnaire lie between these two extremes. Some empirical comparisons of the variances for various lengths of reference periods are presented in the following.

The relative sampling variances from the Israeli 1979/80 and 1986/87 FES's were compared for various major groups of expenditures, for which most of the data were obtained from the questionnaires, and for income. In the 1979/80 FES, the reference period for these data was a 12-month period, whereas in the 1986/87 survey it was reduced to 3 months. The results of this comparison (after adjusting for the different size of the sample in the two surveys) are shown in Table 4.5.

These results may provide only some indication of the effect of reducing the reference period on the sampling variances. Because the two surveys do not relate to the same time, changing purchasing behavior may affect the results, and also because these surveys were not conducted under the same conditions with the same survey design, the comparison is of only limited validity.

Table 4.5  
Comparison of Relative Sampling Variances for a 12-Month and a 3-Month Reference Period  
1979/80 and 1986/87 Israeli FES's

| Expenditure major group              | Percent of HH reporting in 1986/87 | Expenditure percent from questionnaire | Relative sampling variance - 3-months against 12-months ( $\alpha$ ) |
|--------------------------------------|------------------------------------|--|--|
| <u>Total Income</u>                  | 100                                | 100                                    | 1.2  |
| Income from work                     | 76                                 | 100                                    | 1.6  |
| Income from pensions                 | 16                                 | 100                                    | 1.1  |
| <u>Total Expenditure(*)</u>          | 100                                | (*) 55                                 | 1.3  |
| Health insurance                     | 94                                 | 100                                    | 1.1  |
| Jewellery & Watches                  | 22                                 | 72                                     | 1.3  |
| Electrical equipment                 | 41                                 | 91                                     | 1.9  |
| Durables for Culture & Entertainment | 83                                 | 81                                     | 2.1  |
| Bedding & Towels                     | 17                                 | 78                                     | 2.2  |
| Furniture                            | 17                                 | 99                                     | 2.9  |
| Dental treatment                     | 28                                 | 100                                    | 2.9  |
| Dwelling repairs & renovation        | 67                                 | 83                                     | 4.1  |
| Dues & Donations                     | 80                                 | 78                                     | 7.3  |
| Legal & Financial services           | 16                                 | 78                                     | 8.6  |

(\*) Including also expenditures for which a 12-month reference period was used in both surveys.

Yet, despite these limitations the results show that for items with less variation between quarters, the increase in the sampling variance is moderate, e.g., for Total Income  $\alpha = 1.2$ , for Pensions and for Health Insurance  $\alpha = 1.1$ . For items with higher variation over time, the use of a shorter reference period leads to a substantial increase in variance, e.g., for Furniture and for Dental treatment  $\alpha \approx 3$ , for Dwelling repairs & renovation  $\alpha \approx 4$ , and for Legal & Financial services  $\alpha \approx 9$ . Such comparison for expenditures on more detailed items between the two surveys may be misleading, but it is reasonable to expect that reducing the reference period will result in a considerable increase in the sampling variances for specific items.

For some groups of expenditures, comparison of the sampling variances for different lengths of reference periods was made, by using only the data of the 1986/87 FES. For the same households in the sample, variances were calculated for sub-periods of the reference periods. For a few items for which a 12-month reference period was used, variances were calculated for four reference periods of 3, 6, 9 and 12 months, and for the other items with a 3-month reference period, for 1, 2, and 3 months. Some of the results are presented in Tables 4.6 and 4.7.

The results in Table 4.6 show that if the reference period for these items would have been reduced to 3 months, the sampling variances would have been increased by factors in the range 2.5-3.3. Just because a large increase in variance was expected for these items, a 12-month reference period was chosen, under the assumption of moderate recall effects (which was found to be not true in a later study). Although, as the results in Table 4.5 show (despite their limitations), that even larger increase in variances are associated with other groups of items, nevertheless, it was decided to use a shorter reference period for them, because of expected serious recall effects.

Table 4.6  
Relative Sampling Variance by Length of Reference Period  
Items of 12-Month Reference Period in the Israeli 1986/87 FES

| Expenditure item        | Sampling variance relative to 12-month period |          |          |           |
|-------------------------|---|----------|----------|-----------|
|                         | 3 months                                      | 6 months | 9 months | 12 months |
| Total                   | 2.74  | 1.52     | 1.12     | 1.00      |
| Recreation & Excursions | 2.64  | 1.58     | 1.18     | 1.00      |
| Hotels                  | 2.46  | 1.49     | 1.11     | 1.00      |
| Stay abroad             | 3.34  | 1.79     | 1.17     | 1.00      |
| Trips abroad            | 3.34  | 1.67     | 1.11     | 1.00      |

The results in Table 4.7 show for the total expenditure on all items (in the recall effects study), that if the 3-month reference period would have been reduced to two months, the variance would have increased by 1.4 and to one month, by 2.4. The length of the reference period has, as can be expected, different effects on the variances of different items, depending on the intra-months correlation. For example, with Household help, the length of the period has almost no effect on the variance, whereas for Durables for Culture & Entertainment the effect is substantial.

Table 4.7  
Relative Sampling Variance by Length of Reference Period  
Items of 3-Month Period in the Israeli 1986/87 FES

| Expenditure item                     | Sampling variance relative to 3-month period |         |         |
|--------------------------------------|--|---------|---------|
|                                      | 1 month                                      | 2 month | 3 month |
| All items included in study          | 2.38   | 1.36    | 1.00    |
| <u>Selected expenditure items</u>    |  |         |         |
| Jewellery                            | 2.47   | 1.34    | 1.00    |
| Watches                              | 1.74   | 1.00    | 1.00    |
| Quilts & Blankets                    | 2.40   | 1.32    | 1.00    |
| Bedding & Towels                     | 2.69   | 1.26    | 1.00    |
| Dwelling repairs & renovation        | 1.75   | 1.08    | 1.00    |
| Household help                       | 1.02   | 1.05    | 1.00    |
| Furniture                            | 2.68   | 1.50    | 1.00    |
| Electrical equipment                 | 2.86   | 1.40    | 1.00    |
| Durables for Culture & Entertainment | 3.25   | 1.43    | 1.00    |
| Pictures & Frames                    | 1.51   | 1.15    | 1.00    |
| Dental treatment                     | 2.34   | 1.18    | 1.00    |
| Private physician's fees             | 0.99   | 0.64    | 1.00    |
| Optical rehabilitation aids          | 2.16   | 1.26    | 1.00    |
| Vehicle licence                      | 2.70   | 1.40    | 1.00    |
| Vehicle upkeep & repairs             | 2.31   | 1.45    | 1.00    |

#### 4.4 Optimal Length of the Reference Period

When several alternative lengths of the recall period are available, a prime consideration in choosing among them is a comparison of the MSE (comprising the sampling variance and the square of the bias) of key statistics for a given cost. The sampling variance depends on the number of households in the net sample, the length of the reference period for each sampled household, the sample design, and on the weighting procedure. However, it should be noted that other components of variance may affect the survey estimates such as response variance.

Bias may be introduced into the survey estimates by recall effects, but also by procedural effects and other various factors. Although not only the recall effects will be affected by the length of the recall period, it is reasonable to assume that the other sources of bias can be ignored in choosing among alternative recall lengths.

Estimates of both the sampling variance  $V_k^2(l)$  and the recall bias  $B_k(l)$  for alternative recall lengths,  $l$ , can be used to determine the optimal recall period for each of the required survey estimates,  $k$ . This is done by comparing the  $MSE_k(l)=V_k^2(l)+B_k^2(l)$  for the alternative recall lengths. However, usually, no single recall length is best for all the required estimates. Furthermore, even for a single estimate, the best recall length for the total population is not always the best for various sub-groups.

One possibility is to choose an "average best" recall length for all the investigated variables. This may be done by taking into account the importance of the different variables and the different sizes of their MSE's. Another, is to determine different recall lengths for different groups of variables.

As far as costs are concerned, interviewing costs are about the same for different recall periods. This is not the case when the recall period is short and the reference period is long, i.e., several interviews are conducted for each household instead of one.

The sampling variances for various recall periods were discussed in section 4.3 and selected results from the Israeli FES's were presented under the assumption of no change in cost. The bias due to recall effects can be estimated, only if it can be assumed that the "recent" month is free of recall effects. If this is true, recall biases from different lengths of recall are estimated by comparing the estimates obtained by using the data of each recall period, against the estimates derived from the recent period. These results must be interpreted with some caution, since they may not distinguish between the bias due to the two types of recall effects, telescoping and memory effects. Thus, for example, the estimated bias of a 3-month recall period may include telescoping effects in the "distant" month, while for a 2-month or a 1-month recall period, telescoping effects are avoided because their data are bounded by the previous months. Furthermore, the intra-month telescoping cannot be distinguished from other recall effects. It should also be remembered that the estimates of the recall biases are subject to very high sampling variances.

Despite all these limitations, let us examine some of the results that were obtained from the data of the Israeli 1986/87 FES. Some examples are shown in Table 4.8 for the expenditures from a 3-month recall period and in Table 4.9 for those from a 12-month period.

For the estimate of Total expenditures on all the items that were included in the study, the sampling variance is the dominant component of the MSE, and the 3-month period is the best: the MSE of a 2-month period is 1.4 times higher, and of a 1-month period 2.2 times as high. A similar situation was found with Furniture and with Durables for Culture & Entertainment and Vehicle upkeep & repairs. However, for many other groups of items, the recall bias is the dominant component of the MSE for the 3-month period: for Dental treatment  $MSE(1)/MSE(3) \approx 0.2$ , for Household help the ratio is 0.4, for Watches the ratio is 0.5, and for Private physician's fees, 0.6.

Table 4.8  
Relative Errors of 3-Month Period and Comparison of MSE by Length of Reference Period  
Israeli 1986/87 FES

| Expenditure groups                     | 3 months period             |                          |                           | Relative MSE relative to 3 months |          |         |
|--|-----------------------------|--------------------------|---------------------------|-----------------------------------|----------|---------|
|  | Relative sampling error (%) | Relative recall bias (%) | Relative $\sqrt{MSE}$ (%) | 3 months                          | 2 months | 1 month |
|  |                             |                          |                           |                                   |          |         |
| <u>All items included in the study</u> | 2.8                         | - 0.7                    | 2.9                       | 1.00                              | 1.37     | 2.25    |
| <u>Selected expenditure items</u>      |                             |                          |                           |                                   |          |         |
| Jewellery                              | 8.1                         | - 6.7                    | 10.5                      | 1.00                              | 1.02     | 1.46    |
| Watches                                | 9.3                         | -15.0                    | 17.7                      | 1.00                              | 0.67     | 0.48    |
| Quilts & Blankets                      | 12.3                        | - 3.2                    | 12.7                      | 1.00                              | 1.25     | 2.25    |
| Bedding & Towels                       | 8.5                         | - 7.4                    | 11.3                      | 1.00                              | 0.72     | 1.51    |
| Dwelling repairs & renovation          | 10.7                        | -10.0                    | 14.6                      | 1.00                              | 0.58     | 0.94    |
| Household help                         | 8.2                         | -10.6                    | 13.4                      | 1.00                              | 0.40     | 0.38    |
| Furniture                              | 6.1                         | + 0.1                    | 6.1                       | 1.00                              | 1.77     | 2.69    |
| Electrical equipment                   | 4.7                         | - 3.2                    | 5.7                       | 1.00                              | 0.96     | 1.96    |
| Durables for Culture & Entertainment   | 7.4                         | - 1.3                    | 7.5                       | 1.00                              | 1.51     | 3.19    |
| Pictures & Frames                      | 14.9                        | -11.3                    | 18.7                      | 1.00                              | 1.19     | 0.96    |
| Dental treatment                       | 6.1                         | -18.6                    | 19.5                      | 1.00                              | 0.20     | 0.23    |
| Private physician's fees               | 9.5                         | + 7.8                    | 12.3                      | 1.00                              | 1.66     | 0.59    |
| Optical rehabilitation aids            | 6.6                         | - 7.3                    | 9.8                       | 1.00                              | 0.58     | 0.98    |
| Vehicle licence                        | 5.2                         | -24.3                    | 24.9                      | 1.00                              | 0.31     | 0.11    |
| Vehicle upkeep & repairs               | 5.7                         | - 0.2                    | 5.7                       | 1.00                              | 1.49     | 2.30    |

For Recreation & Excursions, it was found that the worst recall period is 12-months, while 3, 2 or 1 quarters provide similar MSE's, e.g.  $MSE(9)/MSE(12) = 0.2$ . For Trips abroad the best recall period was found to be 3-months with a ratio of 0.4. These results are not easy to explain, unless the recall bias estimates are biased themselves, as was explained earlier in section 4.2.

Table 4.9  
Relative Errors of 12-Month Period and Comparison of MSE by Length of Reference Period  
Israeli 1986/87 FES

| Expenditure group       | 12 months period            |                          |                           | Relative MSE relative to 12 months |          |          |          |
|-------------------------|-----------------------------|--------------------------|---------------------------|------------------------------------|----------|----------|----------|
|                         | Relative sampling error (%) | Relative recall bias (%) | Relative $\sqrt{MSE}$ (%) | 12 months                          | 9 months | 6 months | 3 months |
|                         |                             |                          |                           |                                    |          |          |          |
| Total                   | 3.70                        | -13.6                    | 14.1                      | 1.00                               | 0.18     | 0.25     | 0.21     |
| Recreation & Excursions | 3.71                        | -16.0                    | 16.4                      | 1.00                               | 0.13     | 0.19     | 0.17     |
| Trips abroad            | 4.84                        | -10.2                    | 11.3                      | 1.00                               | 0.56     | 0.49     | 0.37     |

These results show that no single length of the recall period is best for all the target estimates of the survey. However, further research is required to provide more reliable estimates of recall biases.

## 4.5 Variability Estimation

FES's, like various other surveys, aim to supply estimates of the distribution and of parameters of variability of the population for a given period, for example, measures of inequality of the distribution of households by income. Another example is the distribution of households by the percentage of expenditure on food out of total expenditure giving some indication on the standard of living.

When any measure of variability of the population is calculated on the basis of data collected for a reference period that does not cover the whole target period for each unit, it includes not only the variability between the population units, but also the variability between partial periods of each unit within the given period. Thus, the shorter the reference period and the higher the variability between the partial periods, the wider the variation between the population units being estimated.

In many FES's it was preferable to reduce the reference period (both of the diary and the interview) in order to obtain better estimates of averages and weights for updating the CPI, at the expense of obtaining estimates of the distribution of the population and related measures of expenditures.

This is not the case with income, because it is important to classify households according to their income and to provide estimates of expenditures by level of income.

A preliminary study, on an experimental basis, was carried out in Israel to study the effects of reducing the reference period from 12 months to 3 months, on income estimates. For this study, administrative data of monthly salary payments for a few groups of employees were used.

Estimates were derived by simulation for reference periods of  $T$  months,  $T=1,2,3,6,12$ . As was expected, the estimated variability between employees increased as the length of the reference period was shortened. The ratios of the estimated variability relative to the 12-month period were: 1.02 for 6 months, 1.05 for 3 months, 1.12 for 2 months and 1.20 for 1 month.

This experimental study provided also results on the amount of misclassification of units according to deciles of annual income, for the different lengths of reference period. For each unit the true annual decile was determined by the information over 12 months. Then, this was repeated according to the information of shorter reference periods. It was found that 25% were misclassified by using only 6 months, 40% by using 3 months, 47% by 2 months and 53% by the use of only one month. Most of the misclassifications were within a distance of one decile, but some units were wrongly classified by a distance of two or more deciles: 12% with reference period of a one month, 7% with a two-month period, and only 1% with three months.

Although this study did not represent the entire population nor all income components, the results indicate the importance of this issue.

## CHAPTER 5. SAMPLING

In most countries, the sampling design of FES's is often not that much different from that of other household surveys. However, there may be some specific considerations in designing the sample for FES's, such as the importance of sampling in time, the degree of clustering, or the efficiency of drawing either a self-weighted sample or a sample with different probabilities.

### 5.1 Sampling in Two Dimensions

In FES, as in other surveys in which the investigating variables vary over time, the survey population comprises two dimensions. One dimension, is the set of all households,  $\{N\}$ , who belong to the survey target population. The second dimension is the set  $\{T\}$  of all basic "time units", e.g., days, weeks, etc. that cover the entire survey period, which is usually a whole year. The sample for a yearly FES is selected from both dimensions, unless each selected household has to provide all the required data for the entire year. Almost nowhere is a full-year diary used nowadays. However, an interview that relates to a reference period of a year is not unusual. In some countries it relates to the same year for all the sampled households, as in Canada and Sweden for example, but this is not always the case. In other countries, the interviews refer to "moving" reference periods preceding the diary period, e.g., the last 12 months, or the last 3 months.

In almost all FES's the sample has to be drawn with respect to both dimensions and the first issue is how to allocate the sample between these two dimensions. For a given cost, what has to be considered is whether to draw a sample with a large number of households and to sample a short reference period for each sampled household, or to extend the reference period at the expense of reducing the number of households in the sample. These considerations refer primarily to the length of the diary period and have been discussed in detail in chapter 3.

For a predetermined length of a diary period, the entire survey period is divided into  $T$  sub-periods  $p$  ( $p=1,2,\dots,T$ ), for each household  $i$  ( $i=1,2,\dots,N$ ). The survey population, therefore, comprises the two-dimensional set of  $\{N*T\}$  units, and the sample should be selected from this set, so that the Ultimate Sampling Unit (USU) is a certain household  $i$  in a certain investigation period  $p$ .

Theoretically, the following sampling techniques are possible:

- (1) Simple random sampling from  $\{N*T\}$ , which is the most inefficient sampling method in both dimensions, both with respect to cost and sampling variance.
- (2) Sampling  $t$  periods out of  $T$ , where each period is considered as a cluster of all the  $N$  households and then, within each sampled period, a sample of households is selected independently, by whichever sampling technique that is considered to be the best. With this method, sampling variance will increase the higher is the variance between periods. Also, the field work cannot be carried out currently over the entire survey period and this may cause organizational difficulties.
- (3) Sampling  $n$  households out of  $N$ , where each household is a cluster of  $T$  sub-periods, from which one sub-period is sampled at random. This technique will lead to an inefficient sample in the time-dimension.
- (4) Stratified sampling from  $\{N*T\}$ , where the sub-periods,  $T$ , serve as strata and a sample of  $n_p$  is drawn independently from each period  $p$  by whichever sampling technique that is considered to be the best. The samples  $n_p$  are usually of equal size, but a non-proportional sample is optional, like in the U.S.A., where, for the period preceding the New Year, a larger sample is drawn. This technique will result in smaller sampling variances, but may cause operational difficulties in the field work.

- (5) Sampling  $n$  households out of  $N$  and then allocating the sample between all the  $T$  sub-periods, not necessarily equally. This technique is similar in its features to the technique in (4) mentioned above, although the sample is not drawn independently for each period. However, the sample drawn by this technique is easier to work with in the field.

The two phases of the last-mentioned technique, which is often the preferred one, are discussed in the following.

## 5.2 Sampling Frames and Sampling Designs

As with other surveys, the sampling design depends on the features of the sampling frame, i.e., the kind of units in the frame and the information available for each unit.

Four primary types of frame may be used:

- (a) An area frame, in which each unit is a certain geographical area identified by defined boundaries and all the units cover the entire area of the survey target population. Geographical characteristics of each area (such as region, type of locality, etc.) and usually some other information, for example, that obtained in the previous census (such as the number of housing units or socio-economic details of the area as a whole), are available.
- (b) A Housing units frame, in which each housing unit is listed separately. Each housing unit has its identification details (address and name of owner or renter). Geographical characteristics of each unit may be derived from the identification details, and other characteristics may also be available, such as number of rooms, type of housing unit (an apartment, a house, etc.). Sometimes, further information about the households that occupied the housing unit at the time the frame was compiled (e.g., the previous census), may also be included.
- (c) A household frame, which has features like those of the housing units frame, but each unit is identified by details of the household. One housing unit may contain more than one household, and one household may be associated with more than one housing unit. More characteristics of the households may be available in this kind of frame than in the housing unit frame.
- (d) A frame of persons, where each person (all persons or only those of a certain age) is listed separately with identification details, and usually with some additional basic information (such as sex, age, etc.).

With all kinds of frames, almost always a sample of two or more stages is used, allowing various degrees of geographical clustering. When an area frame is used on its own, first areas are sampled and then a list of all the USU's, i.e., households, has to be prepared by special field screening. The area sample may be drawn in one or in several stages, depending on the size of the areas in each stage and on the degree of the clustering planned for the sample. Then the lists which are prepared for each area can be either used as the sampling frames for the final stage of drawing the sample of the USU's, or even as the lists of the USU's in the sample. An area frame is used on its own only when no better alternative is available, because of the extra cost and errors that are involved in preparing the lists of households within the sampled areas.

Nevertheless, when one of the other kinds of frames are being used, almost always, an area sample is first selected, either in one or more stages. As an example, in the first stage, localities which are the Primary Sampling Units (PSU's) are selected, and in the second stage, a sample of enumeration districts of the previous census is selected within each of the PSU's. Then, for each of the selected areas, the frame is used to draw the sample of the USU's, i.e., housing units, households or persons, depending on the type of the sampling frame.

When the housing unit is the USU, all the households that occupy that unit at the time of the survey are included in the final sample. This may increase the amount of clustering, depending on how common is the use of one housing unit by more than one household. With a household frame, most USU's coincide with the investigated units of the survey.

With a persons frame, since the population units are households, each household is represented in the frame by each of its members. Thus, because the inclusion of a certain household in the sample is by means of the inclusion of one of its members, the sampling probabilities of the households will vary with the number of their representatives in the frame. With such a sampling design, households are sampled with probability proportional to their size (PPS), e.g., the sampling probability of households of 4 persons will be twice that of households with 2 persons and four times that of households of 1 person. If a frame that includes only adult persons (say, of age 18 and over) is used, these probabilities will vary less than if all individuals are included.

However, if households have to be sampled with the same probability from a persons frame, it is necessary to define in advance for each household only one representative in the frame, something that is rarely easy to perform. Alternatively, the representative member can be defined by predetermined rules to be applied in the field and only those households for which the selected individual follows these rules are included in the sample: This design is, of course, very costly and thus not efficient. Therefore, if there is a choice, the use of a persons frame is justified only if sampling of the households with probabilities proportional to their size is proved to be much more efficient than sampling with equal probabilities (see section 5.4).

A problem common to all sampling frames is that they may not be fully updated to the period of the survey (although this is not, in general, the situation with an area frame). The more the frame is out of date, the more costly and the less accurate the survey estimates may be. If an out-of-date frame is used, undercoverage may occur and this may result in biases in the survey estimates.

The housing units frame may be subject to undercoverage due to new construction, however usually special separate frames for drawing a supplementary sample are available. With a household frame, undercoverage may occur due to new households, such as recently married couples, divorcees, single adults leaving their parents' home, etc. and usually there are no suitable frames for drawing supplementary samples for these groups.

On the other hand, the use of out-of-date frames may result in ineligible cases in the sample: with a housing units frame, vacant dwellings, businesses, etc. and with households and persons frames, due to deaths, emigration, etc.

When either a frame of households or of persons is used and the addresses are not up-to-date, the rate of non-response due to non-contacts will increase, because the interviewer may have difficulty to find or to trace the household in its new address. With a housing units frame there are no changes in the location of the units, but if streets have been renamed, difficulties may arise in locating sampled units. Similarly, if the occupants of the sampled unit have changed, problems of location can arise when the identification of the housing unit is by the resident household.

Both non-contacts and ineligible cases in the sample increase the cost of the survey. But, whereas, ineligible cases in the sample have a minor effect only on the sampling variances of most of the survey estimates, undercoverage and non-contacts may be a serious source of bias. Therefore, when it is possible to choose, a frame of housing units is usually preferred to a frame of households, although the latter may provide more suitable data for a better stratification.

### 5.3 Sampling of Households

As with all surveys which are based on visiting the investigated units, i.e., excluding telephone or mail surveys, the extent of the geographical clustering of the sample has to be considered. For a given sample size, a balance has to be found between the negative effect of clustering on the sampling variances (when a positive high intra-cluster correlation exists), and the positive effect of decreasing the survey cost. The sampling design must take into account the size of the interviewers assignment for one period, which depends on the field organization and on the survey design, e.g., the length of the diary period, the number of the interviewer's visits to the household during the recording period, the average time required for the interview, etc.

One-stage sampling of households with no clustering in the population dimension may be too costly, even when clustering in the time dimension is introduced by grouping the sample into geographical clusters for the interviewers' assignments. Thus, the sample of two or more stages is often used with clustering in the household dimension as well. Nevertheless, as far as possible, clustering should be avoided, since high positive intra-cluster correlation within small areas is often found for the kinds of variables under investigation in FES's.

As an illustration of a typical sampling design for the FES, let us examine, for the sake of simplicity, two-stage sampling with uniform final sampling probabilities.

Let us denote by  $N_j$  the number of households in PSU j ( $j=1,2,\dots,M$ ), where  $N=\sum N_j$  is the number of households in the entire survey population, and let  $n$  be the size of the entire sample,  $f=n/N$  the final sampling probability, and  $\bar{m}$  the size of each interviewer's assignment per period.

First, PSU's with  $N_j > \bar{m}/f$  are included in the sample with certainty. Then, within each of these PSU's, a sample of the USU's is drawn independently with probability  $f$ . Thus, with respect to that part of the population,  $N(1)$ , included in this group of PSU's,  $M(1)$ , the sample is actually a one-stage stratified sample of size  $n(1)=fN(1)$ , each of the large PSU's being a separate stratum.

A two-stage sample is then drawn from all the other  $M(2)$  PSU's comprising a population of size  $N(2)=N-N(1)$ .

In the first stage, PSU's are selected with probability proportional to their size (PPS), i.e.,  $f_{j1} = N_j m(2)/N(2)$ , where  $m(2)=[n-n(1)]/\bar{m}$  is the number of sampled PSU's from this group.

In the second stage, a sample of USU's is selected within each of the sampled PSU's with the probability  $f_{j2}=f/f_{j1}=fN(2)/m(2)N_j$ , so that the final probability is equal for all the USU's.

The sample of the USU's in each PSU will be at least of size  $\bar{m}$ :  $n_j=N_j f_2 \bar{m}$  for PSU j selected with certainty, and  $n_j=N_j f_{j2} = \bar{m}$  for each of the other PSU's. This will be the case if  $N_j$  are accurate, otherwise  $n_j$  will vary according to the ratio of the real size of the PSU and its size in the frame and the actual size of the total sample may be somewhat different from the planned one.

By such a sampling technique, the size of the sample within the PSU's that were included with certainty, if either too large for one interviewer's assignment and too small to be split between two interviewers, may be inconvenient for the field organization. In order to solve this problem, the sample design can be modified by grouping such PSU's so that their accumulated population size will give a sample size of approximately a multiple of  $\bar{m}$ . Then, the number of clusters in the sample in each of the pooled PSU's is determined by PPS sampling and the sampling probability within each of the PSU's is adjusted to correspond to a final probability  $f$ . For example, when one PSU is of size  $N_1=1.4\bar{m}/f$  and a second PSU is of size  $N_2=1.6\bar{m}/f$ , the two PSU's are pooled together so that  $(N_1+N_2)f=3\bar{m}$ . Then, a PPS sample of size 3 is selected, so that from one of these PSU's 2 clusters will be selected and only 1 cluster from the other PSU.

Stratification in the sampling of the PSU's, as well as of the USU's is usually applied according to the available information in the sampling frame.

Following are a few typical examples of sampling designs for different FES's:

- (a) In Israel a two-stage sample design is used with uniform final probabilities, but most of the sample (about 80% of the population) is a one-stage stratified sample with a fixed cluster size modification. Each of the interviewer's assignments is planned to include about 12 eligible households (of which about 2-3 are expected to be non-respondents). Each locality of a size that corresponds to at least one interviewer assignment is included in the sample with certainty. Small localities are selected by PPS with stratification according to type, size, region, and socio-economic index. For each selected locality, the municipal tax file is arranged by address and is used as the sampling frame of dwelling units. Although these files are continuously updated, a supplementary sample is subsequently drawn to cover new construction since the time of the selection of the primary sample and throughout the survey period. Additional samples are selected from special frames of semi-institutional populations, such as student dormitories, new immigrant absorption centers, etc., that are not covered in the main frame. The sample size is about 7,500 dwelling units.
- (b) In the United Kingdom (Elliot, 1991) a two-stage sample design with uniform final probabilities is used: at the first stage, areas are sampled (each area containing roughly 2,500 households) and at the second stage, addresses are sampled within the sampled areas. The sample of the PSU's is stratified by region, degree of urbanisation and indicators of socio-economic status derived from the most recent census of population. The addresses are selected from the Postcode Address File, i.e., a list of postal delivery points throughout the country. Each of the annual samples includes about 10,000 households.
- (c) In Hungary (Elteto, 1991) the sample of the FES is a sub-sample of a master sample consisting of census enumeration districts. The sample is drawn in three stages. In the first stage localities are sampled: cities with 25,000 or more inhabitants are included in the sample with certainty, and the other localities are selected with probability proportional to the number of dwellings, after stratification by size and occupational structure. The population in small villages have a greater probability to be selected than in cities. In the second stage, enumeration districts are selected by PPS according to size, after stratification by occupational structure. In the third stage, households are selected within each of the sampled districts from a sampling frame stratified by size and social group. The sampling frame is updated in the sampled districts before selection of the sample by correcting the list of dwellings for demolitions and new construction. The FES sample covers more than 1,300 census districts in nearly 400 localities. In each enumeration district 9 households are selected, thus making the total sample size almost 12,000.
- (d) In Italy (Russo and Coccia, 1991) two-stage sampling is used with households as the USU's instead of housing units. In the first stage, localities are used as the PSU's: all localities with more than 50,000 inhabitants are selected with certainty, and smaller localities are selected with probability proportional to size, after stratification by size, zone and prevalent economic activity. In the second stage, households are selected systematically from lists of registered households within the PSU's. The use of a household frame leads to a very high rate of non-contact and thus a process of substitution is used. Every month 3,200 families are included in the sample, 38,500 in a year.

## 5.4 Non-Proportional Sampling of Households

In general, when the primary purpose of a survey is to obtain one estimate for the entire population, or various estimates that are highly correlated, a non-proportional sample may be more efficient than a proportional one. In that case, for a fixed sample size,  $n$ , optimal selection probabilities,  $f_h(\text{opt})$ , within strata  $h$  can be calculated according to the Neyman allocation:

$$f_h(\text{opt}) = \frac{n_h(\text{opt})}{N_h} = n \frac{s_h/\sqrt{c_h}}{\sum N_h s_h/\sqrt{c_h}}$$

where, for each stratum  $h$ ,  $N_h$ ,  $s_h$ , and  $c_h$  are the size of the population, the standard deviation and the cost per unit, respectively, and  $n_h(\text{opt})$  are the optimal sizes of the sample in strata  $h$ . This means that optimal selection probabilities will be higher, the higher is the variability within the stratum and the cheaper is the investigation per unit of that stratum. If the cost per unit does not differ much between strata, and the  $s_h$  are of similar magnitude, a proportional sample is preferable.

FES's are multi-purpose and usually there is no one optimal set of  $f_h$  for all the different estimates that have to be derived from the survey data. Therefore, most of the samples for FES's are selected with equal probabilities. However, there may be other considerations for drawing a non-proportional sample for FES's.

Sometimes, there is a special need to supply separate estimates for certain sub-populations which are rather small and thus, the samples for these groups are selected with higher probabilities to ensure samples of a sufficient size. For example, rural areas are sampled with higher probabilities in several countries, because special estimates are required for this sub-population. Another situation is when for certain groups of the population much higher non-response rates are expected, e.g., households of size 1. Then, in order to obtain a proportional sample of the respondents, a higher selection probability may be considered for the groups with poor response rates. This can be done, in practice, only if the sampling frame includes the necessary information for such stratification.

Another form of non-proportional sampling (which is in use in Sweden, for example) is to select the entire sample of households with PPS according to the household's size. This is possible, either when a sampling frame of households is available with updated information on each household's size, or when a frame of individuals is used. With the latter, a sample of individuals is selected, and the households to whom these individuals belong are included in the sample (as was explained above in section 5.2).

Let us examine the efficiency of this sampling method.

Let  $Z_i$  be the size of household  $i$ , which may be defined by all its members, or only by its adult members, say, of age 18 years and over. Then, household  $i$  is selected with probability  $\pi_i = Z_i/Z$ , where  $Z$  is the total number of individuals in the population.

Let  $X_i$  be the value of the investigated variable(say, total expenditure) for household  $i$  of the sample ( $i=1,2,\dots,n$ ) and let  $\bar{X} = \sum X_i/N$  be the average expenditure per household in the population.

Let  $A_i = X_i/Z_i$  be the average per person in household  $i$ , and  $\bar{A} = \bar{X}/Z$  the average per person in the entire population, where,  $Z = Z/N$  is the average household size.

The simple estimate of the average per household is  $\bar{X}_{\text{ep}} = \sum X_i/n$  with equal probabilities, and

$\bar{X}_{\text{pps}} = (1/n)(1/N)\sum X_i/\pi_i = \bar{Z}(1/n)\sum A_i = \bar{Z} \bar{A}_{\text{ep}}$  with PPS, providing  $N$  is known.

The relative sampling variance (assuming for simplicity, sampling with replacement) is

$$R.Var(\bar{X}_{ep}) = \frac{\sum_{i=1}^n (\bar{X}_i - \bar{X})^2}{nN^2} = \frac{C_x^2}{n} \quad \text{with equal probabilities, and}$$

$$R.Var(\bar{X}_{pps}) = \frac{\sum_{i=1}^n (\bar{x}_i / \pi_i - \bar{N})^2}{nN^2} = \frac{\sum_{i=1}^n (A_i - \bar{A})^2}{n\bar{A}^2} = \frac{C_A^2}{n} \quad \text{with PPS.}$$

Thus, PPS sampling will be better if  $C_A^2 < C_x^2$ , i.e., the relative variation between the households' expenditure is greater than that between per person expenditure. This may be valid for some of the variables under investigation in FES's, but not for all of them, and not even for all of the most important ones. One may expect that  $C_A^2 < C_x^2$  for expenditure on Food, where the amount of expenditure depends on how many persons have to be fed. But, this may not be the case with Health expenditure, which is likely to be higher in households of one or two persons, many of whom are elderly people. Household expenditure on Entertainment may decrease with the size of the family and this may also be the case with Trips abroad. The relationship between  $C_A^2$  and  $C_x^2$  depends, of course, on the definition of the size of the household that is chosen for determining the selection probabilities - all the household members or only the adult ones.

The efficiency of a PPS sample of households was examined in Israel by studying the 1986/87 FES data. The sampling variances of the survey estimates for the entire population were estimated first under the actual sampling design, i.e., with equal selection probabilities. Second, sampling variances for the survey estimates for each household size  $h$  ( $h=1,2,\dots,6+$ ), were calculated. These variances were used to estimate the sampling variance for estimates of the entire population, if a stratified sample by size of household would have been used with selection probabilities that correspond to a PPS sample of households. This could be used as a good approximation, since post-stratification by size of household was used for weighting and most of the effects of the weighting procedure and of the non-response rates (which will be discussed in later chapters), were taken into account in the calculation of the sampling variances.

Let us denote by  $\sigma_h^2(1)$  and  $\sigma_h^2(2)$  the sampling variances of stratum  $h$ , i.e., households of size  $h$ , with a proportional stratified sample and with a non-proportional stratified sample, respectively. It should be noted that for households of size 7 and over, the selection probabilities were chosen to be the same as for households of size 6 in order to avoid too much differentiability. Such a sampling design can be implemented in practice by including only 6/7 of the households of size 7, 6/8 of size 8, etc.

Let  $n_h(1)$  and  $n_h(2)$  be the size of the net sample with a proportional sample and with a non-proportional sample, respectively, for a given size of the overall net sample  $n$ , i.e.,  $n = \sum n_h(1) = \sum n_h(2)$ .

Let  $N_h$  be the number of households in stratum  $h$  in the population, and  $w_h = N_h/N$ .

It can be shown that  $n_h(2) = nn_h(1)h/\sum n_h(1)h$ , where  $h=6$  for  $6+$ .

Table 5.1 presents the distribution of the net sample in the Israeli 1986/87 FES by size of household, from the proportional sampling actually used and the distribution that would have been obtained if PPS sampling would have been used. The ratios  $B_h = n_h(2)/n_h(1) = \sigma_h^2(1)/\sigma_h^2(2)$ , i.e., the ratio between the sampling variance of the proportional sampling and of the PPS sampling for sub-groups by size of household, are also presented in the Table 5.1.

Table 5.1  
Population Weights and Net Sample Size by Household Size and Sampling Design  
Israeli 1986/87 FES

| Population weight<br>&<br>Net sample size                           | Size of household h |      |       |      |       |       |       |
|---|---------------------|------|-------|------|-------|-------|-------|
|   | Total               | 1    | 2     | 3    | 4     | 5     | 6+    |
| Weight  | 100.0               | 17.5 | 22.3  | 13.6 | 17.2  | 14.4  | 15.0  |
| $n_h(1)$ : Proportional   | 5,000               | 659  | 1,081 | 698  | 963   | 764   | 835   |
| $n_h(2)$ : PPS  | 5,000               | 188  | 614   | 591  | 1,096 | 1,086 | 1,425 |
| $B_h = \frac{n_h(2)}{n_h(1)} = \frac{\sigma^2_h(1)}{\sigma^2_h(2)}$ | 1.00                | 0.28 | 0.57  | 0.85 | 1.14  | 1.42  | 1.71  |

For estimates for sub-groups of the population by size of household (one of the primary aims of FES's), it is obvious that sampling variances of households of size 1, 2, or 3 will increase substantially with a PPS design ( $B_1=0.28$ ,  $B_2=0.57$  and  $B_3=0.85$ ), whereas for larger households of size 4, 5, and 6+, the variances will be reduced ( $B_4=1.14$ ,  $B_5=1.42$  and  $B_{6+}=1.71$ ).

For the total population, the ratio of the sampling variances will thus be

$$B = \frac{\sigma^2(1)}{\sigma^2(2)} = \frac{\sum W_h^2 \sigma^2_h(1)}{\sum [W_h^2 \sigma^2_h(1)/B_h]} \quad \text{and the PPS design would be more efficient when } B > 1.$$

The efficiency of PPS sampling by size of household (values of B) and for PPS sampling by number of household members of age 18 and over (values of B\*), were calculated for expenditure and income groups. Some of these results are presented in Table 5.2.

The results show very clearly that for the majority of the estimates, sampling with equal probabilities is more efficient than both kinds of PPS sampling. Furthermore, for almost all the estimates, for the entire population, PPS by the size of the household is less efficient than PPS by the number of adults.

For Total Expenditure  $B=0.73$  and  $B*=0.96$ , for Total Income  $B=0.81$  and  $B*=0.93$ . Even with Food  $B=0.84$  and  $B*=0.81$ , which may be explained by the consumption of cheaper food by large households and more expensive food by small households. As was expected, B is comparatively small for Health ( $B=0.64$ ), whereas  $B^*$  is much higher ( $B^*=0.98$ ) which is partly explained by the system of health insurance payments in Israel.

A few examples for sub-groups of expenditure are presented here. For Bread, B is close to 1 and also for Meat & Poultry, for Dairy Products & Eggs and for Water. With most of the other sub-groups  $B < 1$  and also  $B^* < 1$ . For example, for Dental treatment  $B=0.66$ , for Trips abroad  $B=0.66$ , for Recreation & Excursions  $B=0.65$  and for Meals away from home  $B=0.56$ . An exception is with Children's outerwear, for which  $B=1.25$ .

Also, for estimates of proportions, e.g., households' possessing durables, the PPS sampling leads, obviously, to higher sampling variances. Thus, for most of the estimates of the total population, as well as of sub-groups, it was found without any doubt that PPS sampling is far from optimal.

**Table 5.2**  
**Ratio of Sampling Variances of Proportional and of PPS Sampling**  
**Total Population - Israeli 1986/87 FES**

| Income / Expenditure group           | Ratio of Sampling Variances of proportional and PPS by number of household members |                   |
|--------------------------------------|--|-------------------|
|                                      | All members (B)  | Adults (18+) (B*) |
| <b>TOTAL INCOME</b>                  | 0.81   | 0.93              |
| Net income                           | 0.74   | 0.96              |
| Income from work                     | 0.80   | 0.89              |
| Income from capital                  | 0.61   | 0.97              |
| Income from pensions                 | 0.51   | 0.72              |
| Income from assistance               | 0.71   | 0.78              |
| <b>TOTAL EXPENDITURE</b>             | 0.73   | 0.96              |
| <b>MAJOR EXPENDITURE GROUPS:</b>     |  |                   |
| Food (excluding Vegetables & Fruit)  | 0.84   | 0.81              |
| Vegetables & Fruit                   | 0.83   | 0.93              |
| Housing                              | 0.67   | 0.71              |
| Household maintenance                | 0.71   | 0.88              |
| Furniture & Household equipment      | 0.77   | 0.88              |
| Clothing & Footwear                  | 0.87   | 1.00              |
| Health                               | 0.64   | 0.98              |
| Education, Culture & Entertainment   | 0.75   | 0.89              |
| Transport & Communication            | 0.73   | 0.84              |
| <b>SUB EXPENDITURE GROUPS:</b>       |  |                   |
| Bread, Cereals & Pastry products     | 1.01   | 0.93              |
| Meat & Poultry                       | 0.91   | 1.06              |
| Dairy products & Eggs                | 0.93   | 1.02              |
| Meals away from home                 | 0.56   | 0.71              |
| Water                                | 0.97   | 0.88              |
| Children's outerwear                 | 1.25   | 0.99              |
| Footwear                             | 0.84   | 0.93              |
| Health insurance                     | 0.65   | 0.91              |
| Dental treatment                     | 0.66   | 0.93              |
| Entertainment, Culture & Sport       | 0.76   | 0.87              |
| Recreation & Excursions              | 0.65   | 0.81              |
| Durables for Culture & Entertainment | 0.59   | 1.01              |
| Public transport                     | 0.80   | 0.88              |
| Trips abroad                         | 0.66   | 0.71              |
| Post & Telephone                     | 0.68   | 0.70              |

## 5.5 Allocating the Sample Over Time

First, let us look at the model of the sample allocation between the survey periods. What distinguishes various allocation models is the degree of overlap between periods, which may have different effects on the survey estimates due to possible biases from the diary recording (see section 3.2).

For a two-week diary, three possible models are as follows:

- 1) The non-overlapping model - The sample is divided into 26 sub-samples, each of them being allocated to a different two-week period. All households of the same sub-sample start their diary keeping on the same day.
- 2) The weekly overlapping model - The sample is divided into 52 sub-samples and each week, on the same day, another sub-sample starts. Thus, each week (except for the first and the last), two sub-samples are in the survey, one in its first recording week and the other in its second recording week.
- 3) The daily overlapping model - The sample is divided into 365 sub-samples and each day, a new sub-sample starts recording. Thus, every day, except for the beginning and end of the survey period, 14 sub-samples participate in the survey, each on a different serial day of recording.

There may, of course, be other allocation models with different overlapping periods. None of these models can be implemented in practice without some distortions, especially the daily model.

Let us examine the effects of these models on the survey estimates on the assumption that each sub-sample is a random sample and that either there is no non-response, or, if there is, then it is of a random nature.

Let us denote by  $d$  ( $d=1,2,\dots,7$ ) and by  $d+7$  ( $d+7=8,9,\dots,14$ ) the serial day of recording in the first week and the second week of the diary, respectively. So, if for example, the starting day of the diary is Tuesday,  $d=1,8$  refer to all Tuesdays,  $d=6,13$  refer to all Sundays, etc.

For any two-week period, let  $\bar{X}_d$  and  $\bar{X}_{d+7}$  denote the true expenditure per household on day  $d$  and day  $d+7$  of the diary recording.

Let  $\bar{X}_d^*$  and  $\bar{X}_{d+7}^*$  denote the respective reported values. Then the true and the reported average expenditures for the entire survey period are:

$$\bar{X} = \sum_d (\bar{X}_d + \bar{X}_{d+7}) \text{ and } \bar{X}^* = \sum_d (\bar{X}_d^* + \bar{X}_{d+7}^*)$$

Then, the relative recording biases of  $\bar{X}_d^*$ ,  $\bar{X}_{d+7}^*$  and  $\bar{X}^*$ , are:

$$\beta_d = (\bar{X}_d^* / \bar{X}_d) - 1, \quad \beta_{d+7} = (\bar{X}_{d+7}^* / \bar{X}_{d+7}) - 1 \text{ and } \beta = (\bar{X}^* / \bar{X}) - 1.$$

It can be shown that the relative bias  $\beta$  is not the same for the three allocation models above:

With the non-overlapping model  $\beta(1) = \sum_d (U_d \beta_d + U_{d+7} \beta_{d+7}) = \sum_d W_d \bar{\beta}_d^w$ ,

$$\text{with the weekly model} \quad \beta(2) = \sum_d \frac{\beta_d + \beta_{d+7}}{2} = \sum_d W_d \bar{\beta}_d$$

$$\text{and with the daily model} \quad \beta(3) = \frac{1}{7} \sum_d \frac{\beta_d + \beta_{d+7}}{2} = \frac{1}{7} \sum_d \bar{\beta}_d$$

where,  $W_d = (\bar{X}_d + \bar{X}_{d+7}) / \bar{X}$  are the weights of the weekdays, and

$$U_d = \bar{X}_d / (\bar{X}_d + \bar{X}_{d+7}) \quad \text{and} \quad U_{d+7} = \bar{X}_{d+7} / (\bar{X}_d + \bar{X}_{d+7})$$

are the weights of the first and the second weeks for each of the weekdays.

There would not be any difference between the biases of the three models if for all days  $W_d = 1/7$  and also  $U_d = U_{d+7} = 1/2$ . However, as was demonstrated in Section 3.2, the weekly expenditure is not expected to be evenly distributed between the different days of the week. Thus, in both the non-overlapping model and the weekly model the biases will usually be different than in the daily model. The biases of the non-overlapping and the weekly models will not be the same if  $U_d \neq U_{d+7}$ , e.g., if the households spending pattern is affected by the timing of salary payments where they are based on a two-week cycle. Except for the daily model, the bias depends on the interaction between the daily recording biases and the daily weights of the expenditure.

An empirical study could indicate which is the least-biased model. However, evaluation of  $B_d$  is not so simple, although there is evidence that bias in the diary recording varies with the day of recording. If the  $B_d$  are higher the higher are the  $W_d$ , the daily model is preferable. The weekly model is better than the non-overlapping model if  $B_d$  are higher in that week with higher  $U_d$ . Although there is no evidence of such interactions, they probably do exist and thus the daily model seems to be the best one. However, this model is difficult, if not impossible to apply in practice.

An approximation of the daily model can be achieved with a weekly model, if there is no fixed starting day for all households, and interviewers are instructed to place the diaries, as evenly as possible, over the weekdays. This is the model that is being used in the Israeli 1992/93 FES. It should be noted that in the previous FES, a non-overlapping model was used, and it was found that the interviewers placed most of the diaries in the first week of the two-week assignment, thus every second week there were very few starting days. The use of a daily overlapping model, even if not fully implemented in practice, will probably provide data for a study that may lead to a better understanding of the recording biases.

For the sample allocation over time, first the number of periods is determined by the allocation model. For any number of periods,  $T$ , the final clusters of the sample can be assigned into  $T$  sub-samples at random, or by some stratification, and each sub-sample is designated at random to one of the  $T$  periods. Such a technique may be very difficult to apply because it does not take into account the organization of the field work over the year. Therefore, a more sophisticated way of allocating the sample between the periods is required, under the field organization constraints. Although this may lead to some deviation from a pure randomized allocation, it is preferable because, anyway, many deviations from the desired allocation are likely to occur when field constraints are not taken into account.

Prior to the allocation of the sample between the survey periods, all households in the sample are either grouped straightaway into the final clusters, each comprising one interviewer's assignment for one period, or, have to be further grouped within those clusters that contain more than one assignment. Sometimes, in order to decrease the degree of clustering in time within the very large PSU's, the sample is first systematically divided into two or more sub-samples and the grouping into the assignments is made within each sub-sample separately.

As an example, let us look at the way the allocation was made in the Israeli 1992/93 FES, where a two-week diary is being used. In each one-period assignment, the interviewer works for four weeks, i.e., placing the diary over the first two weeks, then visiting the families during their record keeping, and conducting the final interviews. Thus, each interviewer may have as many as 13 assignments, although he may have less, depending on when he starts and completes his task.

## CHAPTER 6. NON-RESPONSE

### 6.1 Effects on the Survey Estimates

In each survey, the sample can be divided into five groups:

- (1) Respondents - Units who belong to the survey population and for whom all the required information for all the items under investigation was obtained and was found to be suitable to be included in the survey processing, i.e., did not fail the editing process.
- (2) Respondents with item non-response - Units who belong to the survey population and for whom only a limited number of items of the required data is missing. For these cases the missing data are either imputed, or are left as "unknowns" and thus they are included in the survey estimates.
- (3) Non-respondents - Units who belong to the survey population but for whom no information was obtained (because of refusal, dropping-out during the investigation period, inability to co-operate because of illness or other limitations, absence, etc.), or for whom the collected data were not sufficient for use, or were of poor quality and failed to pass the editing procedures.
- (4) Ineligible cases - Units outside the scope of the survey population that were sampled from groups that could not be excluded from the sampling frame, such as vacant dwelling units. Ineligible units in the sample have to be omitted and should not be regarded in any way as non-response.
- (5) Non-contacts - Units for whom no information whatsoever was obtained because of the failure of the interviewer to locate them, or even because no attempt was made to contact them (e.g., no easy access, overloaded interviewer's assignment, etc.). Some of the non-contacts could be ineligible units and thus, this group as a whole cannot be regarded as non-response. Absentees are usually regarded as non-response, although they can be also considered as non-contacts.

The distribution of the sample by these five groups depends on the sampling frame, on the investigation procedures, on the qualifications of the field staff, as well as on the kind of data and on the population under investigation.

Non-response reduces the net size of the sample and thus increases the sampling variance. But, often the initial sample size,  $n^*$ , is planned in advance to take into account an expected response rate,  $\phi$ , so that  $n^*=n/\phi$ , where  $n$  is the target size of the net sample. However, this increases the field costs because the non-respondents,  $n^*-n$ , have to be contacted and efforts are made to get them to participate in the survey. Furthermore, the non-response rates may not be distributed evenly between the interviewers' assignments and the efficiency of the field work declines and thus becomes even more costly. The alternative approach of substituting for non-response in order to obtain a fixed sample size is more complicated and thus probably more expensive.

However, the more serious problem of non-response lies in the possible biases that it may introduce into the survey estimates and substituting for non-response, usually, does not solve this problem.

If  $\mu_1$  and  $\mu_2$  are the means of the investigated variable of the respondents and of the non-respondents, respectively, and  $\phi$  and  $1-\phi$  are the proportions of these groups, the mean of the entire population is  $\mu = \phi\mu_1 + (1-\phi)\mu_2$ .

Thus, the bias of the survey estimate  $\hat{\mu}$ , which is based only on the respondents, will be

$$\text{Bias}(\hat{\mu}) = E(\hat{\mu}) - \mu = \mu_1 - [\phi\mu_1 + (1-\phi)\mu_2] = (1-\phi)(\mu_1 - \mu_2).$$

The amount of bias is a product of the proportion of the non-response rate and the difference between the means of the respondents and the non-respondents. The survey provides estimates of  $\mu_1$  and of  $\phi$ , but provides no information on  $\mu_2$ .

Many studies in various surveys have shown that non-respondents often differ from respondents, i.e.,  $\mu_2$  differs from  $\mu_1$ . It may happen that  $\mu_2$  is similar to  $\mu_1$  for the entire population, but this may not be the case for different sub-groups of the population, e.g., for one sub-group  $\mu_2 > \mu$  and for the other group  $\mu_2 < \mu$  so that the biases for the entire population may cancel each other out.

It should be mentioned that the presence of ineligible units in the sample leads also to an increase in the survey cost (because, usually, the fact that they do not belong to the survey population is discovered only after they are contacted in the field) and may increase the variance of the survey estimates. However, ineligible units do not cause bias in the survey estimates, providing substitution for them (in order to keep the sample size fixed) is avoided.

## 6.2 Non-response Rates

Because of the heavy demands which FES's impose on the respondents and because of the type of the investigated variables, the non-response rates in these surveys are almost everywhere much higher than in other household surveys. In FES's, non-response due to refusals and drop-outs comprise the majority of the total non-response, whereas in other household surveys they play a much smaller role. In FES's, some households are not capable of recording in the diary or even of providing the interviewer with the required data.

Non-response rates due to absentees and non-contacts are usually smaller in FES's than in other household surveys. This is because interviewers usually make repeated attempts (much more than in other surveys) to contact temporary absentees, if they have returned home within the assigned period for investigation, or if delays are permitted. Also, much more effort is invested in FES's to reduce the number of non-contacts.

Non-response rates are not the same in different FES's, depending on the survey design, on the general reluctance of the population to reveal details of personal and household finances, on the quality of the field work and on other factors. A comparison of non-response rates as reported in different countries may be somewhat misleading, because of differences in the definition of non-response (e.g., if non-contacts are included or not), and due to alternative ways of calculating non-response rates (e.g., out of all the sampled units or only out of eligible cases). Furthermore, in FES's where more imputation is permitted for units giving poor information, less non-response are left. In some FES's, units who provide data for either the diary or the questionnaire are regarded as respondents, whereas in other surveys respondents are only those who completed both. Nevertheless, it is interesting to point out that many of the developed countries report similar non-response rates, as in the following few examples.

In the United Kingdom, the non-response rates vary little from year to year and they are in the range of 26% to 32%, the majority of which is due to initial refusals and drop-outs during the two-week diary period, or immediately after the initial interview (Elliot, 1991).

In Australia, where a personal two-week diary is used with additional information collected by interviewing each adult member of the household and the household as a whole, the non-response rate in the 1988/89 survey was 21%. But, for a further 15.6% only some information had been obtained and the missing data were imputed (Harrison, 1991).

In Hungary, in the 1989 survey, with a two-month diary and a supplementary interview, the non-response rate was about 25%, of which 70% originated from refusals (Elteto, 1991).

In France, the non-response rate of the 1978/79 FES, where the data were collected by a personal 10-day diary and by three interviews, was about 25%. In a pilot survey that was conducted in 1977, for a 7-day diary the non-response rate was 20%, for a 10-day diary 25%, for a 14-day diary 30%, and slightly over 50% for a one or two-month diary. In this pilot survey with a monthly diary the drop-out rate after starting to keep the diary was very low and most of the non-response was from households who immediately refused to co-operate, or refused when the interviewer asked them to keep diaries (Glaude, 1982).

In the United States, where two separate surveys are conducted, the non-response rate is about 15% on the average for the four consecutive quarters of the Interview Survey. In the Diary Survey, about 18% do not complete even one of the two one-week diaries and about 9% complete only one (Tucker, 1992).

There are, of course, exceptions, for example, in the Netherlands an extremely high non-response rate of about 80% (!) is reported (de Heer, 1991).

In Israel, non-response rates steadily increased until 1979/80. In the 1986/87 FES, the diary period was reduced from a full month to two weeks and the reference period of the recall interview was also reduced from 12 months to 3 months for most of the items. These changes in the survey design together with some other improvements, resulted in a substantial reduction of the non-response rate. The non-response rates of the four recent FES's conducted in Israel since 1968/69 are presented in Table 6.1, by primary group of non-response.

Table 6.1  
Response and Non-response Rates by Main Reasons  
1968/69-1986/87 Israeli FES's

| Non-response Group               | 1968/69     | 1975/76     | 1979/80     | 1986/87     |
|----------------------------------|-------------|-------------|-------------|-------------|
| <u>Total eligible households</u> | 100.0       | 100.0       | 100.0       | 100.0       |
| <u>Total non-respondents:</u>    | <u>28.2</u> | <u>34.7</u> | <u>41.0</u> | <u>21.5</u> |
| Refusals                         | 19.4        | 15.9        | 18.4        | 13.6        |
| Drop-outs                        | 3.5         | 11.8        | 16.4        | 1.5         |
| Not capable                      | 4.7         | 5.5         | 4.8         | 3.6         |
| Not at home                      | 0.6         | 1.5         | 1.4         | 2.8         |
| <u>Total respondents:</u>        | <u>71.8</u> | <u>65.3</u> | <u>59.0</u> | <u>78.5</u> |
| Unsuitable for processing        | 2.0         | 3.0         | 3.2         | 2.4         |
| Suitable for processing          | <u>69.8</u> | <u>62.3</u> | <u>55.8</u> | <u>76.1</u> |

In the three earlier surveys, the overall rate of non-response increased from 28.2% in the 1968/69 survey, to 34.7% in the 1975/76 survey, and was very high, 41%, in the 1979/80 survey. Although all three surveys were of a similar design, the main increase was due to drop-outs 3.5%, 11.8% and 16.4%, respectively. The drop-outs were either when the respondent failed to complete the full-month diary (unless there were only a few missing days), or even when the diary was fully completed, but the final interview failed. The percentage of cases whose data were found to be unsuitable for processing also slightly increased.

The increasing rate of non-response over these three surveys could be explained by the increasing rate of price inflation. The annual rate of price inflation was less than 3% in 1968/69, about 30% in 1975/76 and about 140%(!) in 1979/80. There is no doubt that the higher the price inflation is, the more difficult is the diary recording and providing data on expenditure and income by recall for the interview. Furthermore, when there is a very high rate of inflation, the public may be more reluctant to provide information on its economic activities to a department of a government which is, in their opinion, responsible for the economic situation.

In 1986/87, the non-response rate was reduced to 21.5%, and this was mostly due to a considerable reduction in drop-outs (only 1.5%). Cases of non-capable and of not-suitable for processing were also less, each by about 1%. This was probably largely due to shortening both the recording period and the recall period, but this was not the only reason. The number of interviewer's visits to the households during the diary period was increased, so that the interviewer visited the household five times on average, besides the initial and the final interview. The number of visits to each family was determined in the field according to the help that the family needed in recording. In some of the households even a daily visit was made. Also, for the 1986/87 survey a different strategy of incentives was used and found to be very effective (see section 6.4). In addition, this survey was carried out under a moderate level of yearly price inflation of about 20%.

### 6.3 Characteristics of Non-response

As was mentioned above, the bias due to non-response depends not only on the non-response rate but also on the difference between the means of the respondents and the non-respondents. Thus the problem of the high rate of non-response in FES's is of considerable importance when non-response is not evenly distributed between sub-groups of the population that differ with respect to the variables under investigation.

While, there is no problem in determining the overall non-response rate from the survey itself, there is no way to obtain unbiased estimates of the survey variables for the non-respondents. However in order to reduce non-response bias as far as possible, e.g., by weighting procedures, attempts are made to characterize the non-respondents by characteristics that are assumed or found to be correlated with the survey variables. This is not an easy task, because for most of the non-response the survey itself does not provide such information. Usually, only a very limited amount of information from the sampling frame is available for each unit of the original sample. Attempts that have been made to collect basic data in FES's (and in most other surveys) for the non-response have not usually been too successful.

A study of the different non-response rates between different sub-groups can be conducted when the FES is carried out at the same time (or close to) a population census, or when a population register with detailed data for each unit is available. Such studies were performed in the United Kingdom for the 1971 and the 1981 FES's, by matching the survey sample (respondents as well as non-respondents) with the census of the same year. About 95% of the eligible households were matched in both surveys. The censuses provided many characteristics of each matched unit and two sets of distributions by each census variable were obtained, one for the respondents and the second for the non-respondents.

Univariate comparisons were made (Kemsley, 1975 and Redpath, 1986) and also a multivariate analysis, based on the same data, was carried out by examining the joint effect of the households' characteristics on response (Elliot, 1991). Some of the results of these studies for the 1981 survey (based on a sample of about 2,500 households) are presented in Table 6.2.

The impression from these studies was that non-response rates decline with the size of the household, coinciding with the rate that decreases with the number of dependent children.

On the other hand, childless households, particularly those containing older people seemed less likely to co-operate. Substantial differences were associated with the age of the head of household. A linear regression of response rates and age groups (for the 1971 survey) attained a regression coefficient of -2.22 with a correlation coefficient of 0.92. This is not unacceptable, since often older people are either reluctant to undertake the heavy task of diary-keeping, or are unable to understand the task, or are afraid of making mistakes. High non-response rates were also found among households comprising adults who live relatively independently of each other who may not want to commit themselves to co-operate in budgeting or participating jointly in the survey. The lower response of households of self-employed vis-a-vis households of employees, may be due to a greater reluctance to reveal details of income and expenditures to a government survey.

Table 6.2  
Non-response Rates by Household Characteristics  
United Kingdom 1981 FES

| Household characteristics                  | Non-response rate |                   |
|--|-------------------|-------------------|
|  | Percent           | Relative to total |
| TOTAL                                      | 26                | 1.00              |
| <u>Household size:</u>                     |                   |                   |
| 1  | 29                | 1.12              |
| 2  | 25                | 0.96              |
| 3  | 29                | 1.12              |
| 4+   | 23                | 0.88              |
| <u>Age of head of household:</u>           |                   |                   |
| 16 - 35                                    | 16                | 0.62              |
| 36 - 45                                    | 23                | 0.88              |
| 46 - 55                                    | 31                | 1.19              |
| 56 - 65                                    | 30                | 1.15              |
| 66 - 70                                    | 25                | 0.96              |
| 71+  | 35                | 1.35              |
| <u>Type of household:</u>                  |                   |                   |
| Lone parent with dependent children        | 17                | 0.65              |
| Married couple with dependent children     | 22                | 0.85              |
| Married couple without children            | 27                | 1.04              |
| Lone parent with non-dependent children    | 31                | 1.19              |
| Married couple with non-dependent children | 39                | 1.50              |
| <u>Employment status:</u>                  |                   |                   |
| Employees                                  | 24                | 0.92              |
| Self-employed                              | 35                | 1.35              |
| <u>Educational Qualifications:</u>         |                   |                   |
| Some                                       | 14                | 0.54              |
| None                                       | 25                | 0.96              |

Sources: Kemsley (1975), Redpath (1986) and Elliot (1991).

As was mentioned above, characterization of the non-respondents by matching with a census is possible when FES's are carried out at the same time as the census. This is not the case in many countries where FES's are conducted only once in several years and rarely close to the census. Thus the characterization of the non-response can be done only indirectly, when reliable external estimates of the distribution of the population by household characteristics are available. Such an analysis was made with the Israeli FES's.

The expected size of the sample of group k is  $E(n_k) = N_k f$ , where  $N_k$  is the known (or estimated) population size of group k and f is the sampling probability of the survey sample. The estimated non-response rate of group k, is  $(1-\phi_k) = 1 - r_k / E(n_k)$ , where  $r_k$  is the number of respondents in group k. These estimates are subject, of course, to sampling variance. Such estimates of non-response rates, for example, by household size, are presented in Table 6.3, for three Israeli FES's.

Table 6.3  
Non-response Rates by Household Size  
1975/76-1986/87 Israeli FES's

| Size of household | 1975/76 |                   | 1979/80 |                   | 1986/87 |                   |
|-------------------|---------|-------------------|---------|-------------------|---------|-------------------|
|                   | Percent | Relative to total | Percent | Relative to total | Percent | Relative to total |
| TOTAL             | 37.7    | 1.00              | 44.2    | 1.00              | 23.9    | 1.00              |
| 1                 | 45.9    | 1.22              | 60.3    | 1.36              | 41.1    | 1.72              |
| 2                 | 28.2    | 0.75              | 48.4    | 1.10              | 26.0    | 1.09              |
| 3                 | 37.7    | 1.00              | 48.5    | 1.10              | 20.7    | 0.87              |
| 4                 | 11.1    | 0.29              | 27.8    | 0.63              | 14.2    | 0.59              |
| 5                 | 32.6    | 0.86              | 37.2    | 0.84              | 18.0    | 0.75              |
| 6+                | 50.0    | 1.33              | 36.3    | 0.82              | 15.0    | 0.63              |

Households of size 1 have high non-response rates in the Israeli FES, as in the United Kingdom FES. However, in the Israeli surveys for different household sizes, the differences in response rates are much more substantial. These rates for households of size 1 relative to the overall non-response rate were about 1.1 in the U.K. 1971 and 1981 FES's, whereas in Israel they were 1.2 for the 1975/76 survey, 1.4 for the 1979/80 and 1.7 for the 1986/87 survey. This increase may be partly explained by the increased proportion of the population of households with one elderly person among households of size 1. By contrast, households of size 4 consistently co-operated more than others. It can be seen that the non-response rates by household size were not consistent over time, e.g., non-response rates of large households with 6 and more members, were lower than the average in the two recent surveys, but had the highest non-response rate in the 1975/76 survey. Maybe, this could be explained by the proportion of less-educated people in large families which was quite high in 1975/76 and since then has dramatically declined.

Estimates of non-response rates were similarly obtained for some other characteristics of the households found to be highly correlated with the household level of expenditure. As an example, the estimated non-response rate of groups of households defined by size of household and number of rooms of the dwelling in the Israeli 1986/87 FES, are presented in Table 6.4. These results show that within each household size, the non-response rate declines with the density of persons per room. This is a fairly good indicator that households of low socio-economic status were less likely to co-operate in the 1986/87 survey.

The above examples of the differential non-response rates between groups that can be expected to have different levels of expenditure, indicate how serious could be the bias in the survey estimates due to non-response. Thus, much effort is invested in attempting to reduce the non-response rate: a wide publicity campaign is mounted before and during the survey; a letter is mailed to each selected household just before the interviewer's visit explaining the importance of the survey and how important it is to participate in the survey; and, the interviewers supply brochures with further explanations of the purposes and uses of the survey.

Furthermore, the interviewers are specially trained to persuade the families to participate in the survey and assure the respondents of confidentiality. Households which initially were unwilling to co-operate are sent personal letters from the central statistical office and are visited again by the local supervisor in order to encourage them to co-operate. In addition, various kinds of incentives are often used, and their effect is discussed in the following section.

Table 6.4  
Estimates of Non-response Rates by Household Size and Number of Rooms  
1986/87 Israeli FES

| Household size | Number of rooms | Non-response rate |               |                  |
|----------------|-----------------|-------------------|---------------|------------------|
|                |                 | Percent           | Relative to   |                  |
|                |                 |                   | Total HH size | Total population |
|                | TOTAL           | 24                | ..            | 1.0              |
| Size 1         | <u>Total</u>    | <u>41</u>         | <u>1.0</u>    | <u>1.7</u>       |
|                | 1.0-1.5         | 52                | 1.3           | 2.2              |
|                | 2.0-2.5         | 40                | 1.0           | 1.7              |
|                | 3.0+            | 39                | 0.9           | 1.6              |
| Size 2         | <u>Total</u>    | <u>26</u>         | <u>1.0</u>    | <u>1.1</u>       |
|                | 1.0-2.0         | 32                | 1.2           | 1.3              |
|                | 2.5-3.0         | 26                | 1.0           | 1.1              |
|                | 3.5+            | 14                | 0.5           | 0.6              |
| Size 3         | <u>Total</u>    | <u>21</u>         | <u>1.0</u>    | <u>0.9</u>       |
|                | 1.0-2.5         | 26                | 1.3           | 1.1              |
|                | 3.0             | 18                | 0.8           | 0.7              |
|                | 3.5+            | 21                | 1.0           | 0.9              |
| Size 4         | <u>Total</u>    | <u>14</u>         | <u>1.0</u>    | <u>0.6</u>       |
|                | 1.0-2.5         | 24                | 1.7           | 1.0              |
|                | 3.0             | 17                | 1.2           | 0.7              |
|                | 3.5+            | 10                | 0.7           | 0.4              |
| Size 5         | <u>Total</u>    | <u>18</u>         | <u>1.0</u>    | <u>0.8</u>       |
|                | 1.0-2.5         | 29                | 1.6           | 1.2              |
|                | 3.0-3.5         | 23                | 1.3           | 1.0              |
|                | 4+              | 12                | 0.7           | 0.5              |
| Size 6+        | <u>Total</u>    | <u>15</u>         | <u>1.0</u>    | <u>0.6</u>       |
|                | 1.0-3.0         | 27                | 1.8           | 1.1              |
|                | 3.5-4.0         | 19                | 1.3           | 0.8              |
|                | 4.5+            | 10                | 0.7           | 0.4              |

## 6.4 Incentive Effects

Usually, official statistical organizations do not give gifts or pay respondents for participating in surveys. However, because of the relatively high rate of non-response in FES's and because of the heavy burden that these surveys impose on the participants, incentives have been always considered, cash payments or various kinds of gifts. The timing of offering the gift, on the first visit as an encouragement to agree to participate, or at the end of the investigation as a token of appreciation, have also been considered. Several experiments were carried out to evaluate the effects of different kinds of incentives on improving the response rate.

One of the first experiments on the effects of incentives in FES's was conducted in 1951 in the United Kingdom (Kemsley and Nicholson, 1960) in two surveys each with a sample size of about 1,500 addresses:

- (1) In the first survey (conducted in January 1951), each adult (16 years and over) in each household (except those still at school) was first requested to be interviewed and to give details of regular and infrequent purchases, and was then asked to record all expenditures in the following week in a diary. At the end of the week, the interviewers collected the diary and asked further information about income. The sample was split into three sub-samples of the same size and each interviewer was allocated the same number of addresses in each of the three sub-samples. Varying rates of payment for co-operation were offered at the end of the first interview: nothing to the first sub-sample; 5s. to each person co-operating in the second, and 10s. to each person co-operating in the third sub-sample.
- (2) In the second survey (carried out in July-September 1951), the sample was split into two sub-samples. In the first sub-sample, all adults were asked to keep a one-week diary, whereas in the second one, all adults were asked to keep a four-week diary. Adults in both sub-samples were asked to complete an additional questionnaire giving full details of their income.

The payments were conditional on filling in both income and expenditure schedules. In the first sub-sample, the rate of payment to each person co-operating was 2½s. with an additional 7½s. per person if all adults in the household co-operated, making a maximum payment of 10s. per person. In the second sub-sample the rate of payment was 2½s. per person per week with an additional 2½s. per person if all adults in the household co-operated in that week, and an additional 5s. per person if they all co-operated for all 4-weeks, making a maximum payment of 25s. per person.

The results of these experiments are presented in Table 6.5 and they show clearly that payments to those who co-operate in surveys of this kind bring a substantial increase in response rates.

In the first survey, the effect of the first 5s. appears to be slightly greater than that of the additional 5s., which is nevertheless substantial (and statistically significant). The payment of 10s. raised the rate of full co-operation from 36% to 53%. In that survey, payments were made to each individual who co-operated regardless of whether other members of the same household co-operated. As a result, payments were made to about 14% of the households from which only partial co-operation was obtained.

In the second survey, higher payments were made to each person only if full co-operation of the household was obtained. This led to a considerable reduction in the rate of partial co-operation to about only 3% of all households for a 1-week diary.

Since the data obtained from partial co-operation cannot be used, these cases cannot be regarded as response. Nevertheless, the results indicate that the payment of a maximum 10s. per person conditional on the participation of the other members of the household results in a similar rate of full response as the unconditional payment of 10s. per person, but the latter is more expensive. A reduced payment of 5s. per person resulted in a lower rate of full response but is, of course, much less costly.

The effect on the response rate of asking households to keep a diary for more than one week, was evaluated in the second survey. The rates of payment were not proportional to the amount of work involved. It is not surprising, therefore, that a lower rate of full co-operation was obtained with a maximum payment of 25s., as compared to the one-week diary with a maximum of 10s. Even so, 46% of the households co-operated with a four-week diary, when payments were offered, whereas only 36% co-operated with a one-week diary when no incentive was offered. The general conclusion of this experiment was that offering payments to FES participants undoubtedly has positive effects on the response rate.

Table 6.5  
Response Rates with Monetary Incentives  
1951 Experimental FES's, U.K. Central Statistical Office

| Sub-sample                                       | Total households contacted |         | Percentage Co-operating |        |      |
|--|----------------------------|---------|-------------------------|--------|------|
|  | Number                     | Percent | Fully                   | Partly | Non  |
| <u>First survey:</u>                             |                            |         |                         |        |      |
| 1) <u>1-week diary</u> with No payment           | 479                        | 100.0   | 35.7                    | 16.5   | 47.8 |
| 2) <u>1-week diary</u> with 5s./person           | 490                        | 100.0   | 46.7                    | 14.7   | 38.6 |
| 3) <u>1-week diary</u> with 10s./person          | 500                        | 100.0   | 53.4                    | 13.8   | 32.8 |
| <u>Second survey:</u>                            |                            |         |                         |        |      |
| 1) <u>1-week diary</u> with max. 10s./person(*)  | 466                        | 100.0   | 55.6                    | 3.0    | 41.4 |
| 2) <u>4-week diary</u> with max. 25s./person(**) | 898                        | 100.0   | 46.4                    | 5.8    | 47.8 |

(\*) 2½s./person-week + 7½s. bonus; (\*\*) 2½s./person-week + 2½s. bonus/week + 5s. bonus  
Based on Kemsley and Nicholson (1960).

Sudman and Ferber (1971) conducted a pilot study for the U.S.A. BLS that was aimed at studying the effects of offering gifts, rather than money compensation. Four random sub-samples received different treatment:

- (1) No gift, except for a plastic folder and a pen for use in recordkeeping (a value of about 1.00\$) not referred to as a gift by the interviewer although the households were allowed to keep them.
- (2) American flag with holder, posters or one of 40 books (value of about 5.00\$) was offered.
- (3) A report of the household purchases summarized by major categories and compared with other families with similar incomes, was offered.
- (4) A large padded stationery holder with a pen (retailing for about 5.00\$) was given before the household started record keeping.

Table 6.6 shows that co-operation was consistently lower in the group that received no gift, regardless of the length of the record keeping period and that the magnitude of the differences increased as the record keeping period lengthened.

For the four-week period, less than half as many respondents in the group who received no gift completed all four diaries (23%), as compared with all the groups who received some gift (49%). Among the gifts, the immediate gift (the stationery holder) was more effective than either of the other delayed gifts (except for the four-week period). Of the two delayed gifts, the offer of a report seems slightly better than the offer of a flag or a book. This study's results suggest that an initial gift is more effective in the short run, and some reinforcement is required for a longer period.

Table 6.6  
Co-operation Rates by Type of Gift and by Length of Diary Keeping  
percentages(\*)

| Extent of co-operation | With<br>no<br>gift | With<br>a<br>gift | Type of gift       |                   |                               |
|------------------------|--------------------|-------------------|--------------------|-------------------|-------------------------------|
|                        |                    |                   | Flag<br>or<br>book | Summary<br>report | Large<br>stationery<br>holder |
| Kept at least:         |                    |                   |                    |                   |                               |
| 1 week                 | 77                 | 85                | 79                 | 86                | 90                            |
| 2 weeks                | 62                 | 75                | 71                 | 71                | 83                            |
| 3 weeks                | 54                 | 67                | 64                 | 67                | 70                            |
| 4 weeks                | 23                 | 49                | 51                 | 55                | 39                            |

\*) out of those asked to keep diaries for at least that length of time.

Based on Sudman and Ferber (1971).

Recently, an experiment with incentives was conducted in Sweden (Lindstrom, 1991), during the 1988 FES. The Swedish FES's have, since 1958, given some kind of compensation to families that take part in the survey. The value of the compensation has, however, gone down in each successive survey, one reason being that it was found that what is important is the promise of some reward, rather than its value. However, whereas, previous FES's attained around a 70% response rate, the response rate of the 1988 survey fell considerably, only 55% in the first three months of the survey, and even less in the subsequent month (45%). While the survey was in progress, a new strategy of incentives was adopted, using incentives in advance and giving a pocket calculator as a gift. In addition, a controlled experiment was built in the ongoing survey for evaluating the effect of such incentives on the response rate. The conclusion was to continue with the new strategy.

Incentives may also lead to better reporting of the respondents. An experiment for evaluating effects of this kind was conducted in the U.S. Bureau of the Census with the 1972/73 FES (Cowan, 1977) for 8 weeks of the survey, with three treatments randomly assigned to three sub-samples: no payment, 5\$ and 10\$. Each interviewer handled only one treatment, thus interviewers' effects were not fully controlled. Each sub-sample included about 500 households that completed one or both one-week diaries. The average expenditure per household, the proportion of respondents who reported on expenditure and the average of the reported expenditure of the three sub-samples were compared for each of the 72 main expenditure categories.

Examples of the results obtained in this study for frequent expenditure categories for which estimates are primarily derived from the diary, are presented in Table 6.7. For all these categories (except one) the proportion of the respondents who reported on expenditure increased with the incentive, and for almost all of them the increase was larger, the greater the amount of incentive. The average weekly expenditure for the reported items was also found to rise with the amount of incentive in most of the categories. As a consequence, the effects of the incentives on the average weekly expenditure was not negligible, e.g., for Total Food the increase was by 1.13, for Tobacco by 1.22, for Dairy products by 1.1, etc.

However, this does not mean necessarily that incentives reduce reporting biases. The assumption that "more is better" is not always correct. As an example, if non-response is more frequent among households of size 1, and if their average weekly expenditure is lower than other households, incentives should result in a lower average if the response rate is increased.

**Table 6.7**  
**Effects of Incentives - U.S.A. 1972-73 Consumer Expenditure Survey**

| Selected expenditure categories   | Amount of incentive |              |              | Ratio to no payment |             |             |
|---|---------------------|--------------|--------------|---------------------|-------------|-------------|
|   | 0\$                 | 5\$          | 10\$         | 0\$                 | 5\$         | 10\$        |
| <b>1. Percentages of Respondents Who Reported on Expenditure Category</b> |                     |              |              |                     |             |             |
| Total Food at Home:   | <u>83.4</u>         | <u>89.1</u>  | <u>87.1</u>  | <u>1.00</u>         | <u>1.07</u> | <u>1.04</u> |
| Bakery products   | 78.7                | 82.3         | 81.6         | 1.00                | 1.05        | 1.04        |
| Dairy products  | 82.4                | 85.3         | 84.4         | 1.00                | 1.04        | 1.02        |
| Meat  | 75.4                | 77.1         | 78.9         | 1.00                | 1.02        | 1.05        |
| Vegetables  | 67.3                | 68.8         | 71.3         | 1.00                | 1.02        | 1.06        |
| Non-alcoholic beverages   | 67.5                | 71.5         | 73.5         | 1.00                | 1.06        | 1.09        |
| Food & Beverages away from home   | 64.5                | 69.0         | 75.2         | 1.00                | 1.07        | 1.17        |
| Alcoholic beverages   | 28.5                | 29.6         | 30.0         | 1.00                | 1.04        | 1.05        |
| Tobacco   | 42.6                | 50.6         | 49.2         | 1.00                | 1.19        | 1.15        |
| Household supplies  | 70.7                | 73.9         | 79.0         | 1.00                | 1.05        | 1.12        |
| Clothing & Footwear   | 43.2                | 45.3         | 49.0         | 1.00                | 1.05        | 1.13        |
| <b>2. Average Reported Weekly Expenditure</b>                             |                     |              |              |                     |             |             |
| Total Food at Home:   | <u>21.39</u>        | <u>22.68</u> | <u>23.15</u> | <u>1.00</u>         | <u>1.06</u> | <u>1.08</u> |
| Bakery products   | 2.21                | 2.33         | 2.38         | 1.00                | 1.06        | 1.08        |
| Dairy products  | 3.24                | 3.36         | 3.48         | 1.00                | 1.04        | 1.08        |
| Meat  | 7.16                | 7.41         | 7.43         | 1.00                | 1.03        | 1.04        |
| Vegetables  | 2.21                | 2.24         | 2.31         | 1.00                | 1.01        | 1.05        |
| Non-alcoholic beverages   | 2.41                | 2.46         | 2.60         | 1.00                | 1.02        | 1.08        |
| Food & Beverages away from home   | 10.05               | 10.14        | 10.97        | 1.00                | 1.01        | 1.09        |
| Alcoholic beverages   | 6.42                | 6.79         | 8.30         | 1.00                | 1.06        | 1.29        |
| Tobacco   | 4.46                | 4.27         | 4.72         | 1.00                | 0.96        | 1.06        |
| Household supplies  | 3.41                | 3.55         | 3.52         | 1.00                | 1.04        | 1.03        |
| Clothing & Footwear   | 20.69               | 21.55        | 21.94        | 1.00                | 1.04        | 1.06        |
| <b>3. Average Weekly Expenditure</b>                                      |                     |              |              |                     |             |             |
| Total Food at Home:   | <u>17.84</u>        | <u>20.21</u> | <u>20.16</u> | <u>1.00</u>         | <u>1.13</u> | <u>1.13</u> |
| Bakery products   | 1.74                | 1.92         | 1.94         | 1.00                | 1.10        | 1.11        |
| Dairy products  | 2.67                | 2.87         | 2.94         | 1.00                | 1.07        | 1.10        |
| Meat  | 5.40                | 5.71         | 5.86         | 1.00                | 1.06        | 1.09        |
| Vegetables  | 1.49                | 1.54         | 1.65         | 1.00                | 1.03        | 1.11        |
| Non-alcoholic beverages   | 1.63                | 1.76         | 1.91         | 1.00                | 1.08        | 1.17        |
| Food & Beverages away from home   | 6.84                | 7.00         | 8.25         | 1.00                | 1.08        | 1.27        |
| Alcoholic beverages   | 1.83                | 2.01         | 2.49         | 1.00                | 1.10        | 1.36        |
| Tobacco   | 1.90                | 2.16         | 2.32         | 1.00                | 1.14        | 1.22        |
| Household supplies  | 2.41                | 2.62         | 2.78         | 1.00                | 1.09        | 1.15        |
| Clothing & Footwear   | 8.94                | 9.76         | 10.75        | 1.00                | 1.09        | 1.20        |

Based on Cowan (1977).

Results from other experiments that have been carried out in order to measure the effectiveness of incentives, indicate that in general incentives do improve response rates. Especially, is this the case in diary surveys where the respondents are required to commit themselves to a continuing task, rather than the usual one-time interview.

Thus, in many countries, payments of various amounts of money or other kinds of incentives, or both, are used. For instance, in the United Kingdom each person is paid; in Ireland the family can choose between each person being paid or participating in a lottery with high money prizes; in Denmark each family is paid and also participates in a lottery with 100 money prizes. In the Israeli 1986/87 survey a kitchen-scale was given to each sampled household on the first visit of the interviewer even before they agreed to participate in the survey. There were no doubts that such an incentive was very effective and, according to the interviewers' reports, saved them a lot of effort in persuading the families to take part in the survey. Although the kitchen-scale was a very impressive gift, the interviewers who had to carry it complained because of its size. Thus, for the 1992/93 survey, a quite sophisticated calculator of a similar value, but much easier to carry, was chosen as the gift. It should be mentioned that in previous surveys the gift was offered only after the final interview and was not very effective. On the other hand, there are countries with no incentives at all, like the U.S.A., Italy, Greece and others.

In conclusion, although it is clear that incentives reduce non-response, and probably also improve the reporting quality, there is no easy way to quantify the benefits derived using incentives versus their cost. One alternative possibility is to use incentives only for certain sub-groups of the population which are subject a priori to higher non-response rates (however, this can be seen as unfair). There may always be doubts whether it would not be better to invest the incentive expense in improving the field work, such as in better training of interviewers, or in increasing the number of the interviewer's visits to the households during the diary period.

## CHAPTER 7. WEIGHTING

### 7.1 Weighting in FES's

In FES's, as with all sampling surveys, a weight is determined for each responding unit of the sample to express the number of units in the population represented by that unit. The weighting procedure aims to minimize the M.S.E. of the survey estimates, comprising the sampling variance and the biases in the net sample, primarily due to non-response. Because of the relatively high rates of non-response in FES's and because these surveys aim to supply detailed information, the weighting is of special importance.

For weighting, first, the inverse of the selection probability of each household determines an initial weight. Then, the initial weights are adjusted and the final weights are usually determined by a complex multi-stage weighting procedure.

One adjustment often used is that of correcting for differential non-response rates within geographical areas. This may not be effective in FES's since the non-response is almost entirely due to refusals and the residential area does not necessarily characterize their expenditures.

Another adjustment often required is that of correcting for an uneven distribution of the net sample over the survey period. In practice, even when the sample is initially distributed evenly over the sub-periods of the entire survey period, the net sample is not balanced over time. Since the variables that are investigated in FES's vary over time, such a distortion can lead to biases in estimates of a seasonal nature. For example, postponement because of absentees can cause an upward bias in estimates of vacation expenses. There are several ways to reduce the bias in time by the weighting process, by assigning different weights to the different periods investigated for the same household, or by having one weight for each household.

Sometimes, additional adjustments for certain sub-groups that were not fully covered in the sample are also required. One example is an adjustment to compensate for not updating the sample of new construction throughout the survey period.

Important stages in the weighting are adjustments for key characteristics that are correlated with the investigated variables. This can be done by using the survey data themselves, only when special efforts are made to collect these key characteristics from the non-response and they are successful, or when the sampling frame can supply the relevant data. Then, correcting factors are applied to the respondents to compensate for similar non-respondents.

Alternatively, adjustments that make use of an external distribution of the survey population by key characteristics, i.e., post-stratification and its variants, are widely used in FES's. Some methodological issues related to such adjustments are discussed in the following sections.

### 7.2 Effects of Post-stratification

In general, for a simple random sample of size  $n$  out of a population of size  $N$  with  $r$  respondents in the survey ( $r \leq n$ ), the simple estimate is actually the estimate of the mean of the respondents' population:

$$\bar{y}_S = \sum y_i / r.$$

Post-stratification can be applied if each respondent can be classified into one of the post-strata,  $h=1,2,\dots,H$ , providing the population weights  $w_h=N_h/M$  are known. Then, the post-stratification estimate of the population mean is

$$\bar{y}_{PS} = \sum w_h \bar{y}(R)_h, \text{ where } \bar{y}(R)_h \text{ are the simple estimates based on } r_h \text{ respondents in strata } h.$$

Post-stratification can also be applied by using reliable estimates of  $w_h$  which can be derived from another survey of much larger scale, and substituting  $\hat{w}_h$  for  $w_h$ .

Although post-stratification can be applied to more complicated sampling designs and to other types of estimates, the following presentation is, for the sake of simplicity, limited to the simple case without loss of generality.

If the response rates are  $\phi_h=r_h/n_h$  and  $\phi=r/n$  for strata  $h$  and for the entire sample, the simple estimate can be expressed as

$$\bar{y}_S = \sum w_h (\phi_h/\phi) \bar{Y}(R)_h, \text{ where } w_h=n_h/n.$$

Whereas with the simple estimate, all respondents have implicitly the same weight, by post-stratification different weights might be given to the different post-strata. The weights are not necessarily the same, first, since the  $n_h$  are random variables, and second, due to possible variation of the response rates in different strata.

Let us examine the effects of post-stratification, by a given classification into post-strata, on the survey estimates.

#### a) The Effects on Sampling Variance

The ratio of the expected sampling variances of the post-stratification estimate and of the simple estimate can be presented as comprising three components:

$$\frac{\bar{E}[\text{Var}(\bar{y}_{PS})]}{\bar{E}[\text{Var}(\bar{y}_S)]} = (1-R^2)(1+\delta)(1+D).$$

In the following it is assumed that  $N_h$ , the population sizes of the strata are large enough, i.e.,  $1/N_h$  are negligible,  $n$  also is large enough, but the fpc's are negligible, and  $S^2_h \approx S^2(R)_h$ , i.e., the variances of the respondents and the non-respondents are similar.

- (1) The first component,  $1-R^2$ , is the reduction of the sampling variance which would have been obtained by proportional stratified sampling as compared with a simple random sampling, under full response.

$$1-R^2 = \frac{\text{Var}(\bar{y}_{ST} | \phi=1)}{\text{Var}(\bar{y}_S | \phi=1)} \approx S^2_w / S^2 = 1 - S^2_b / S^2$$

where  $S^2_w$  is the within-strata variance,  $S^2_b$  the between-strata variance and  $S^2$  the overall variance, thus  $0 \leq R^2 \leq 1$ . A classification into post-strata which produces a higher value of  $R^2$ , i.e., with higher divergence between the strata means, will result in a greater reduction of the variance.

- (2) The second component,  $1+\delta$ , is the increase in variance due to the fact that under post-stratification the sample is not selected a priori as a stratified one.

$$1+\delta = \frac{\bar{E}[\text{Var}(y_{PS})]}{\bar{E}[\text{Var}(y_{ST})]} \approx \Omega / n,$$

where  $n$  is the average sample size per stratum, and  $\Omega$ , which may not be negligible, is the ratio between  $(1/H)\sum S_h^2/\phi_h$  and  $\sum W_h S_h^2/\phi_h$ .  $\Omega=1$  when  $W_h=1/H$ , or when  $S_h^2$  and  $\phi_h$  are the same for all  $h$ . However, the second condition is not likely to be fulfilled. Thus, for a given sample size,  $\delta$  is minimized, first, by refraining from too many post-strata, providing the average sample size in the final post-strata will be large enough. Second, by minimizing  $\Omega$ , or at least avoiding  $\Omega>1$  by constructing the post-strata with similar sizes, for example, for  $n_h \approx 25$ , the variance will increase in comparison with a proportional stratified sample approximately by 1.04.

- (3) The third component,  $1+D$ , is the effect on the variance resulting from the variation in the response rates between the strata.

$$1+D = \left[ \frac{\text{Var}(y_S | \phi=1)}{\bar{E}[\text{Var}(y_S)]} \right] \left[ \frac{\bar{E}[\text{Var}(y_{ST})]}{\text{Var}(y_{ST} | \phi=1)} \right].$$

It can be shown that

$$1+D \approx \phi \frac{\sum W_h S_h^2 / \phi_h}{\sum W_h S_h^2} \approx \frac{\text{Var}(y_{OPT}) [1 + \text{Relvar}(S_h G_h)]}{\text{Var}(y_{ST})}$$

where,  $G_h=N_h/E(r_h)$  are the expected weights and  $\text{Var}(y_{OPT})=(\sum W_h S_h)^2/n$  is the sampling variance for an optimal stratified sampling of size  $n$ . For the case in which  $S_h^2$  are similar for all  $h$ ,  $1+D \approx S^2 [1 + \text{Relvar}(G_h)]$ , and is widely used as an approximation for the contribution of different weights to the increase in sampling variance. Hence, in order to avoid increasing the variance it is desirable to avoid post-strata with too high non-response rates.

In the rare case, where there is a high positive correlation between  $S_h$  and  $\phi_h$ , so that  $S_h/\phi_h$  are nearly the same for all strata, the sample allocation between the post-strata may become de facto optimal and  $1+D<1$ . If so, this component may even contribute to a decrease of the variance.

- (4) An additional increase in sampling variance would be when estimates of the population weights,  $\hat{W}_h$ , are used. This additional increase is approximately  $S^2 b/n'$ , where  $n'$  is the size of the sample from which  $\hat{W}_h$  are estimated. Thus, the sampling variance will increase as  $n'$  is smaller, and as the variance between the post-strata is larger, namely, when  $R^2$  is larger. It should be noted that for estimates of sub-groups this additional effect will be smaller, the smaller the sub-group is.

Since the increase in sampling variance from using estimates of  $W_h$  can be dominant over the decrease of the variance that is achieved by a better classification into post-strata, the use of post-stratification is justified only if for a given ratio  $n/n'$ ,

$$R^2 > \frac{(1+\delta)(1+D) - 1}{(1+\delta)(1+D) - n/n'}$$

b) The Effects on Non-response Bias

The reduction of the non-response bias by post-stratification depends on the following components:

- (1) The first component is the divergence of the response rates,  $\phi_h$ , between the post-strata. In the rare case of full response for all strata ( $\phi_h=1$ ), the simple and the post-stratification estimates are both unbiased. When the response rates are the same for all strata, both estimates have the same bias. Therefore, the use of post-stratification for reducing the non-response bias should be considered only when the post-strata are constructed so that the response rates are not the same for all strata.
- (2) The second component is the divergence between the post-strata means. The higher this divergence, i.e., the more homogeneous the post-strata are, the more reduction of the bias can be achieved by post-stratification.
- (3) The third component depends on the differences between respondents and non-respondents within post-strata. More similarity between respondents and non-respondents within post-strata is expected the more homogeneous are the post-strata, and thus the less biased is the post-stratification estimate. Nevertheless, these differences might not be negligible and the post-stratification estimate will remain biased.
- (4) The forth component depends on the relationship between the above three components. Post-stratification will reduce the bias only under certain conditions, otherwise, it may produce greater biases than with the simple estimate.
- (5) Another source of bias is when biased  $W_h$  are used for post-stratification. If the bias from this source is substantial, the use of post-stratification might not be justified.

All the above components of both the effects on variance and bias should be taken into account when the post-strata are constructed and when deciding whether post-stratification should be used or not. Both the sampling variance and the non-response bias would be reduced only with sufficiently high  $R^2$  and this can be achieved by constructing homogeneous post-strata under the above mentioned constraints. However, in general, a trade-off between the effects on the bias and on the variance has to be made.

### 7.3 Classification Trees of Homogeneous Sub-groups

Usually, the post-strata are determined by a complete cross-classification of the units by a number of classifying variables, for which external data or reliable estimates of the size of the population are available. However, for a given sample size, the number of classifying variables and their categories is limited, because each ultimate group has to be sufficiently large. Another weighting method of "raking", by which marginal distributions of the population are being controlled, allows the use of a greater number of variables, although with no complete cross-classification.

Alternatively, the determination of the post-strata can be based on a hierarchical classification by constructing a tree of homogeneous groups, in which each branch of the tree can be further divided by a different variable and not necessarily to the same number of sub-branches. This makes possible the use of a greater number of classifying variables and potentially provides a better classification into homogeneous post-strata for a given sample size (Kantorowitz, 1990).

Two procedures for constructing a classification tree of homogeneous groups for one single estimated variable are outlined in the following.

- a) A Stepwise Multi-branch Procedure - This classification procedure is performed hierarchically according to a set of potential categorical classifying variables. At the first step the classifying variable which gives the maximum reduction for the within-variance of the entire population, i.e., maximum  $R^2$ , is selected. The second step is to determine for each of the groups defined by the first step, a classifying variable which will give maximum  $R^2$  within that group. Similarly, the third step is to determine the best classifying variables, separately for each of the groups determined by the second step, and so on, step by step.

At each step, for each group intended for further classification, the number of categories and the boundaries of the classifying variables not used in the previous steps, are redefined. This is done on the basis of their sample sizes, on the similarity of their averages, and according to their definitions (for example, close age groups, similar occupations, etc.).

A stopping rule is applied at each step for each group, taking into account the accumulated and the marginal effects mentioned above in order to decide whether to stop the classification at the previous step, or to combine groups. However, in practice, this is done only at the advanced stages.

By this procedure, a tree is constructed whose various branches at different stages do not necessarily have the same number of sub-branches since this depends on the number of categories of the classifying variable used for each sub-branch. Also, not every branch reaches the top, because the "thickness" of the branch not always allow the use of all the classifying variables. A classifying variable can be used only once in the different nodes of the same branch.

This classification is not unique because many judgmental considerations are involved. For example, the choice of the classifying variable at each step obviously depends on the categorization of the different classifying variables. At every step, it is also possible to consider whether to classify according to one variable with a considerable degree of categorization or to limit the number of categories and so, make possible further classification by an additional classifying variable at the next step. Furthermore, similar to the situation for stepwise regression, such a classification is not necessarily the optimal one. This does not, however, diminish the advantages of using such classification even if it is always possible to construct a better one.

- b) A CART-based Binary-tree Procedure - First, the CART software (Breiman, Friedman, Olshen and Stone, 1984), is used to obtain a binary tree, i.e., when a branch of the tree is divided, it is always into two sub-branches. Thus, the same classifying variable may be used more than once in different nodes of the same branch. The classifying variables can initially be quantitative or categorical, and for each of them at each node, the best two sub-groups are determined and the best classifying variable is chosen, in order to minimize the within variance of that node, i.e., to maximize  $R^2$ .

Because a too-detailed tree is not suitable for practical uses, CART "prunes" the tree, according to a predetermined pruning rule that depends on a balance between the marginal gain in reducing the within-variance and a certain measure of the complexity of the tree. Some other optional constraints can be used for stopping rules, such as defining the maximal sample size for further classification.

The CART tree provides an optimal classification (under certain constraints), however, with respect only to  $R^2$ . Thus, the resulting tree is often not satisfactory for post-stratification, because there is no way to control the other effects, as with the Stepwise multi-branch tree. Therefore, the CART tree must be modified. For example, in order to obtain ultimate groups of a certain range of a sample size, a very detailed tree is needed and manual pruning has to be applied.

The CART procedure is applicable only for a self-weighted sample, whereas the Stepwise Multi-branch procedure can be applied for data of a sample with unequal weights.

Both procedures are suitable for one single target variable. However, most surveys are multi-purpose and supply estimates for a variety of variables. Usually there would not be a single best classifying tree for all important target variables, and obviously classification into homogeneous groups with regard to only one variable will be beneficial for that variable but it might be ineffective for the others, unless they are highly correlated with the chosen one.

One way is to choose one estimated variable to "represent" the others (the most important variable, or among all the important variables, that which is most liable to have the highest sampling error, etc.). Another way is to construct a single synthetic variable based on a number of "representative" estimated variables and then construct a single tree. Various types of synthetic variables can be constructed, one example being the use of the Principle Component technique.

An alternative solution is to use one or several "combined" criteria for choosing the best classifying variable for each group. However, this is applicable only under the Stepwise multi-branch procedure. Such criteria may be based on functions of  $R^2$  (or on  $(1-R^2)(1+\delta)(1+D)$ ) that are obtained for the different target variables in each step of the tree construction. Let us look at some simple criteria.

Let  $R^2_{ij}$  be the value of  $R^2$  for the  $j$ -th estimated variable ( $j=1,2,\dots,J$ ) and for the  $i$ -th classifying variable ( $i=1,2,\dots,I$ ). The classifying variable  $i$  with the highest value of one of the following is the best.

- (1) The number of estimated variables for which  $\text{Max}_j(R^2_{ij})$  is obtained - by this all the estimated variables have the same influence, even those for which  $R^2$  is small for all the classifying variables.
- (2) The average  $\bar{R^2_i} = \sum_j R^2_{ij} / J$  - by this, estimated variables for which  $R^2$  are high will contribute more to determination of the classifying variables. This is equivalent to minimizing the total "losses" of all the estimated variables if use is made of a classifying variable different from the best one for each of them.
- (3) The average  $\bar{RR^2_i} = [\sum_j R^2_{ij} / \text{Max}_j(R^2_{ij})] / J$  - by this, even estimated variables for which  $R^2$  is low for all the classifying variables, but this ratio is high, may have much influence on the choice of the classifying variable. This is equivalent to minimization of the total "relative losses" of all the estimated variables by choosing a classifying variable different from the best one for each of them.

Variations of these criteria may be used by weighting the estimated variables, for example, by rating their importance, or by their variance, or by the remaining variance left at every step of the classification. Combined criteria of this kind can be applied to only a limited number of estimated variables, because the efficiency might decrease for each of the estimated variables separately and also because of technical difficulties.

## 7.4 Illustration of Classification Trees

Constructing a tree is illustrated here first for one estimated variable as in the Israeli Income Survey. This survey resembles the FES in many respects and is therefore used here for illustrative purposes. Afterwards, tree construction for a set of estimated variables is illustrated as in the Israeli Family Expenditure Survey. For both surveys the Stepwise procedure was applied.

a) Construction Tree for Net Income per Household in Income Survey

Until 1984, post-stratification in Income Surveys was based on a cross-classification of three variables: sex and age of economic head of household, and size of locality. Since 1985, post-stratification has been implemented using a classification tree and the classifying variables chosen were: size of household, number of employed persons in household, size of locality, and characteristics of economic head of household - employment status, age, continent of birth, years of study, occupation, economic branch, and weekly hours of work.

At the first stage, the classification was according to three primary groups of household heads: employees, self-employed and not working, since all estimates of the survey are given separately for these groups. At the following stages, it turned out that every classifying variable was used, except size of locality and weekly hours of work because they contributed the least to the reduction of the variance.

The cumulative values of  $R^2$  are given in Table 7.1, corresponding to the various steps of the classification, for all households and for each of the three primary sub-groups.

Table 7.1  
Cumulative  $R^2$  by Classification Stage and by Three Sub-groups  
Net Income per Household in the Israeli Income Survey

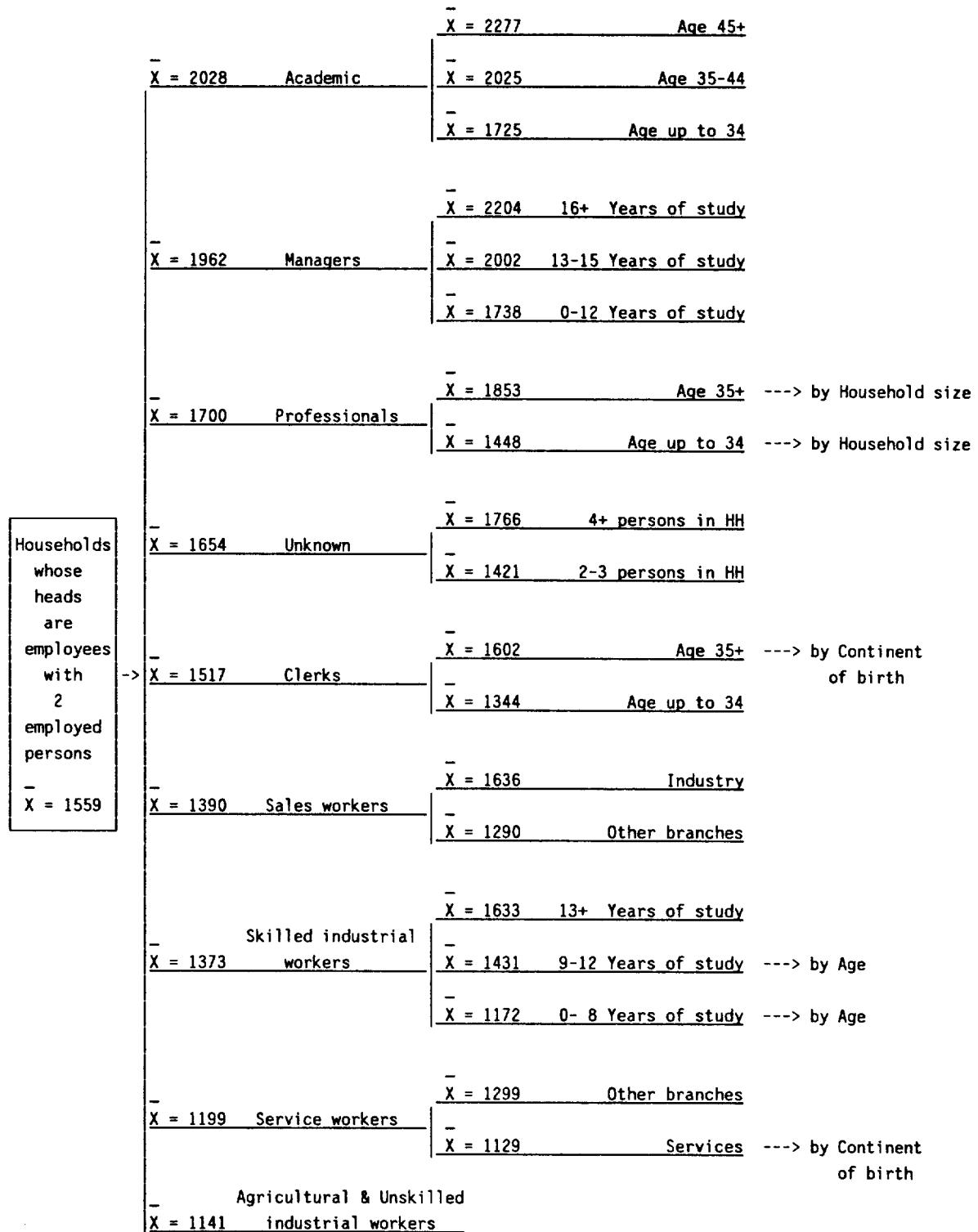
| Stage | Total         |                  | Household heads           |                  |                              |                  |                             |                  |
|-------|---------------|------------------|---------------------------|------------------|------------------------------|------------------|-----------------------------|------------------|
|       |               |                  | Employees<br>$w = 63.4\%$ |                  | Self-employed<br>$w = 7.7\%$ |                  | Not working<br>$w = 28.9\%$ |                  |
|       | No. of groups | Cumulative $R^2$ | No. of groups             | Cumulative $R^2$ | No. of groups                | Cumulative $R^2$ | No. of groups               | Cumulative $R^2$ |
| 1     | 3             | 0.277            | 1                         | -                | 1                            | -                | 1                           | -                |
| 2     | 9             | 0.452            | 3                         | 0.279            | 3                            | 0.132            | 3                           | 0.095            |
| 3     | 40            | 0.527            | 23                        | 0.395            | 9                            | 0.180            | 8                           | 0.165            |
| 4     | 84            | 0.560            | 57                        | 0.447            | 13                           | 0.199            | 14                          | 0.192            |
| 5     | 100           | 0.569            | 70                        | 0.463            | 14                           | 0.201            | 16                          | 0.198            |

The values of  $R^2$  obtained by the tree were: for the overall population 0.57; for households whose head is an employee 0.46; and for the two other groups, only 0.20. Households whose heads are self-employed constitute less than 8% of the overall population and bearing in mind the constraints imposed by the sample size, a more detailed classification could not be used.

The group of households whose head does not work, although larger, was classified only into 16 categories despite the fact that this group constitutes about 30% of the overall population. A better classification of that group could have been achieved if more suitable classifying variables would have been chosen. A revised tree (with more classifying variables) that has recently been constructed using the CART software, achieved a higher value of  $R^2$  (about 0.35) for that group.

It is difficult to present here the complete classification tree with its 100 sub-groups. As an example, Figure 7.1 presents one of the branches of households whose heads are employees with 2 employed persons in the household. This sub-group was first classified by 9 occupation groups. Then, academics, professionals and clerks were classified by age, managers and skilled industrial workers by number of years of study, services and sales workers by economic branch, while the group of agricultural and unskilled industrial workers were not further classified.

**Figure 7.1**  
**Classification Tree for Net Income per Household ( $X$ ) in the Israeli Income Survey**  
**Example of one sub-branch**



b) Constructing a Tree for a Set of Estimated Variables in Family Expenditure Surveys

Since the 1979/80 survey, classification trees have been used for post-stratification with respect to a set of 11 estimated variables: Total expenditure of the household and the 10 major expenditure groups. The potential classifying variables were: size of household, size of locality, and head of household characteristics - age, continent of birth, last school attended, occupation and employment status (period of investigation was used at a separate stage of weighting). In previous surveys the weights were determined by post-stratification based on cross-classification, by size of household, size of locality and period of investigation.

At each stage, the values of  $R^2$  were calculated for each of the 11 estimated variables according to each of the classifying variables not yet used at that stage. Five different combined criteria were examined at each stage: the three mentioned in the previous section and also two weighted criteria. The weights were relative to the sampling variances of the estimated average expenditure per household in the total population, even though it would have been better to determine these weights relative to the residual sampling variance in the sub-group. At some stages there was one classifying variable that was best according to all criteria, while at others, judgment was required to decide which criterion to use.

To illustrate the process, the determination of the classifying variables for only a few of the groups in the first stages is demonstrated. At the first stage all the classifying variables took part and Table 7.2 presents the contribution of each classifying variable.

Table 7.2  
Determination of the First Classifying Variable for Total Population  
Israeli FES

| Average Expenditure per Household (Estimated variable) | Weight for criteria | Classifying variable |                  |             |             |                   |      |                    |
|--|---------------------|----------------------|------------------|-------------|-------------|-------------------|------|--------------------|
|  |                     | Size of H.H          | Size of locality | Last school | Occupation  | Employment status | Age  | Continent of birth |
| Total expenditure                                      | 1                   | 14.3                 | 2.1              | 12.1        | <u>18.4</u> | 9.5               | 13.4 | 4.2                |
| Food (excl. Veggies. &Fruit)                           | 1                   | <u>41.1</u>          | 1.1              | 1.3         | 9.7         | 7.7               | 16.7 | 6.0                |
| Vegetables & Fruit                                     | 1                   | <u>32.5</u>          | 0.6              | 0.5         | 4.7         | 4.1               | 11.8 | 4.7                |
| Household maintenance                                  | 3                   | 10.8                 | 1.7              | <u>11.6</u> | 10.9        | 4.1               | 7.5  | 3.8                |
| Transport & Communication                              | 8                   | 4.7                  | 2.0              | <u>16.2</u> | 15.5        | 5.6               | 7.0  | 3.8                |
| Education, Culture & Ent.                              | 5                   | 9.0                  | 0.6              | 9.1         | <u>13.8</u> | 8.7               | 11.4 | 2.8                |
| Housing  | 2                   | 1.3                  | 4.0              | 5.0         | <u>6.2</u>  | 1.2               | 2.3  | 2.8                |
| Clothing & Footwear                                    | 6                   | <u>10.4</u>          | 0.4              | 1.7         | 4.7         | 3.8               | 6.3  | 1.4                |
| Furniture & HH Equipment                               | 10                  | 3.6                  | <u>8.6</u>       | 3.2         | 5.5         | 4.5               | 7.7  | 1.8                |
| Misc. Goods & Services                                 | 4                   | 4.3                  | 0.5              | 4.5         | <u>7.1</u>  | 4.3               | 6.0  | 1.8                |
| Health   | 5                   | 3.1                  | 1.7              | <u>4.5</u>  | 4.1         | 2.0               | 1.6  | 0.7                |
| a) Number of Max( $R^2$ )                              | 3                   | 1                    | 3                | <u>4</u>    | 0           | 0                 | 0    |                    |
| b) Simple Average of $R^2$                             | <u>12.3</u>         | 2.1                  | 6.3              | 9.1         | 5.0         | 8.3               | 3.1  |                    |
| c) Weighted Average of $R^2$                           | 7.3                 | 2.9                  | 6.9              | <u>8.8</u>  | 4.8         | 7.1               | 2.5  |                    |
| d) Simple Average of $R^2/\text{Max}(R^2)$             | 0.69                | 0.24                 | 0.58             | <u>0.75</u> | 0.39        | 0.59              | 0.23 |                    |
| e) Weighted Average of $R^2/\text{Max}(R^2)$           | 0.60                | 0.34                 | 0.63             | <u>0.79</u> | 0.44        | 0.64              | 0.22 |                    |

Size of household is the best classifying variable, only according to criterion (b) and it dominates for Food, Vegetables & Fruit, and Clothing & Footwear. Occupation is the best classifying variable according to all the other four criteria. Despite this, size of household was chosen as the first classifying variable, since this is the main sub-division by which the estimates are presented and also because of the different non-response rates in these groups. It can be seen that size of locality, which served as a classifying variable in previous surveys, contributed very little to the classification according to all the combined criteria, and for all the estimated variables, except one.

At the second stage, the classifying variables were determined within each one of the groups of household size. The selection of the second classifying variable is demonstrated here for only one group - size 1. As may be seen from the data of Table 7.3, last school is the best classifying variable by all criteria. It should be noted that this variable was also preferable for households of sizes 2 and 3, but for households of size 4 and more, occupation was the best classifying variable by all criteria.

Table 7.3  
Determination of the Second Classifying Variable - Within Households of Size 1  
Israeli FES

| Average Expenditure per Household (Estimated variable) | Weight for criteria | Classifying variable |                  |             |             |                   |             |                    |
|--|---------------------|----------------------|------------------|-------------|-------------|-------------------|-------------|--------------------|
|  |                     | Size of H.H          | Size of locality | Last school | Occupation  | Employment status | Age         | Continent of birth |
| Total expenditure                                      | 1                   | -                    | 2.7              | <u>16.8</u> | 13.3        | 8.8               | 8.3         | 0.4                |
| Food (excl. Veggies. & Fruit)                          | 1                   | -                    | <u>6.8</u>       | 2.2         | 4.8         | 5.3               | 4.7         | 5.7                |
| Vegetables & Fruit                                     | 1                   | -                    | 4.7              | <u>1.1</u>  | 2.0         | 2.9               | 7.2         | <u>14.0</u>        |
| Household maintenance                                  | 3                   | -                    | 1.1              | <u>4.7</u>  | 3.9         | 0.2               | 3.5         | 2.3                |
| Transport & Communication                              | 8                   | -                    | 0.5              | <u>15.2</u> | 11.4        | 8.4               | 13.1        | 0.5                |
| Education, Culture & Ent.                              | 5                   | -                    | 5.0              | 10.2        | <u>10.3</u> | 9.5               | 10.0        | 0.2                |
| Housing  | 2                   | -                    | 8.7              | <u>9.7</u>  | 4.7         | 1.3               | 1.0         | 4.2                |
| Clothing & Footwear                                    | 6                   | -                    | 0.8              | 6.7         | 3.7         | 1.7               | <u>7.8</u>  | 0.5                |
| Furniture & HH Equipment                               | 10                  | -                    | 5.7              | 9.5         | 11.3        | 8.3               | <u>11.5</u> | 0.5                |
| Misc. Goods & Services                                 | 4                   | -                    | 0.5              | 2.2         | <u>2.3</u>  | 1.5               | 1.5         | 0.1                |
| Health   | 5                   | -                    | <u>3.9</u>       | 2.1         | 3.0         | 0.9               | 1.2         | 1.4                |
| a) Number of Max( $R^2$ )                              | -                   | 2                    | <u>4</u>         | 2           | 0           | 2                 | 1           |                    |
| b) Simple Average of $R^2$                             | -                   | 3.7                  | <u>7.3</u>       | 6.4         | 4.4         | 6.3               | 2.7         |                    |
| c) Weighted Average of $R^2$                           | -                   | 3.2                  | <u>8.3</u>       | 7.5         | 5.2         | 7.8               | 1.2         |                    |
| d) Simple Average of $R^2/\text{Max}(R^2)$             | -                   | 0.45                 | <u>0.78</u>      | 0.72        | 0.45        | 0.67              | 0.30        |                    |
| e) Weighted Average of $R^2/\text{Max}(R^2)$           | -                   | 0.39                 | <u>0.85</u>      | 0.80        | 0.50        | 0.78              | 0.16        |                    |

At subsequent stages, further classification was performed by using the same procedure, although with less classifying variables. For example, at the third stage, for households of size 1 whose heads had the lowest level of education, the best classifying variable by all criteria was continent of birth, whereas for those whose heads of household had the highest education the best classifying variable was age.

The resulting accumulated  $R^2$  at every stage for each of the 11 estimated variables are presented in Table 7.4.

**Table 7.4**  
**Accumulated  $R^2$  by Classification Stages**  
**Israeli FES**

| Average expenditure<br>per household<br>(Estimated variable) | Accumulated $R^2$ |         |         |         |
|--|-------------------|---------|---------|---------|
|  | Stage 1           | Stage 2 | Stage 3 | Stage 4 |
| Total expenditure  | 0.143             | 0.282   | 0.391   | 0.409   |
| Food (excl. Veggies. & Fruit)                                | 0.411             | 0.429   | 0.479   | 0.491   |
| Vegetables & Fruit   | 0.325             | 0.345   | 0.394   | 0.408   |
| Household maintenance  | 0.108             | 0.225   | 0.288   | 0.308   |
| Transport & Communication                                    | 0.047             | 0.199   | 0.263   | 0.289   |
| Education, Culture & Ent.                                    | 0.090             | 0.186   | 0.257   | 0.270   |
| Housing  | 0.013             | 0.072   | 0.190   | 0.199   |
| Clothing & Footwear  | 0.104             | 0.128   | 0.178   | 0.193   |
| Furniture & HH Equipment                                     | 0.036             | 0.080   | 0.167   | 0.190   |
| Misc. Goods & Services                                       | 0.043             | 0.089   | 0.138   | 0.155   |
| Health   | 0.031             | 0.061   | 0.117   | 0.148   |
| Number of Sub-groups   | 6                 | 26      | 80      | 117     |

The  $R^2$  achieved by this classification is relatively high for estimates of the average Total expenditure, for Food and for Vegetables & fruit (between 40% and 50%), but it is lower for all the other estimates (between 15% to 30%).

The first three stages contribute most of the homogeneity and the contribution of the last stage is minimal even though the number of groups noticeably increased. The marginal contribution of each stage is not similar for the various expenditure groups. For Food, Vegetables & Fruit, the first stage contributes most of  $R^2$ , whereas in other groups, the marginal contribution of the last stages is relatively high. For the 1992/93 FES, a revised tree with a larger set of potential classifying variables is planned to be constructed, so as to obtain a better classification. Also, the efficiency of using the CART software with a synthetic variable will be studied.

## 7.5 Efficiency of Using Classification Trees

Usually, the tree is constructed by using data from a certain survey and is applied for post-stratification in a subsequent survey. If the sample of the later survey would have been initially stratified by the same post-strata of the tree, the expected reduction in sampling variance would have been by the factor  $1-R^2$ , providing no changes had taken place in the population with regard to the estimated variable. This would also be the case with a simple sample with post-stratification by the same post-strata if their sample sizes are sufficiently large, if the response rates are similar in the post-strata and if the external estimates are known. However, not all these conditions are fulfilled in practice, thus the actual empirical efficiency of post-stratification by classification trees has to be examined.

a) For the 1988 Income Survey empirical results were studied by comparing three methods:

- A: Self-weighting (one group).
- B: Post-stratification by cross-classification (100 groups) - the old method.
- C: Post-stratification by classification tree (100 groups) - the new method.

Table 7.5 presents a comparison of the relative sampling variance of the estimates of net income per household from the three methods.

Table 7.5  
Ratios of Relative Sampling Variances Between Weighting Methods  
Estimates of Net Income per Household - Israeli 1988 Income Survey

| Sub-group                                  | Ratio Between Relative Sampling Variances |      |      |                        |      |       |                   |      |
|--|---|------|------|------------------------|------|-------|-------------------|------|
|  | With known weights                        |      |      | With estimated weights |      |       | $\epsilon$ effect |      |
|  | B/A                                       | C/A  | C/B  | B'/A                   | C'/A | C'/B' | B'/B              | C'/C |
| Total                                      | 0.85                                      | 0.55 | 0.65 | 0.92                   | 0.69 | 0.75  | 1.08              | 1.25 |
| <u>Employment Status of Household Head</u> |   |      |      |                        |      |       |                   |      |
| Employee                                   | 0.92                                      | 0.62 | 0.67 | 0.96                   | 0.75 | 0.79  | 1.04              | 1.22 |
| Self employed                              | 1.05                                      | 0.92 | 0.87 | 1.05                   | 0.93 | 0.89  | 1.00              | 1.02 |
| Not working                                | 1.34                                      | 0.80 | 0.60 | 1.35                   | 0.82 | 0.61  | 1.01              | 1.03 |
| <u>Years of Study of Household Head</u>    |   |      |      |                        |      |       |                   |      |
| 0- 8                                       | 0.95                                      | 0.79 | 0.83 | 0.98                   | 0.94 | 0.95  | 1.04              | 1.19 |
| 9-12                                       | 1.04                                      | 0.68 | 0.65 | 1.06                   | 0.74 | 0.70  | 1.02              | 1.10 |
| 13+  | 0.90                                      | 0.62 | 0.69 | 0.93                   | 0.74 | 0.80  | 1.04              | 1.20 |
| <u>Size of Household</u>                   |   |      |      |                        |      |       |                   |      |
| 1  | 1.09                                      | 1.02 | 0.94 | 1.10                   | 1.08 | 0.98  | 1.01              | 1.06 |
| 2  | 0.99                                      | 0.94 | 0.95 | 1.01                   | 1.00 | 0.99  | 1.02              | 1.06 |
| 3  | 1.02                                      | 0.96 | 0.93 | 1.03                   | 0.99 | 0.96  | 1.01              | 1.04 |
| 4  | 1.09                                      | 0.71 | 0.65 | 1.09                   | 0.74 | 0.67  | 1.00              | 1.03 |
| 5  | 1.02                                      | 0.89 | 0.88 | 1.02                   | 0.95 | 0.93  | 1.01              | 1.07 |
| 6+   | 1.03                                      | 0.97 | 0.95 | 1.03                   | 1.02 | 0.99  | 1.01              | 1.05 |

It can be seen that if the  $W_h$  were known (methods B and C), then for the population as a whole, the relative sampling variance in the old method is reduced to 0.85 compared with 0.55 in the new method. Thus, by the new method there is a gain of 0.65 against the old method. The reduction of the sampling variance by the new method (C) is smaller than the initial value 0.43 of  $1-R^2$ , mainly because of the  $\delta$  and D effects. Since, however,  $W_h$  are estimated (methods B' and C') from the Labour Force Survey, an additional factor,  $\epsilon$ , contributes more to the increase of the variance in the new method than in the old one, because  $S^2_b$  is larger. For the overall population,  $\epsilon$  in the new method, increases the relative sampling variance by 1.25 as against 1.08 in the old method. Thus, in practice, the real gain in efficiency achieved with the new method is only 0.75.

For estimates within sub-groups, the old method reduces only slightly the sampling variance and sometimes may even increase it as a result of differential rates of non-response. With the new method, sampling variances are reduced but, in general, by a smaller amount than in the overall population. As expected, in both methods  $\epsilon$  is smaller, the smaller the groups are, and conversely, larger the larger is  $S^2_b$  within the group.

The estimates themselves obtained from the three methods were also compared both for the estimated variable used to determine the classification (net income per household) and for two other variables (gross income per household and net income per person), and are presented in Table 7.6.

Table 7.6  
Comparison between Estimates by and Ratio of M.S.E. by Weighting Methods (Percentages)  
Israeli 1988 Income Survey

| Sub-group                                  | Comparison between Estimates |          |                            |          |                          |          | Ratio of M.S.E. of Net income per household |      |      |
|--|------------------------------|----------|----------------------------|----------|--------------------------|----------|---|------|------|
|  | Net income per person        |          | Gross income per household |          | Net income per household |          |   |      |      |
|  | (C'/A)-1                     | (B'/A)-1 | (C'/A)-1                   | (B'/A)-1 | (C'/A)-1                 | (B'/A)-1 | B'/A  | C'/A | C'/B |
| Total                                      | 2.0                          | -0.2     | 6.4                        | 3.6      | 5.0                      | 3.0      | 0.18  | 0.03 | 0.14 |
| <u>Employment Status of Household Head</u> |                              |          |                            |          |                          |          |   |      |      |
| Employee                                   | 1.9                          | -0.1     | 2.2                        | 1.0      | 1.9                      | 0.8      | 0.46  | 0.14 | 0.31 |
| Self employed                              | 0.2                          | -1.0     | 1.6                        | 0.7      | 1.7                      | 0.6      | 0.88  | 0.69 | 0.78 |
| Not working                                | -2.9                         | -3.5     | 0.1                        | 1.0      | -0.2                     | 0.9      | 1.50  | 0.82 | 0.55 |
| <u>Years of Study of Household Head</u>    |                              |          |                            |          |                          |          |   |      |      |
| 0- 8                                       | 1.6                          | -1.1     | 7.1                        | 4.4      | 5.7                      | 3.7      | 0.20  | 0.09 | 0.47 |
| 9-12                                       | 0.4                          | -0.3     | 4.6                        | 2.5      | 3.4                      | 2.0      | 0.31  | 0.12 | 0.38 |
| 13+  | 2.6                          | -0.9     | 5.0                        | 1.8      | 3.9                      | 1.4      | 0.46  | 0.10 | 0.21 |
| <u>Size of Household</u>                   |                              |          |                            |          |                          |          |   |      |      |
| 1  | 3.5                          | 2.7      | 5.4                        | 3.7      | 3.5                      | 2.7      | 0.43  | 0.39 | 0.90 |
| 2  | 4.2                          | 2.2      | 5.9                        | 3.1      | 4.2                      | 2.2      | 0.34  | 0.15 | 0.44 |
| 3  | 3.9                          | 0.9      | 5.1                        | 1.2      | 3.9                      | 0.9      | 0.68  | 0.23 | 0.34 |
| 4  | 2.2                          | 0.7      | 3.1                        | 0.9      | 2.2                      | 0.7      | 0.78  | 0.37 | 0.48 |
| 5  | 1.3                          | 0.2      | 2.0                        | 0.4      | 1.3                      | 0.2      | 0.93  | 0.65 | 0.70 |
| 6+   | 2.9                          | 0.0      | 3.6                        | 0.4      | 2.8                      | 0.2      | 0.91  | 0.38 | 0.42 |

It can be seen that on the whole the estimates from the new method (C') are higher than from the old method (B') and both are usually higher than estimates obtained without the use of weighting (A). The exception to this is in households whose heads do not work, where the opposite is true.

If it can be assumed that estimates in the new method are less biased, then the reduction in bias by the new method is considerable when compared to the old method. This is the case both for the overall population and for most of the sub-groups. However, it is not clear whether the bias in each of the groups is corrected to the same extent, since the efficiency of the classification varies from group to group, a point which needs further examination.

The relative M.S.E. of the estimates of net income per household by the three methods, under the assumption that by method C the estimates are not biased, are also presented in Table 7.6. Since the relative biases are higher than the relative sampling errors, the M.S.E. in the new method is much smaller than in the old method, and, of course, than that of the simple estimate. Thus, the new method is very efficient for the overall population but less so for small sub-groups (such as the self-employed).

b) For the 1979/80 Family Expenditure Survey the difference between the two estimates obtained by the old weighting procedure and the new one, relative to the estimate by the new method, were calculated. Table 7.8 presents these ratios for estimates of the average expenditure per household for the total population and for size of household. These ratios may indicate a reduction of biases (due to non-response) by the new method because it is expected to be less-biased.

Table 7.7  
Reduction of Bias by the New Weighting Relative to the Old Weighting by Size of Household  
Israeli 1979/80 FES  
(Percentage)

| Expenditure Group                              | Household size |       |       |       |       |       |       |
|--|----------------|-------|-------|-------|-------|-------|-------|
|  | Total          | 1     | 2     | 3     | 4     | 5     | 6+    |
| Estimates of Average Expenditure per Household |                |       |       |       |       |       |       |
| Total expenditure                              | + 4.7          | + 9.2 | + 7.9 | + 2.2 | + 4.1 | + 3.7 | + 0.7 |
| Vegetables & Fruit                             | + 1.0          | - 0.2 | + 2.4 | + 0.4 | + 2.1 | - 0.4 | - 0.6 |
| Food (excl. Veggies. &Fruit)                   | + 1.6          | - 0.6 | + 1.0 | + 0.6 | + 0.5 | - 0.1 | - 1.3 |
| Furniture & HH Equipment                       | + 2.2          | - 3.2 | + 8.8 | - 1.3 | + 2.9 | + 2.5 | - 4.0 |
| Clothing & Footwear                            | + 3.6          | + 1.8 | +10.8 | + 0.1 | + 4.0 | + 2.0 | + 1.0 |
| Housing  | + 4.9          | + 8.1 | + 8.1 | - 1.0 | + 4.9 | + 3.5 | + 5.2 |
| Household maintenance                          | + 5.0          | + 2.5 | + 7.4 | + 3.9 | + 5.4 | + 6.2 | + 1.7 |
| Health   | + 5.4          | +12.4 | + 6.1 | + 7.5 | + 3.2 | + 2.5 | + 1.7 |
| Misc. Goods & Services                         | + 6.6          | +20.4 | +11.8 | + 1.0 | + 5.1 | + 1.5 | + 3.7 |
| Education, Culture & Ent.                      | + 8.3          | +18.1 | +14.7 | + 8.3 | + 4.3 | + 6.8 | + 1.2 |
| Transport & Communication                      | + 9.3          | +28.9 | +11.3 | + 5.8 | + 7.5 | + 9.7 | - 1.1 |
| Estimates of Expenditure Weights               |                |       |       |       |       |       |       |
| Total expenditure                              | 0.0            | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   |
| Vegetables & Fruit                             | - 3.5          | - 8.6 | - 5.1 | - 1.8 | - 1.9 | - 4.0 | - 1.3 |
| Food (excl. Veggies. &Fruit)                   | - 3.0          | - 9.0 | - 6.4 | - 1.6 | - 3.5 | - 3.7 | - 2.0 |
| Furniture & HH Equipment                       | - 2.4          | -11.4 | + 0.8 | - 3.4 | - 1.2 | - 1.2 | - 4.7 |
| Clothing & Footwear                            | - 1.1          | - 6.8 | + 2.7 | - 2.1 | - 0.1 | - 1.6 | + 0.3 |
| Housing  | + 0.2          | - 1.0 | + 0.2 | - 3.1 | + 0.8 | - 0.2 | + 4.5 |
| Household maintenance                          | + 0.3          | - 6.1 | - 0.5 | + 1.7 | + 1.2 | + 2.4 | + 1.0 |
| Health   | + 0.7          | + 2.9 | - 1.7 | + 5.2 | - 0.9 | - 1.2 | + 1.0 |
| Misc. Goods & Services                         | + 1.8          | +10.3 | + 3.6 | - 1.2 | + 1.0 | - 2.1 | + 3.0 |
| Education, Culture & Ent.                      | + 3.4          | + 8.2 | + 6.3 | + 6.0 | + 0.2 | + 3.0 | + 0.5 |
| Transport & Communication                      | + 4.4          | +18.0 | + 3.2 | + 3.5 | + 3.3 | + 5.8 | - 1.8 |

It can be seen that by the old method the estimate of Total expenditure was 4.7% higher than by the new method, but biases differ for various sizes of household. For example, for size 1, the relative bias is +9.2% and for size 3 it is only +2.2%. For different groups of expenditure, different relative biases are obtained, these being particularly high for Transport & Communication (+9.3%) and for Education & Culture (+8.3%), and relatively small for Food (+1.6%) and for Vegetable & Fruit (+1.0%).

Thus, the estimated weights for the Consumer Price Index are biased too, as shown in Table 7.7. For the whole population, it can be clearly seen that the weight of expenditure on Food is downward biased (3.0%) and the weight of expenditure on Education & Culture; Transport & Communication are upward biased (3.4% and 4.4%, respectively).

The reduction of biases by using the classification tree for the total population is influenced by the entire tree, including, of course, the first classification by size of household. An examination of the biases of the weights within every size of household reveals the effects of using the classification tree within each of the groups by size of household. Specially high relative biases of these weights were found for size 1 households, where for Food the estimate is downward biased by about 9.0%; Furniture & Household equipment downward by about 11%, and Clothing & Footwear downward by about 7%. The weight of expenditure for Transport & Communication are biased upward by about 18% and for Education & Culture by about 8%.

To summarize, these empirical results (and additional findings not presented here), show that there is no doubt that a classification tree gives a better classification into homogeneous groups than cross-classification and thus the efficiency of post-stratification is increased. It is also reasonable to assert that by using the classification tree, a reduction of the bias arising from non-response is achieved. However, it is not clear whether the correction of biases is the same in each of the sub-groups and so it is unclear whether estimates of differences between groups are less biased. This is a problem common to most weighting methods.

## 7.6 External Weights of the Households

For persons, current reliable external estimates are usually available for cross-classification by basic characteristics, such as sex, age, religion or ethnic group, type or size of locality, etc. However, usually, no current data of households' characteristics are available. The current Labour Force Survey (LFS), or a similar current household survey, is often the only source of such estimates. The LFS estimates can be used for weighting in other household surveys that are based on samples smaller than that of the LFS and, often having a considerably higher rate of non-response.

LFS's in many countries are carried out as panel surveys, so that each household participates in the survey for a number of rounds. This is also the case in Israel, where the LFS is conducted continuously on a quarterly panel basis, with each dwelling unit investigated for two consecutive quarters and, after a break of two quarters, for two additional consecutive quarters. The weighting procedure of the LFS for individuals is based primarily on ratio-estimates to ensure consistency with independent current estimates of the distribution of the population by sex-age-religion-region groups. Thus, in general, different weights are assigned to different members of the same household.

The estimates of the number of households and their characteristics are obtained in Israel at present by assigning to the households the weight of their head as was determined by weighting for individuals. There is definitive evidence that the estimates of households obtained by such weighting are biased, e.g., the estimated percentage of 1-person households is severely downward-biased. Neither the use of weights of other "representative" persons, nor the use of various combinations of the individual weights of members within the households, were satisfactory.

A different approach to this problem, not based on individual weights, has been used in Israel for the Family Expenditure Surveys and for some other surveys since 1980, and is now being adopted for use on a current basis. The stages of the procedure are outlined in the following:

- (a) Preliminary adjustments of the initial weights are made to correct for certain types of undercoverage in the survey sample (such as new construction and temporary housing units of new immigrants).

- (b) Imputations are made for "temporary" non-respondents, i.e., non-respondents in one or more rounds, who respond at least once during the four rounds. The known data, for a selection of variables suitable for imputation of this kind, from the nearest quarter are imputed to the missing quarter, after adjustments, if required, for the time-lag between the quarters.
- (c) Adjustments are made to compensate for "permanent" non-respondents, i.e., households that did not respond in any round. There are several alternatives: assigning a uniform adjusting factor to either all respondents or only to the "temporary" non-respondents; or, assigning different factors to different groups of "temporary" non-respondents (according to reasons for non-response) to compensate for the corresponding groups of "permanent" non-respondents.
- (d) Special treatment is applied for panels not yet investigated in all the four rounds, especially for the panel investigated only once.
- (e) Controlling of the external distribution of the persons by sex-age-religion-region groups (see section 7.7).

Even by this procedure, the estimates for households may still be somewhat biased. An alternative approach is to use the data from the latest population census in conjunction with estimates of changes over time from LFS's. Since it is reasonable to believe that the improved LFS's estimates are less biased than other estimates in estimating changes over time, better external estimates for post-stratification can be obtained.

## 7.7 Controlling Person Distribution with Household Weighting

Weighting procedures based on post-stratification (or its variants), that control the households' distribution, do not necessarily control the auxiliary distribution of individuals by key characteristics, e.g., sex-age groups. The survey estimates relating to persons may not agree with the controls even when the external estimates of households used for weighting households are unbiased. Besides an increase in the variance, this may also result in bias both in the survey estimates relating to households (since a correlation between the expenditure on various groups of items and the household composition according to the characteristics of its members is expected) and in the survey estimates relating to specific groups of persons.

Therefore, adjustment to ensure compatibility with the auxiliary distribution of persons has to be considered as an additional stage in the weighting procedure of households, so as to avoid possible biases in the survey estimates from this source.

Adjustments of this kind have been used in the Israeli 1979/80 and 1986/87 FES's (as well as with some other household surveys). These were made after the weights of the households were determined by a multi-stage procedure, i.e., post-stratification (described in previous sections), in conjunction with raking procedures for other adjustments (for example, to correct for the uneven distribution of the sample in time).

Usually, a reliable auxiliary distribution of the households by their members' age-sex groups is not available. Therefore, the adjustment to match controls of persons is done indirectly by means of an iterative process. In principle, profiles of households are defined by sex-age composition of their members and a series of correction factors is assigned to each household according to its profile, in several cycles until convergence is attained, as is explained in the following.

Let  $N_k$  be the number of persons in the population in group  $k$ ,  $k=1,2,\dots,G$ . These are the auxiliary data and they are usually reliable.

Let  $n_{ik}$  be the number of persons in household  $i$  who belong to group  $k$ , and  $n_i$  the total size of household  $i$ . Then, for each household  $i$ , a profile is defined by the vector  $\{n_1, n_2, \dots, n_G\}_i$ , and  $F$  different profiles ( $f=1, 2, \dots, F$ ) are possible.

The number of persons in the sample in all households of profile  $f$  that belong to group  $k$  is  $n_{kf}$  and the total number of persons in all households of profile  $f$  is  $n_f$ .

Let  $N_k^*$  be the estimated size of group  $k$  obtained by the weighting of households before the adjustment for persons is made.

If  $N_k^* \approx N_k$  for all  $k$ , no adjustment is required, but usually this is not the case. Among a number of alternatives, two adjusting procedures for persons distribution are outlined here: the first procedure does not give a complete fit to the auxiliary data of households, and the second, which is a raking-type procedure, does. The choice between these two procedures depends on the quality of the external distribution used for weighting the households. When it is believed that these data are reliable and unbiased, the second procedure is preferred.

In both procedures, sometimes, an overall preliminary adjusting factor has to be assigned to all profiles to correct for either possible overall undercoverage or overcoverage of the external estimates of households.

#### With the first procedure:

In the first round,  $A_k(1) = N_k/N_k^*$  are calculated for each  $k$ . Then, for each profile  $f$ , the first adjusting factor is calculated

$$\alpha_f(1) = \sum A_k(1)n_{kf}/n_f$$

and each household's weight is multiplied by this factor and new estimates of the number of persons in each group  $k$  are obtained,  $N_k^*(1)$ .

In each subsequent round,  $r$ , a set of adjusting factors is obtained by the same way:

$$\alpha_f(r) = \sum A_k(r)n_{kf}/n_f,$$

where,  $A_k(r) = N_k/N_k^*(r-1)$  and  $N_k^*(r-1)$  is the estimate obtained in the previous round.

#### With the second procedure:

In each round, there are two steps: in the first, the  $\alpha_f(r)$  are calculated, as in the first procedure, and in the second, adjusting factors,  $B_h(r)$ , are given to each household group  $h$ ,

$$\beta_h(r) = L_h / L_h^*(r),$$

where,  $L_h$  is the external number of households in group  $h$ , and  $L_h^*(r)$  is the new estimate of this number as obtained by the adjusted weights after the first step of round  $r$ .

Both procedures stop when a sufficient fit is achieved, or no further improvement in fit is obtained. But, when the relvariance of the weights becomes too large, the procedure can be stopped at earlier stages at the cost of less controlling for the person distribution, i.e., a trade-off has to be made between these two factors. The first procedure may also stop to avoid further deterioration of the fit to the external distribution of households.

It should be noted that sometimes a satisfactory adjustment may not be obtained, for example, with a too small sample, or when the external data for households are significantly biased, so that convergence is not always achieved for a reasonable increase in sampling variance.

These kinds of adjustment can be applied for the whole sample as a one group, or for several sub-groups of the population separately, which may be preferable, providing large enough samples are available for each sub-group.

It should be mentioned that both procedures can also be applied by using "binary-profiles" of households instead of "full-profiles", as above. With this, for each household  $i$ , indicator variables,  $\delta_{ik}$ , are defined for each group  $k$ , where,  $\delta_{ik}=1$  if  $n_{ik} \geq 1$ , and  $\delta_{ik}=0$  otherwise. The vector  $\{\delta_1, \delta_2, \dots, \delta_B\}_i$  defines the "binary-profile" of household  $i$  and  $B$  different "binary-profiles" are possible ( $b=1, 2, \dots, B$ ) and  $B \leq F$ , since each of the "binary-profiles" may comprise several "full-profiles". The adjusting procedure works in a similar way as with "full-profiles", where  $n_{kb}$  and  $n_b$  substitute  $n_{kf}$  and  $n_f$ . Although the use of such grouped profiles is not as good as the use of full profiles, it is used when the sample is not sufficiently large.

## CHAPTER 8. CODING

For each non-quantitative variable that is investigated in a survey (or a census), a categorical variable with a pre-specified set of exhaustive and mutually exclusive categories has to be defined, and a code (usually a numerical one) for each category is determined. Only after all the verbal information of each non-quantitative variable has been converted to appropriate codes, can the data be processed.

### 8.1 Coding Techniques in General

Various coding techniques can be used, dependent on the type of the investigated variable and on the structure of the forms used for data collection, as in the following.

- (1) Pre-coding techniques - When closed-type questions are used for data collection, all the categories with all the possible answers are printed on the survey form together with their codes. The code that corresponds to the chosen answer is marked during the data collection. Pre-coded questions are widely used, but are feasible only for variables with a limited number of categories and for answers that can be phrased shortly without loss of clarity.

Another pre-coding technique is when for each category, a certain space is allocated in the questionnaire and is used for entering additional information. The code of each category is either pre-printed and fed into the computer together with the other numerical information, or, the location of the information on the category in the questionnaire determines the code and keying the code is unnecessary.

- (2) Manual coding - The codes of verbal information collected by open-type questions are determined by coders by using code-books, i.e., an alphabetically-arranged list of the possible verbal answers with their categories and codes, and a list of all codes in numerical order, with the possible answers for each code. The appropriate codes are entered in the schedules and are later inputted into the computer together with all the other numerical information.

Another manual technique is when the codes are keyed directly into the computer (instead of entering them first into the schedules and then feeding them into the computer), thus saving one stage in the survey process. However, this technique is suitable only for variables with a limited number of categories. Another version of this technique is when a short dictionary is displayed on the screen and the coder uses a cursor for choosing the appropriate code. This is done when the use of pre-coded questions is not suitable either because the number of different codes is not small enough, or because the structure of the survey form does not allow it.

- (3) Automatic coding - In automatic coding, first, a "dictionary", that takes the place of the code-book used for manual coding, is stored in the computer. Because human judgement cannot be used, it must be as comprehensive as possible, including all possible verbal information expected to appear in the survey schedules (unless very infrequent) and ambiguous descriptions must be avoided. Second, the verbal information collected in the survey has to be entered into the computer and then for each item description, a search is made in the dictionary. If a matching is achieved, the corresponding code is assigned, otherwise, a supplementary process is referred to, i.e., manual coding. The search algorithm and the matching rules must be adapted to the type of the coded variable and to the structure of the language.

In general, matching may be of two kinds: exact or inexact. With exact matching, the standardized description of the item must be identical with one of the dictionary entries and only then the associated code is assigned.

Inexact matching allows the associated code to be assigned even if there is only a partial similarity between the item description and one of the dictionary entries. The degree of similarity can be specified by using various sets of matching rules based on empirical probability patterns and on the structure of the language, such as ignoring word order, assigning heuristic weights to words according to each word's content, etc. (as in Appel and Hellerman, 1983). In general, exact matching is expected to produce less coding errors than inexact matching, but leads to a higher rate of uncoded descriptions. However, this is not always the case with variables of a simple structure, like goods and services in FES's, where a high rate of exact matching can be achieved.

Automatic coding can be performed in two main ways:

- (a) Computer-assisted coding - The verbal information of one item at a time is entered into the computer. Then, all descriptions in the dictionary similar to that item appear on the screen with their codes. The coder chooses from the ranked alternatives the code which seems to him the most appropriate and enters it into the computer by keying or by using a cursor. Sometimes, the coder has access to additional information related to the coded item that can be used for further guidance and assistance from the computer. A version of this technique is when the coder enters only the first characters of the first word (say the first 3 letters) and then a subset of the dictionary entries is displayed on the screen for the choice of the coder (Embry, 1988).
- (b) Fully automated coding - The verbal information of all the schedules of a certain coding batch are keyed consecutively into the computer. Then, the file that includes all the verbal information of all the items is matched with the dictionary via a special search algorithm adapted specifically for this operation. Then, codes are assigned automatically for those entries that were matched. A file of entries which were not coded automatically is referred to manual coding, or first, to computer-assisted coding and then, if necessary, to manual coding.

A good automatic coding procedure is one that minimizes the rate of erroneous coding, the number of entries requiring manual coding and the computer search-time per entry. Clearly, there is a conflict between these targets, i.e., minimizing the rate of erroneous coding may result in an increase in the amount of manual coding and vice versa, and reduction of the computer search-time, for example, by reducing the number of referred items in the dictionary, may result in an increase in the rate of errors. Thus, a trade-off must be made between these different targets for choosing the most efficient method.

A good automatic coding process also ensures flexibility to incorporate information accumulated in the course of the coding. Thus, continuing evaluation of the process is required for updating the dictionary and for improving the process.

With both manual and automatic coding, checking of the coding and the keying operations is needed and can be based on standard quality control sampling procedures. However, unlike control systems used for industrial manufacture, or for keying, there is the additional problem of determining the true code. Thus, special methods have been developed for controlling the coding operation.

There are two main verification procedures: a dependent and an independent verification. In dependent verification, a verifier inspects the code assigned by the coder and decides whether it is correct or not. In independent verification, the verifier assigns codes without seeing the initial codes. The initial codes are then matched (by computer) with the controlled codes, and any discrepancy is judged by an expert who decides which is the correct code, or even chooses another code. The dependent verifier tends to be influenced by the initial code and thus the rate of incorrect codes remaining undetected can be substantial.

Independent verification has a much higher rate of detection of erroneous codes, but is more expensive and more difficult to implement than dependent verification. With any type of verification more errors would be discovered if repeated many times, but because of limited resources, usually only one verification is performed.

Each of the coding techniques has its advantages and disadvantages.

Pre-coding techniques avoid the process of coding and thus are free of coding errors, but their use is limited to certain variables and they can be used only with a certain type of structure of the survey schedule.

Manual coding is usually subject to a high rate of erroneous coding and keying, thus it must be checked thoroughly by quality control procedures. Furthermore, manual coding is both costly and time-consuming, besides which, the recruiting, training and managing of a large number of coders, impose a burden on management.

The primary advantage of automatic coding lies in its potential to provide a better quality of coding than with manual coding. Furthermore, this can be attained with reduced costs and within a shorter time. Automatic coding saves most of the clerical work of manual coding, but not all, since automatic coding must usually be followed by some manual coding, except for certain variables for which full automatic coding can be attained. On the other hand, keying textual responses required in automatic coding is more expensive than keying only codes in the manual process. Automatic coding reduces also the managerial burden, since keyers are readily available, need little special training and their work is much easier to check. Automatic coding is not free of errors and verification cannot be avoided, but this is not as costly and complicated as the verification needed for manual coding.

Computer-assisted coding is sometimes preferable to fully-automated coding, because it combines the intelligence of the human mind with the capacity of the computer. But, whereas fully-automated coding allows for untrained coders, computer-assisted coding needs coder training.

## 8.2 Coding Goods and Services in FES's

In FES's, several coding techniques are used for the different non-quantitative variables. Pre-coded questions are used, for example, for the characteristics of the household members, providing they comprise a limited number of categories (Sex, Last school attended, etc.) and the technique of pre-coded categories is used often for Expenditure items of the questionnaire. The other verbal information is either coded manually or automatically.

In Israel, some variables are coded manually and their codes are written first in the survey forms and then fed into the computer (such as Occupation and Industry), while others are coded manually but the codes are keyed directly into the computer (such as Type of store from the diaries). The coding of goods and services of the diaries is performed automatically in Israel and in a few other countries, but in most countries manual coding is still used. In the following we will focus on the coding of this verbal information, since this is specific to FES's.

### (a) Scope and Difficulties

Although the samples of FES's comprises usually only a few thousands households (e.g., 5,000 in the Israeli 1986/87 survey, 6,000 in the Swedish 1985 survey, about 7,000 per year in the U.S.A. current survey), the coding of goods and services is on a much larger scale than the coding of information in other sample surveys. This is because the average number of items recorded in the diary per household is quite large. For example, about 750,000 diary items had to be coded in the Israeli 1986/87 FES, since an average of about 150 items were recorded in each of the two-week diaries.

Not only is the coding of these items a very large undertaking, but also other difficulties are encountered. Since the majority of the recordings in the diaries are made by the respondents themselves (except for the rare cases where they are made by the interviewer), the coders or the keyers have to contend with thousands of different handwritings. Furthermore, coding is even more difficult in countries where the population comprises groups of different origins, like new immigrants, or in multilingual countries, especially when languages with different characters are used, like in Israel, where Hebrew, Arabic and recently, also Russian, are allowed for recording in the diaries. Sometimes the goods are described by the respondents only by their commercial names without further details needed for classification and this can be more frequent in developed countries where there is a wide variety of goods of the same type. Although the interviewers are instructed to check the recordings in the diaries, many such deficient descriptions remain.

(b) The Classification System

All items in FES's relating to financial transactions are classified by one comprehensive hierarchical system in order to simplify aggregation of the more detailed groups. In general, the hierarchical classification system comprises four primary classes: (1) Income; (2) Non-consumption expenditures; (3) Consumption expenditures; and (4) Savings. Within these four classes, major groups are defined and within them, sub-groups that are further split into the ultimate groups (comprising a specific item or a group of items).

As an example, the classification system used in the Israeli FES is based on a hierarchical 5-digit code (plus a control digit for reducing keying errors when manual coding is required). The number of classes are presented in Table 8.1. In some countries the number of codes is much larger, such as in The Netherlands (about 3,000) and in Denmark (about 1,500), whereas, other countries use a lot less, like in the United Kingdom (about 400) and in France (about 350) (Statistical Office of the European Communities, 1981).

Table 8.1  
Number of Classes in the Classification System in the Israeli 1986/87 FES

| Primary class<br>(1st digit) | Major groups<br>(2nd digit) | Sub-groups<br>(3rd digit) | Ultimate groups<br>(4th & 5th digits) |
|------------------------------|-----------------------------|---------------------------|---------------------------------------|
| Income                       | 4                           | 15                        | 41                                    |
| Non-consumption expenditure  | 2                           | 4                         | 6                                     |
| Consumption expenditure      | 10                          | 70                        | 686                                   |
| Savings                      | 3                           | 11                        | 14                                    |
| All primary classes          | 19                          | 100                       | 747                                   |

Although attempts are made to decide on consistent classification rules, this is not easy to achieve. Different countries may use different classification systems with different numbers of ultimate classes, adapted to their own needs and to their specific consuming habits. Even in one particular survey, the classification rules are not always consistent. This is because FES's are multi-purpose and it is not always possible to construct one hierarchical classification system suitable for all the survey uses without there being many exceptions.

One general rule that could be applied, as an example, would be to classify a commodity according to its use. According to this rule, all food products are classified under one major expenditure group, although sometimes, vegetables and fruit are considered as a separate major group.

Within these major groups, food products are classified according to their general nature, to their nutritional value, to their method of preparation and according to some other specific considerations. This is not always practical, for example, expenditure on "food eaten away from home" cannot be split among each item consumed and thus is classified according to place of eating (restaurant, kiosk, cafeteria at work, etc.) and by type of meal (sandwich, full meal, etc.).

If the rule of classifying a commodity according to its use would have been applied, a television set, for example, would be classified under "entertainment" or "education", a refrigerator and dishes under "food", a washing machine under "clothing", or "bedding and towels", and so on. Yet, because of considerations of the survey uses this principle cannot be fully applied and so, most electrical equipment (although not all) are usually classified under "electrical equipment" while dishes are under "household appliances".

Another example of the difficulty in classification is when raw materials are acquired from one source and the work is done by another. The overall expenditure of the ultimate product can be classified according to its use, or, be split by its components, i.e., the expenditure is divided between raw materials and labour. For example, when cloth is supplied to a tailor, the expenditure for the cloth is included under "yarns" and the payment to the tailor under "sewing and tailoring". However, if a suit is made to measure, all the expenditure is included under "suits".

An example of inconsistency is when "repairs", "insurance", "licenses", etc. for any kind of equipment are classified under the same class as the equipment itself, whereas "insurance of a dwelling" is classified in one separate class under "housing", and is not split between the different items that this insurance includes.

There may also be different opinions on how to classify various items. An example is that of expenditure on different kinds of sport activities: the question is whether they should be classified under "health" or as part of "education, culture and entertainment". Also, problems arise in classifying expenditure on "travel abroad", whether to classify it under "transport" or "entertainment", or, to classify each part of the expenditure separately under "food", "transport", "accommodation", etc.

As for "gifts" (except cash transfers), one rule may be to classify gifts given to other households as gifts (under non-consumption expenditures) and to classify gifts received from other households according to their type as consumption in kind. Another rule is to classify gifts given to other households as consumption expenditure according to the type of commodity.

It is no wonder that it is impossible to obtain a unique classification system strictly following simple rules. Because of this, correct codes are not easy to determine and thus the coding task is even more complicated.

### 8.3 Evaluation of Manual Coding

Although any evaluation study of coding errors, like with most evaluation studies of response errors, is subject to the difficulty of deciding upon the true values (even if determined by experts who made special efforts to find out the true code), evaluation studies can supply useful results. The manual coding (after verification) of the items in the monthly diaries in the Israeli 1979/80 FES was evaluated (Kantorowitz, 1981). This was done, first, to estimate the effects of erroneous coding on the accuracy of the survey estimates, and second, to study the rates of erroneous coding by type and source of error in order to improve coding in future surveys.

A controlled sub-sample of 120 households was selected, balanced for size of household, education of head of household and month of inquiry. For each of the 27,200 items included in the sub-sample, a "true" code was defined. Two experts checked independently all original codes, recording each case found to be in error. Any disagreement between the two experts was checked again. Extra effort was made to determine "true" codes, for example, individual prices of certain goods were used when the description in the diary was insufficient; additional information was sought about items described only by their commercial names; information in the questionnaires was utilized, etc. For each error found, also the type and source of error were coded. Some of the results of this study are presented in the following.

(a) Coding Errors by Type and Source of Error

Out of all codes, 8.2% were erroneous and their distribution by type and source is presented in Table 8.2. As can be expected, coding errors are more frequent the more detailed is the required classification. Indeed, out of all errors, 70.7% were between the ultimate groups within sub-groups (type A); 14.8% were between sub-groups within major groups (type B); 11.4% were between major groups of expenditures (type C); 2.4% were between consumption and non-consumption groups (type D); and, 0.7% were within non-consumption groups (type E).

Table 8.2  
Coding Errors, by Type and Source of Error (percentages)  
Israeli 1979/80 FES

| Source of Error       | Type of error |       |       |       |       |
|-----------------------|---------------|-------|-------|-------|-------|
|                       | Total         | A     | B     | C     | D & E |
| TOTAL                 | 100.0         | 70.7  | 14.8  | 11.4  | 3.1   |
| (1) Code-book         | 100.0         | 75.8  | 12.6  | 11.2  | 0.4   |
| (2) Coder             | 100.0         | 62.8  | 18.2  | 14.8  | 4.2   |
| (3) Item description: | 100.0         | 76.7  | 12.2  | 7.3   | 3.8   |
| - Avoidable (*)       | 100.0         | 70.0  | 17.1  | 5.9   | 7.0   |
| - Not avoidable       | 100.0         | 78.9  | 10.6  | 7.8   | 2.7   |
| TOTAL                 | 100.0         | 100.0 | 100.0 | 100.0 | 100.0 |
| (1) Code-book         | 24.9          | 26.8  | 21.3  | 24.4  | 2.9   |
| (2) Coder             | 41.9          | 37.2  | 51.4  | 54.3  | 56.5  |
| (3) Item description: | 33.2          | 36.0  | 27.3  | 21.3  | 40.6  |
| - Avoidable (*)       | 8.4           | 8.3   | 9.7   | 4.4   | 18.9  |
| - Not avoidable       | 24.8          | 27.7  | 17.6  | 16.9  | 21.7  |

(\*) If using prices, or available data of the questionnaires.

Source: Kantorowitz (1981).

Code-book deficiencies caused 25% out of all errors, because categories were either not mutually exclusive or not all-inclusive, i.e., when two different items with similar names were not sufficiently differentiated, or when no specific code was assigned in the dictionary for an item. Most of these errors involved only a limited number of items and so avoiding them would have been relatively simple. Errors due to coders' mistakes comprised 42% of all errors and these related to the coders' training and prior experience, and also to the efficiency of the quality control.

The rest, 33% were due to a deficient description of the item in the diary, thereof, 8% could have been avoided and correctly coded had additional effort been made using available prices or information from the household's questionnaire. About 40% of errors of types D and E stem from this source and about half of them could have been avoided by using further available information. A quarter of the other errors of this source could have been avoided if both the interviewers and the respondents would have been more aware of the need for more specific information.

Coding errors were not distributed proportionally over all groups of items. Error rates by type were estimated for different groups of items as a guide for improving future coding operations by concentrating on the most vulnerable groups. The results by major groups are presented in Table 8.3. For example, the group of Transport & Communication had a significantly high rate of erroneous coding (24%), mostly of type A (22%). Clothing & Footwear and Furniture & Household equipment were also subject to relatively high rates of error (about 18% each), but type A comprises less than two thirds for both groups and type B (5% and 2%, respectively). Vegetables & Fruit had comparatively low rates of error for all types (altogether 4%).

Table 8.3  
Coding Errors by Major Expenditure Groups (percentages)  
Israeli 1979/80 FES

| Major Expenditure Groups(*)         | Percentage of all codes | Percentage of erroneous codes | Rate of erroneous coding (**) |
|-------------------------------------|-------------------------|-------------------------------|-------------------------------|
| TOTAL                               | 100.0                   | 100.0                         | 8.1                           |
| Transport & Communication           | 5.7                     | 17.0                          | 24.1                          |
| Furniture & Household equipment     | 1.2                     | 2.7                           | 18.5                          |
| Clothing & Footwear                 | 2.2                     | 4.8                           | 17.7                          |
| Health                              | 1.0                     | 1.9                           | 13.9                          |
| Education, Culture & Entertainment  | 6.0                     | 6.5                           | 8.8                           |
| Household maintenance               | 3.8                     | 4.1                           | 8.7                           |
| Miscellaneous Goods & Services      | 4.8                     | 4.3                           | 7.3                           |
| Food (excluding Vegetables & Fruit) | 54.3                    | 47.3                          | 7.0                           |
| Vegetables & Fruit                  | 21.0                    | 11.4                          | 4.4                           |

\*) Not including Housing because only rarely coding was required.

\*\*) Excluding Type E.

#### (b) Effects of Coding Errors on the Survey Estimates

Coding errors can affect the survey estimates either by increasing their variance, or, if errors do not cancel each other, by causing bias. The relative bias,  $B(k)$ , of the diary estimate due to coding errors for each expenditure group  $k$ , was estimated by the percentage difference between the estimate based on the initial codes ( $I$ ) and that based on the "true" codes ( $T$ ), relative to the "true" estimate by  $B(k) = [X(k)_I/X(k)_T] - 1$ .

The survey estimate of each group is derived either from the diary (like Food), or from the questionnaire (like Electrical equipment), or from both sources. Thus the estimate of the relative bias of the survey estimate due to coding errors is  $B^*(k) = B^*(k)P^*(k)$ , where  $P^*(k)$  is the proportion of the contribution of the diary to the total estimate of each group  $k$ , and  $1-P^*(k)$  of the estimate is free of coding errors.

Table 8.4 presents the estimated relative biases,  $B(k)$  and  $B^*(k)$ , for major expenditure groups and for some selected sub-groups of items. Besides, the relative sampling errors,  $C(k)$ , of the same groups of items of the same survey, are also presented.

Thus,  $E(k) = C^2(k) / [C^2(k)+B^{*2}(k)]$  expresses the decrease in the effective size of the sample due to coding errors (ignoring other sources of error), or, to what extent field costs could be reduced for a given level of accuracy of the survey estimates, if no coding error existed.

Table 8.4  
Relative Bias, Relative Sampling Error and Decrease in Efficiency due to Coding Errors  
Major Expenditure Groups and Selected Sub-Groups  
Israeli 1979/80 FES

| Expenditure Groups                  | Percentage of expenditure from diary<br>$P(k)$ | Relative bias (%)  |                     | Relative sampling error<br>$C(k)$ | Relative efficiency<br>$E(k)$ |
|-------------------------------------|--|--------------------|---------------------|-----------------------------------|-------------------------------|
|                                     |  | of diary<br>$B(k)$ | overall<br>$B^*(k)$ |                                   |                               |
| TOTAL                               | 44.1   | -0.13              | -0.06               | 1.1                               | 1.00                          |
| <u>Major Groups(*):</u>             |  |                    |                     |                                   |                               |
| Food (excluding Vegetables & Fruit) | 100.0  | +0.30              | +0.30               | 1.0                               | .92                           |
| Vegetables & Fruit                  | 100.0  | +0.80              | +0.80               | 1.1                               | .65                           |
| Clothing & Footwear                 | 100.0  | -0.53              | -0.53               | 2.6                               | .96                           |
| Miscellaneous Goods & Services      | 42.8   | +3.63              | +1.55               | 2.5                               | .72                           |
| Transport & Communication           | 33.1   | +1.34              | +0.44               | 2.5                               | .97                           |
| Household maintenance               | 31.9   | -0.17              | -0.05               | 1.4                               | 1.00                          |
| Education, Culture & Entertainment  | 27.8   | +0.14              | +0.04               | 2.3                               | 1.00                          |
| Furniture & Household equipment     | 17.8   | -1.80              | -0.32               | 3.3                               | .99                           |
| Health                              | 8.9  | -0.73              | -0.06               | 2.5                               | 1.00                          |
| <u>Selected Sub-Groups:</u>         |  |                    |                     |                                   |                               |
| Bread, Cereals & Pastry products    | 100.0  | -1.01              | -1.01               | 1.3                               | .62                           |
| Vegetable oils & related products   | 100.0  | -0.07              | -0.07               | 2.9                               | 1.00                          |
| Meat & Poultry                      | 100.0  | +1.10              | +1.10               | 1.5                               | .65                           |
| Dairy products                      | 100.0  | +0.34              | +0.34               | 1.1                               | .91                           |
| Sugar & its products                | 100.0  | -2.47              | -2.47               | 2.0                               | .40                           |
| Meals away from home                | 100.0  | -6.48              | -6.48               | 4.9                               | .36                           |
| Fresh vegetables                    | 100.0  | +0.72              | +0.72               | 1.5                               | .81                           |
| Fruit                               | 100.0  | -0.13              | -0.13               | 1.5                               | .99                           |
| Household help                      | 100.0  | -2.39              | -2.39               | 6.9                               | .89                           |
| Sundry household items              | 100.0  | -0.01              | -0.01               | 2.4                               | 1.00                          |
| Home decorations                    | 40.9   | -9.96              | -4.07               | 6.5                               | .72                           |
| Men's outerwear                     | 100.0  | +14.97             | +14.97              | 7.6                               | .20                           |
| Women's outerwear                   | 100.0  | +0.34              | +0.34               | 4.9                               | 1.00                          |
| Children's outerwear                | 100.0  | +26.21             | +26.21              | 4.5                               | .03                           |
| Footwear                            | 100.0  | -1.16              | -1.16               | 3.8                               | .91                           |

\*) Not including Housing because only rarely coding was required.

As can be seen in Table 8.4, even when  $B^*(k)$  is quite small,  $E(k)$  is not negligible. For instance, although for Vegetables & Fruit  $B^*(k)=+0.8\%$  is small, yet  $E(k)=0.65$ ; for Food (excluding Vegetables & Fruit)  $B^*(k)=0.3\%$  and  $E(k)=0.92$ . For many sub-groups of items,  $E(k)$  was found to be even lower, e.g., for Sugar & its products  $E(k)=0.40$ .

The estimated relative bias for almost all major groups was found to be less than one percent, except for Miscellaneous Goods & Services, where  $B(k)=+3.6\%$  and  $B^*(k)=+1.6\%$ . As for sub-groups, the relative biases were usually higher: some of them suffered from a serious bias, such as Children's outerwear ( $B^*(k)=+26\%$ ), and Men's outerwear ( $B^*(k)=+15\%$ ). For some of the more detailed items (not presented in Table 8.4), bias was even higher, for example, it was as high as 4.3% for the estimates of expenditure on specific kinds of bread, while for the overall group of breads this bias was small (-0.5%), i.e., it was difficult to distinguish between very similar items.

These findings, despite their limitations, indicate how important it is to invest more effort in all means of improving coding accuracy: a better code-book, better training of the coders, comprehensive quality control, and developing automatic coding systems.

## 8.4 Automatic Coding

The first attempts to develop automatic coding, about 20 years ago, were at the U.S. Bureau of the Census (Corbett, 1972, and O'Reagan, 1972). Today, automatic coding is under experimentation in many countries and is already being used or, is planned to be used in the near future, in some national offices. In most statistical offices, the development of automatic coding originated with the complicated coding of Occupation and Industry, although in some countries, automatic coding was developed for the simple coding of geographical variables (Place of residence, Country of birth, Place of work, etc.), of Educational qualification, etc. Automatic coding of goods and services is not as complicated as the coding of Occupation and Industry, but it cannot be considered as simple as that of Country of birth, or Type of store.

In FES's, however, automatic coding for goods and services is in use in only a few countries. The first application of automatic coding of goods was in the 1978 Swedish FES, and since then it has been used for the 1985 and 1988 Swedish FES's, in Denmark, in The Netherlands and in Israel (Lyberg and Dean, 1992). Different versions of automatic coding of goods and services are used in these countries and the one that was first implemented in the Israeli 1986/87 FES (Berg and Har, 1990) is described in the following.

It should be mentioned, that the construction of the dictionary for the automatic coding and some of the rules in the coding algorithm were suited to the special features of the Hebrew language, which differs from English in its syntax and structure. For example, there is no separate word for "the" in Hebrew and instead a letter is attached to the beginning of a noun, thus reducing the number of words that describe an item; there is almost no use of vowels in Hebrew, therefore words are shorter; nouns precede adjectives, thus the search in the dictionary by the first word is in some cases enough for coding; and there is a distinction by gender and also for singular/plural forms, which increases the possibility of identifying an item.

### (a) Construction of the computer-stored dictionary

An initial file of verbal descriptions of goods and services with their codes was prepared, based on several sources: (1) the code-book used for manual coding in the previous survey after updating; (2) a sample of diaries with about 22,000 entries from the previous FES; (3) a sample of diaries from the pre-test of the 1986/87 FES with about 10,000 entries; and (4) lists of commodities with their commercial names from supermarket chains and big drug stores.

Each description in the dictionary was transformed to a standard form of three words at the most, each word with a maximum of eight characters. The appropriate code was added to each description, after a thorough verification. For some descriptions the code could be determined by the first word only, while for others, the first word was not sufficient and the second word was needed, or even the third one.

Thus, in order to make the coding process more efficient, numerical "indicators" were defined for each word of the dictionary description, so that it was possible to tell whether or not the first or the second word were sufficient for coding. For example, all 'WINE' entries had the same code. Therefore, the following entries, as written in Hebrew (where nouns precede the adjectives) had the same code: "WINE RED", "WINE RED SWEET"; "WINE DRY". So, for these entries the first word "WINE" determined the code, and this was indicated in the dictionary. Taking another example, one word was not sufficient for coding different kinds of "cheese". Two words were needed for all types of "WHITE CHEESE", thus "CHEESE WHITE FAT" and "CHEESE WHITE NONFAT" had the same code. But, different types of "YELLOW CHEESE" had different codes, so that the entry "CHEESE YELLOW HARD" was not coded the same as "CHEESE YELLOW SOFT", thus three words were needed for coding.

The dictionary included additional information for most of the items, such as the possible unit weights and possible type of stores. This was used for checking the validity of the assigned codes.

Items in the dictionary were deleted if there was any chance of error arising from ambiguity, unless such items could be checked in later stages of the process, by price, or weight, or type of store. For example, if a "car battery" was recorded in the diary as a "battery" only, this ambiguity could be detected later by its price, and thus was not deleted from the dictionary. Detecting ambiguities was done by listing all records in the dictionary in alphabetical order.

The dictionary was updated in the course of the coding process, especially during its first stages. Uncoded items occurring frequently were listed, examined and added to the dictionary, errors which were discovered were corrected, and ambiguities were deleted. The dictionary initially included about 2,500 items and, by the end of the process, this number increased to about 3,000. It should be mentioned that items appearing rarely in the diaries were excluded from the dictionary, so as to prevent the dictionary being overloaded. In addition, an initial "synonym-file" of about 1,000 words, that are often spelled incorrectly, was prepared and was also continuously updated.

#### (b) The Coding Process

The coding was processed continuously during the survey investigation period in 24 lots, each lot comprising about 200 two-week diaries, altogether with about 30,000 entries. All the information (numerical and verbal descriptions) of all the diaries entries of a certain lot, were entered into the computer by keying. The keyers were instructed how to key the verbal information according to simple rules, e.g., deleting certain prepositions and suffixes.

Each item description was transformed by a computer process into the standard form of three words, when one space at least was the criterion for ending a word, and each word was cut after its first eight characters. Then, each word of the items' descriptions was matched with the synonym-file in order to "translate" it, if necessary, to the correct representative word, as appearing in the dictionary.

The next step was to search for the matching description in the dictionary and so for the appropriate code. The matching was done by an algorithm aimed at minimizing computer search-time and conforming to specific features of the Hebrew language. In principle, both the dictionary items and the standardized description of the diary entries were sorted by alphabetic order. Then, the first word of each standard diary entry was searched for in the first words of the dictionary items. If an exact match was obtained and the indicator assigned to the dictionary word showed that the first word was sufficient for coding, a code was assigned; if not, a search was made for the second word applying the same rules, similarly for the third word, if necessary. If no exact match was obtained for the first word, no further search was made, and similarly with the second and third word when two or three words were required.

Most of the initial codes went through a number of verifications, in order to distinguish between similar item descriptions with different codes. For example, "coffee" can stand for a "bag of coffee" purchased in a supermarket, or for a "cup of coffee" in a cafeteria. Both the type of store and the unit's weight and specification (cup, bag, etc.) were used for this verification. These two details of the item in the diary were matched with those appearing in the dictionary to determine the legitimacy of the assigned code. If they agreed, the initial code was accepted, otherwise, it was deleted and no automatic coding was attained. Not all items of the dictionary were verified either because it was unnecessary, or, because no relevant verification data was available. It should be noted that further verification of the assigned codes were carried out after the coding process, during the advanced stages of the data processing, e.g., the expenditure together with the amount of the purchased item was checked against a range of possible prices.

The automatic coding for each coding batch resulted in two files: one file that included the coded entries and a second file that included entries which failed to be automatically coded. The uncoded entries were displayed by alphabetical order for manual coding, by which the codes were directly keyed into the computer. The coder could often enter a single code which was then copied automatically for all entries with the same description.

Overall, in the 1986/87 survey 81% of all the diary entries were coded automatically and this rate increased from 75% for the first lot to 83% for the last one (in the Swedish 1978 survey, the overall coding rate was 65%, but in their 1985 survey it rose to 82%, and in Denmark the rate was 75% - Lyberg and Dean, 1992). It should be noted that 1% of the items in the dictionary coded 50% of all diary entries, i.e., a few diary items were very frequent such as, various kinds of bread, milk, newspapers, and cigarettes.

#### (c) Quality control and evaluation of the coding process

Quality control of the coding process was carried out for several purposes:

- (1) To control the quality level of the keying operation.
- (2) For further training of the keyers by identifying their common keying errors.
- (3) To detect deficiencies in the dictionary or in the synonym-file in order to improve them for use in the subsequent coding lots.
- (4) To control the manual coding of those entries that failed to be automatically coded.

A 10% sample of all diary entries (either coded automatically or requiring manual coding) in each coding lot was used for the quality control. All the sampled entries were coded again manually and independently of the initial coding and the coder did not know if a code had been assigned automatically or not. Then, the initial codes were matched with the controlled codes. When the two codes agreed it was assumed that there was no error in the assigned code. If the two codes were not the same, an expert decided independently what was the correct code (not necessarily either of the two codes) and it was then determined whether the initial code was erroneous or not. When a wrong code was initially assigned it was corrected and checked for its source of error: the dictionary, the synonym-file, the keying of the item description, or the manual coding. The dictionary and the synonym-file were corrected, if needed, and the keyer and the coder were continuously warned of frequent errors.

In the previous survey, where only manual coding was used, the estimated rate of erroneous codes was 8.2% and in the 1986/87 survey, where automatic coding was used with some supplementary manual coding, this rate was 3.6%. After the correction of wrong codes in the sample of the quality control, the overall estimated rate of erroneous codes was reduced to 3.2% (this fell even more after price checking was carried out after the coding process in a separate stage).

On the average for all the lots, 3.0% of all automatic codes were erroneous: 1.8% due to the dictionary or to the synonym-file; 1.0% due to wrong keying of the item description; and 0.2% due to other reasons. These rates were higher for the earlier stages of the coding process and were lower for the later ones, since the dictionary was continuously corrected and the keyers became more proficient.

As was mentioned above, in the previous survey the overall rate of erroneous manual coding was estimated to be 8.2%, and the rate of erroneous codes of the manual coding in the 1986/87 FES of items not coded automatically was 6.2%. It would be expected that the rate of the errors in the non-automatic coding in the 1986/87 survey would be higher than the overall rate in 1979/80, because of the more complicated coding for items that were not automatically coded, e.g., less frequent items, unusual descriptions of the commodity in the diary, etc.

This can be explained by several reasons: a better code-book was prepared for this survey (based on the results of the evaluation study of the manual coding in the previous survey); better qualified coders in the 1986/87 survey; poorer coding due to fatigue and boredom in the previous survey; assigning the same code for a number of items with the same description resulting in less errors in 1986/87; and, perhaps due to less accurate quality control in 1986/87.

To conclude, the automatic coding proved to be preferable to manual coding. First, the rate of automatic coding was satisfactory and a much smaller rate of erroneous coding was achieved. Second, as compared to the manual coding used in the previous survey (on the same coding scale), 60% of labour cost was saved. However, considerable effort was invested in developing the automatic coding procedure for its first use, but in the long run, this can be considered as an investment.

## REFERENCES

- Appel M. and Hellerman E. (1983). Census Bureau Experience with Automated Industry and Occupation Coding. Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 32-40.
- Berg A. and Har S. (1990). Report on Automatic Coding in the 1986/87 Family Expenditure Survey in the Israel Central Bureau of Statistics, Jerusalem (Unpublished - Hebrew).
- Breiman L., Friedman J., Olshen R. and Stone C. (1984). Classification and Regression Trees. Monterey, California: Wadsworth, Inc.
- Corbert J.P. (1972). Encoding from Free Word Descriptions. Unpublished manuscript, U.S. Bureau of the Census.
- Cowan C. (1977). Incentive Effects on Amounts Reported in an Expenditure Diary Survey. Proceedings of the Section on Social Statistics, American Statistical Association, Part I, pp. 498-503.
- Elliot D. (1991). Weighting Expenditure and Income Estimates from the United Kingdom Family Expenditure Survey to Compensate for Nonresponse. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Elteto O. (1991). Nonresponse Rates and a Way of Adjusting for Nonresponse in Household Expenditure Surveys in Hungary. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Embry B. (1988). The ASCO Computer Assisted Coding System. Paper presented at the Social Research Conference, University of Queensland, Brisbane.
- Glaude M. (1982). A Decade of Methodological Investigations on Family Expenditure Surveys in France. Paper presented at the International Meeting on Analysis of Sample Survey Data, Jerusalem, Israel.
- Harrison R. (1991). Respondent Burden and Respondent Fatigue in the 1988-89 Australian Household Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- de Heer W.F. (1991). A New Method for the Budget Survey of 1988. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Israel Central Bureau of Statistics (1963). Family Expenditure Surveys 1950/51-1956/57-1959/60. Special Series No. 148. Jerusalem, Israel.
- Jacobs E., Jacobs C. and Dippo C. (1989). The U.S. Consumer Expenditure Survey. ISI Proceedings of the 47<sup>th</sup> Session, Paris, Vol. 1, pp. 123-142.
- Kantorowitz M. (1981). Evaluation of Some Aspects of the Israeli Family Expenditure Survey. ISI Proceedings of the 43<sup>rd</sup> Session, Buenos Aires, Vol. 3, pp. 1563-1582.
- Kantorowitz M. (1990). Improving Post-stratification by Constructing Hierarchical Classification Trees of Homogeneous Groups. Paper presented at the IOS/IASS Summer Conference on Survey Design, Methodology and Analysis, Essex, England.
- Kemsley W.F.F. and Nicholson J.L. (1960). Some Experiments in Methods of Conducting Family Expenditure Surveys. Journal of the Royal Statistical Society, Series A, Vol. 123, pp. 307-328.

- Kemsley W.F.F. (1961). The Household Expenditure Enquiry of the Ministry of Labour: Variability in the 1953-54 Enquiry. *Applied Statistics, Journal of the Royal Statistical Society*, Vol. 10 No. 3, pp.117-135.
- Kemsley W.F.F. (1975). Family Expenditure Survey. A Study of Differential Response Based on a Comparison of the 1971 Sample with the Census. *Statistical News*, No. 31, pp. 16-21.
- Kordos J. and Kubiczek A. (1991). Methodological Problems in the Household Budget Surveys in Poland. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Lindstrom H.L. (1991). An Experiment with Incentives. *Research & Development Report, Statistics Sweden*, No. 10, pp. 3-13.
- Lyberg L. and Dean P. (1992). Automated Coding of Survey Responses: An International Review. Paper presented at the Work Session on Statistical Data Editing, Conference of European Statisticians, Working Paper No. 3. Washington D.C.
- Matheson J. (1991). Application of Computer Assisted Interviewing to the United Kingdom Family Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- McWhinney I. and Champion H. (1974). The Canadian Experience with Recall and Diary Methods in Consumer Expenditure Surveys. *Annals of Economic and Social Measurement*, 3/2, pp. 411-437.
- Moller B. (1991). The Relationship Between Data from Accounting Books, Interviews and Registers. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Nevraumont U. (1991). Evaluation of the Canadian Food Diary Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- O'Reagan R.T. (1972). Computer-assigned Codes from Verbal Responses. *Communications of the ACM*, Vol. 15, pp. 455-459.
- Pearl R. B. (1979). Reevaluation of the 1972-73 U.S. Consumer Expenditure Survey. Technical Paper No. 46. U.S. Bureau of the Census, Washington, D.C.
- Prais S.J. (1958). Some Problems in the Measurement of Price Changes with Special Reference to the Cost of Living. *Journal of the Royal Statistical Society, Series A*, Vol. 121 Part 3, pp. 312-332.
- Redpath B. (1986). Family Expenditure Survey: a Second Study of Differential Response, Comparing Census Characteristics of FES Respondents and Non-respondents. *Statistical News*, No. 72, pp. 13-16.
- Russo A. and Coccia G. (1991). On the Characteristics and Effects of Non-responses in the Italian Consumer Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Silberstein A. (1989). Recall Effects in the U.S. Consumer Expenditure Interview Survey. *Journal of Official Statistics*, Vol. 5 No. 2, pp. 125- 142.
- Statistical Office of the European Communities (1981). The 1979 Harmonized Family Budget Surveys. Review of the Methodological Aspects of the Questionnaires in the Ten Member States.
- Statistics Canada (1990). Survey of Family Expenditures in 1990, Interviewer's Manual. Ottawa, Canada.

- Sudman S. and Ferber R. (1971). Experiments in Obtaining Consumer Expenditures by Diary Methods. *Journal of the American Statistical Association*, Vol. 66 No. 336, pp. 725-735.
- Tucker C. (1992). The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey. *Journal of Official Statistics*, Vol. 8 No. 1, pp. 41-61.
- Turner R. (1961). Inter-week Variation in Expenditure Recorded During a Two-week Survey of Family Expenditures. *Applied Statistics*, *Journal of the Royal Statistical Society*, Vol. 10 No. 3, pp. 136-146.
- U.S. Bureau of Labor Statistics (1988). BLS Handbook of Methods. Bulletin 2285. U.S. Department of Labor, Washington, D.C.

# ASPECTOS METODOLOGICOS EN LAS ENCUESTAS DE GASTOS FAMILIARES

## INTRODUCCION

Las encuestas sobre el gasto familiar (EGF) se encuentran entre las encuestas más importantes de vivienda y familia en muchos países y sirven para una amplia gama de utilidades. Estas encuestas pueden tener diferentes nombres en cada país, por ejemplo, Encuestas de Gasto del Consumidor, ó Encuestas de Presupuesto Familiar, pero básicamente son de tipo parecido. aparte de tener tanta importancia, las EGF también se encuentran entre las encuestas más problemáticas, complicadas y, por lo tanto, costosas para realizar, y además una de las más difíciles para conseguir la cooperación de la familia. A consecuencia de las exigencias que impone la encuesta, siempre ha sido muy difícil mantener un nivel de respuestas suficientemente alto. Además, las familias a menudo tienen reticencias a la hora de revelar los detalles de sus ingresos y gastos, considerándolo como una invasión de la intimidad. Como consecuencia, las estimaciones derivadas de las EGF pueden contener graves errores y entonces se pueden interpretar erróneamente, lo que lleva a tomar decisiones equívocas.

Así que, tanto en el pasado como actualmente, se han invertido muchos esfuerzos en evaluar la calidad de los datos de las EGF y en encontrar el origen de los errores en las diferentes fases de la encuesta. Estos estudios de evaluación tienen como objetivo la mejora de la metodología de las EGF, además de conseguir una mejor asignación de recursos entre los componentes de la encuesta.

Se asocia con las EGF una amplia gama de temas metodológicos y esta publicación trata de ellos, aunque sólo se cubre una selección de temas. Aunque los temas tratados en esta publicación están dentro del contexto de las EGF, algunos pueden tener relación con otros tipos de encuestas familiares.

En el Capítulo 1 se describe algo de la historia de las EGF para entender mejor los temas metodológicos. Se describen en este capítulo los propósitos y usos de las EGF, la población objetivo de la encuesta y la gama de variables incluidas en la investigación.

El Capítulo 2 proporciona una descripción general de los métodos de recogida de datos empleados en las EGF y sirve como una introducción a los Capítulos 3 y 4, donde se discuten en detalle los temas metodológicos asociados con el uso de diarios y cuestionarios.

El Capítulo 5 trata de los aspectos de muestreo de las EGF, con respecto a las dos fases de muestreo de las dimensiones de tiempo y de familia. El Capítulo 6 se concentra en los ratios y las características de la no respuesta en las EGF y su implicación en la fiabilidad de las estimaciones de la encuesta. También se incluye en este capítulo las formas de reducir la no respuesta y su diferencialidad entre los subgrupos.

Dado que el nivel alto de la no respuesta en las EGF puede causar un grave sesgo y porque las EGF proporcionan estimaciones en detalle, es esencial emplear un buen procedimiento de ponderación para reducir tanto los sesgos como las varianzas de muestreo. Es el tema del Capítulo 7 donde se discuten los métodos de ponderación, poniendo el énfasis en la mejora de la post-estratificación.

Finalmente, en el Capítulo 8, se presentan las técnicas de codificación de bienes y servicios en los registros de agenda. Se discute la importancia de evitar los errores de código en la codificación manual y se ilustra el uso de la codificación automática.

Para la mayoría de los temas discutidos se presentan los resultados empíricos de estudios relacionados, la mayor parte basada en la experiencia israelí, aunque también se presentan datos de otros estudios.

Las EGF incluyen muchos temas, pero están excluidos del ámbito de esta publicación. Además, como hemos mencionado antes, no se pueden cubrir todos los temas metodológicos aquí, entre ellos, los métodos de imputación para la no respuesta de ítems y los ajustes necesarios para los cambios de precio durante el período de la encuesta y los cambios en la composición de la familia durante el período de referencia.

## CAPITULO 1. ANTECEDENTES

### 1.1. Propósitos y Usos de las Encuestas de Gastos Familiares

El propósito primario de las EGF es de actualizar las ponderaciones para el Índice de Precio al Consumidor (IPC). Sin embargo, también las EGF siempre han tenido el objetivo de reflejar el nivel de vida de la familia y de proporcionar datos estadísticos para el análisis económico. Aunque originalmente estas encuestas se trataban principalmente del gasto de la familia, hoy en día, se recoge información adicional sobre muchas características de cada familia y sus miembros, así como las fuentes de ingreso y algunas veces sobre los ahorros. Como consecuencia, la encuesta ha llegado a ser una encuesta múltiple que proporciona un fondo único de datos sociales y económicos, ampliamente usado por los departamentos gubernamentales para hacer política, por los investigadores académicos para desarrollar y probar teorías y modelos, y por el sector comercial para la investigación del mercado, y por otros.

Más específicamente, las estimaciones derivadas de las EGF se pueden emplear para muchos propósitos:

1. Para vigilar el IPC general con ponderaciones actualizadas del "cesto" de bienes y servicios que contribuyen al cálculo del IPC.
2. Para proporcionar las ponderaciones actualizadas de "cestos" para calcular el IPC para subgrupos específicos de la población, por ejemplo, familias de gente mayor, familias que viven del bienestar social, etc.
3. Para suministrar las estimaciones requeridas para la ponderación o asignación de la muestra de los puntos de recogida de precios. Por ejemplo, la encuesta puede suministrar las estimaciones de la distribución del gasto en comida por el tipo de comercio (alimentación, supermercados, quioscos, puestos al aire libre, etc.) y por el día de la semana. Luego, la muestra para el IPC se puede planificar para representar esta distribución correctamente para reflejar más exactamente los precios medios ponderados en los diferentes puntos de compra y en días diferentes.
4. Para estudiar los hábitos de consumo de la familia por sus características socio-económicas, por ejemplo, tamaño de la familia y composición, educación, ingreso, etc. Esta información se puede utilizar para tomar decisiones políticas en conexión con la planificación social y económica, tales como las pensiones y programas de asistencia social, acuerdos salariales y varios pagos de apoyo. Además, los datos de la encuesta se pueden utilizar para estimar los efectos redistributivos de los impuestos directos e indirectos y de una amplia gama de beneficios sociales en la situación de varios tipos de familia.
5. Para analizar los niveles nutricionales de diferentes subgrupos de familias, usando las cantidades estimadas de comida consumida. Se usan las funciones empíricas junto con las cantidades estimadas consumidas para obtener las estimaciones de los componentes de nutrición. A menudo, las EGF son la única fuente de esta información a nivel nacional, a pesar de sus limitaciones

(a causa de desperdicio de la comida, insuficiente información sobre restaurantes, etc.)

6. Para analizar las elasticidades de la demanda de algunos bienes seleccionados, es decir, cómo la demanda de ciertos bienes está afectada por el nivel de ingresos dentro de subgrupos de la población. Por ejemplo, la demanda de comida esencial como el pan, se espera que sea menos elástico que la demanda de carne que es cara. Este tipo de análisis a menudo puede arrojar una luz sobre el comportamiento consumidor y se puede emplear para comprender mejor la estructura socio-económica de la población. A veces, esta información se puede usar para la política de impuestos o para la asignación de los fondos sociales entre los diferentes bienes y servicios.
7. Para estudiar las condiciones de la vivienda y la posesión de bienes duraderos con relación al ingreso y gasto, tanto para la población entera como para subgrupos seleccionados.
8. Para analizar la distribución de ingreso de la familia por fuentes de ingreso detallado, y la diferenciación entre ciertos grupos de familias.
9. Para estudiar los hábitos de ahorro de diferentes grupos de la población, incluídos los ahorros negativos, es decir, deudas.
10. Para estudiar las diferencias en el coste de la vida y el nivel de vida entre diferentes áreas del país, así como entre diferentes países.
11. Para analizar la variación en el nivel de vida sobre un período de años y las disparidades entre familias en diferentes grupos socio-económicos.
12. Para tomar decisiones comerciales y pronosticar las demandas del mercado para productores y vendedores, así como para hacer más eficiente la publicidad.
13. Para suministrar datos para utilizar en la compilación de las estimaciones oficiales de gasto en consumo en el sistema nacional de Cuentas y Saldos.

Esta lista de los posibles usos de las EGF no es completa y se pueden emplear las estimaciones de la encuesta para propósitos más específicos. Por ejemplo, en los países del Este de Europa, las EGF sirven como una fuente para asesorar la extensión del fenómeno económico reciente de la privatización del sector comercial (Kordos y Kubiczek, 1991).

## 1.2. La Población Objetivo de la Encuesta

Las EGF tienen una larga historia, incluso en los países en vías de desarrollo, y algunos países, por ejemplo, los Estados Unidos, Francia y Polonia, informan de que algún tipo de EGF se ha llevado a cabo durante casi cien años. En la mayoría de los países las primeras EGF tenían como objetivo cubrir solamente algunos subgrupos específicos de la población, por ejemplo, los trabajadores

industriales y sus familias, o los trabajadores de las grandes ciudades. La cobertura de la población objetivo de la encuesta se ha extendido gradualmente sobre los años en la mayoría de los países, y ha habido una tendencia general de cubrir todas secciones de la población, incluyendo familias de los empleados, las personas autónomas, granjeros, las personas sin empleo, etc., tanto en zonas rurales como urbanas.

Por ejemplo, en los Estados Unidos de América, la primera encuesta nacional sobre el gasto se llevó a cabo en 1888-91 para estudiar los hábitos de gasto de los trabajadores. Una segunda encuesta se realizó en 1901, cubriendo las familias de los trabajadores para proporcionar las ponderaciones para un índice de precios de la comida comprada por los trabajadores, que se usaba como un desinflador de los ingresos. Las encuestas de 1917-19 y 1934-36 proporcionaron las ponderaciones para el IPC, incluyendo todo tipo de gasto, pero sólo cubriendo personas con nómina en las ciudades y oficinistas. En 1935-36 la encuesta incluyó tanto las secciones urbanas como rurales de la población, en 1950 la encuesta cubrió sólo los consumidores urbanos, pero desde entonces, en 1960-61, 1972-73 y en la encuesta permanente continua desde 1979, las encuestas han cubierto toda la población civil no institucional (Oficina de Estadísticas de Empleo de EE.UU., 1988).

En Hungría (Elteto, 1991), la primera EGF se realizó a finales de los años veinte, cubriendo trabajadores y empleados y sus familias en Budapest. En 1948 las EGF empezaron con una periodicidad regular cubriendo todos los trabajadores, empleados y labradores, y desde 1976 la población de la encuesta se ha extendido para cubrir también las familias de pensionistas.

En Israel, la primera EGF en 1950/51 (una de las primeras encuestas llevadas a cabo después de haber establecido el estado), y las tres EGF consiguientes en 1956-57, 1959-60 y 1963-64, cubrieron solamente las familias de trabajadores con nómina que vivían en poblaciones de 10.000 ó más habitantes. En 1968-69, 1973-74 y 1979-80 se incluían todas las familias urbanas. La encuesta de 1985-86 incluyó todas las familias en las poblaciones de 2.000 ó más habitantes, igual que la encuesta actual de 1992-93. Sin embargo, no se cubre todavía en Israel en las EGF la población rural en las poblaciones más pequeñas (que consta de menos de 10% de la población entera, de la cual aproximadamente un tercio vive en "kibbutzim" donde las familias no son unidades individuales de consumidores).

La meta de cubrir todos los sectores de la población en las EGF no se ha logrado todavía en todas partes. Por ejemplo, a veces se excluye la población rural de la encuesta, debido a dificultades operacionales (como costes), o debido a problemas de definición (como bienes producidos y consumidos por la misma familia), pero no a causa de una decisión de política.

Dado que las estimaciones de las EGF se relacionan con las familias, ninguna de estas encuestas cubre la población institucional, por ejemplo, pacientes y residentes en residencias de la tercera edad, los enfermos ó los necesitados. Generalmente, las EGF tampoco cubren personas de tránsito en hoteles, misiones y conventos, comunas, cuarteles de tropas y familias en reservas militares, etc. Sin

embargo, las familias que forman unidades económicas independientes, incluso si comparten la vivienda, tienen que estar incluidas en las EGF. Así que, aquellas personas que tienen viviendas subvencionadas para la tercera edad, para estudiantes, para el personal de los hospitales, etc., tienen que estar incluidas en la encuesta. También es el caso para las personas que viven permanentemente en hoteles, parques de casas móviles, y otras estructuras utilizadas como lugar de residencia particular. La inclusión de estos grupos en la encuesta puede implicar algún problema de cobertura, así como los temas de contenido. Por ejemplo, podría haber un problema para distinguir entre los gastos de alquiler, electricidad, calefacción, agua, etc., dado que a menudo está todo incluido en un solo pago. No obstante, la inclusión de estos grupos es importante por su contribución significante a ciertos subgrupos de la población, como la tercera edad.

De vez en cuando, se llevan a cabo Encuestas de Gasto Familiar para cubrir subgrupos específicos de la población. Por ejemplo, en Israel, en 1963-64 y 1968-69 se realizaron EGF de las familias de personas minusválidas y de familias que recibían ayudas de asistencia social de forma regular. Se llevan a cabo encuestas especiales de este tipo para adaptar a los subgrupos específicos los procedimientos de investigación y el conjunto de las variables investigadas, y para usar una muestra mucho mayor de estos grupos que se podría obtener en una muestra de una EGF regular.

### 1.3. Las Unidades de Población

Se emplean términos diferentes para las unidades de población a que se refieren la mayoría de las estimaciones encuestales: "familia", "vivienda", "unidad consumidora", "unidad de gasto", etc., y se los pueden intercambiar en lo siguiente.

La definición de la unidad de población es básicamente igual en la mayoría de los países, aunque hay algunas variaciones entre las diferentes encuestas. Se usan cuatro criterios para definir un grupo de personas como una unidad:

- (a) Provisión común de la mayor parte de la comida y de la mayoría de las cosas esenciales de la vida.
- (b) Juntar los ingresos y compartir los gastos (hasta cierto punto).
- (c) Cohabitar en la misma vivienda.
- (d) Parentesco por sangre, matrimonio, adopción u otro arreglo legal.

Siempre se emplean alguna variación de los criterios (a), (b) y (c), mientras que el criterio (d) se emplea raramente, porque en los últimos años hay más personas que viven juntas como una familia normal, a pesar de no satisfacer este criterio. Sin embargo, el criterio (d) se usa a veces para definir las sub-unidades dentro de la familia para propósitos analíticos. Por ejemplo, en Canadá, además de la "familia" que se define por los primeros tres criterios, también se define una "Familia Económica" usando los cuatro criterios, y que puede incluir personas como niños acogidos,

hijos casados con sus esposos, hermanos, padres, abuelos, familia política, primos, etc. También usan la unidad "Familia del Censo" que refiere a la familia central, es decir, marido y mujer, o pareja no casada, o padre/madre soltero, con o sin hijos que nunca se han casado (sin hacer caso de su edad), donde los hijos adoptivos e hijastros tienen la misma condición que los hijos propios (Statistics Canada, EGF 1990, Manual del Encuestador).

Generalmente se usan los criterios (a) y (b) sin intentar cuantificar hasta qué punto se comparten los ingresos y gastos y la provisión en común de la comida. En cuanto al criterio (c), existe alguna variación al definir las personas que pertenecen a la familia, según la cobertura de la población, la duración de su estancia en la familia, etc. Por ejemplo, en Israel, se excluyen los ciudadanos que han estado en el extranjero durante al menos los tres meses anteriores a la encuesta y los residentes extranjeros (turistas o residentes temporales) están incluidos sólo si han estado en el país durante al menos un período de tres meses anterior a la entrevista (excepto los diplomáticos extranjeros que siempre se excluyen de las encuestas de población). Los miembros de una familia (o la familia entera) con dos o más direcciones residenciales, pertenecen a la dirección donde pasan la mayoría de las noches. Sin embargo, esto no es la norma para las personas casadas que siempre pertenecen a la casa del esposo/a, incluso si por alguna razón (trabajo, estudios, pero no la separación) tienen una dirección diferente. Este criterio no se aplica a los soldados ni a los niños en un internado, que siempre se incluyen en la dirección familiar. Los estudiantes, se consideran como una familia si viven en la universidad o en cualquier otra casa. Las personas de la tercera edad que viven en viviendas subvencionadas están incluidas en la población de la encuesta, si se consideran como familias económicamente independientes.

Generalmente se determina la composición de la familia según la situación durante la visita inicial del encuestador. Sin embargo, pueden ocurrir cambios en la composición familiar durante el período de referencia de la encuesta. En algunos países, existe un tratamiento especial para estos casos, recogiendo sólo la información de los cambios, o recogiendo más información para los ajustes.

En todas las encuestas se establecen normas tanto para poder conseguir unidades tan cerca como sea posible a las unidades de consumo, como para evitar sesgos en las estimaciones de las encuestas debido a una posible duplicación de la información o una falsa rebaja de los gastos. Sin embargo, se debería notar que para la gran mayoría de la población no hay problemas de definición en absoluto.

#### 1.4. Las Variables de Investigación

Para el uso primario de las EGF, es decir, proporcionar estimaciones de ponderaciones para el IPC general, sólo se requiere información detallada sobre los gastos. No obstante, normalmente se recoge mucha más información en las EGF para los demás usos de la encuesta, y, a veces, sólo para mejorar la exactitud de los datos proporcionados por los encuestados sobre los gastos. Así que, la recogida de los

datos en la mayoría de las EGF incluyen un conjunto comprensivo de variables de investigación, con alguna variación entre los diferentes países.

La lista de las variables potenciales de investigación que se presenta a continuación no es exhaustiva ni siempre se incluyen todas las variables de la lista en las EGF.

A. Composición de la familia según las características de sus miembros

- (1) Sexo, edad, estado civil, religión o grupo étnico, país de nacimiento y año de inmigración (en países con mucha inmigración)
- (2) Relación con el cabeza de familia (u otra persona de referencia)
- (3) Duración de pertenencia a la familia en el período de referencia

B. Las características socio-económicas de los miembros adultos de la familia

- (1) Características de educación - número de años de estudios; tipo del último colegio; nivel más alto logrado; estudios actuales; etc.
- (2) Actividades laborales y económicas - número de semanas y horas semanales de trabajo para las personas ocupadas; número de semanas de desempleo pero disponible para trabajar; razones de no estar disponible para trabajar (estudiante, ama de casa, etc.)
- (3) Características del empleo principal (o último empleo si no está actualmente ocupado) - posición (empresario, empleado, autónomo, etc.); industria (rama económica); ocupación; lugar de trabajo, etc.

C. Condiciones de la vivienda y lugar de residencia

- (1) Características de la unidad de residencia - número de habitaciones (o medida de área); comodidades de la vivienda (número de baños, tipo de calefacción, etc.). Estos apartados se emplean como un indicador del nivel de vida y algunas veces también para estimar el valor de la vivienda propiedad de o alquilada por la familia.
- (2) Posesión de bienes duraderos de consumo - vehículos; aparatos eléctricos (frigorífico, cocina, etc.); aparatos de televisión, y otros. Esta información se usa no sólo como un indicador del nivel de vida, sino también para mejorar la exactitud de los datos sobre gastos en bienes duraderos, y parcialmente para la imputación de los valores de los bienes duraderos de propiedad.

(3) Características del lugar de residencia - región; tipo y tamaño de la localidad, etc.

D. Gastos de consumo durante el período de referencia

Para cada gasto en cualquier bien o servicio (incluyendo regalos recibidos o regalados por la familia, artículos producidos y consumidos por la familia, existencias de negocios y de otros poseedores) se requieren los siguientes detalles:

(1) Una descripción detallada del artículo - para clasificarlo como una clase de artículo predeterminado.

(2) El valor de cada gasto - el dinero gastado en metálico o de crédito (y el flujo de los pagos), o el valor del artículo recibido o producido por la familia sin ningún pago directo.

(3) La cantidad comprada - cantidad total, número de unidades y el peso de la unidad. Esta última variable es opcional para usos específicos, pero ayuda a controlar la calidad de los datos sobre gastos.

(4) Fecha del gasto - generalmente definida como el día de recepción del bien o servicio y no como el día del pago mismo.

(5) Fuente de adquisición de los artículos no comprados directamente por la familia (regalos, productos caseros, tomados de las existencias de negocio, etc.).

(6) Tipo de comercio donde se compró el artículo - opcional para usos específicos adicionales de la encuesta, pero también se puede emplear esta variable para mejorar la calidad de los datos sobre gastos.

(7) Lugar de la compra - opcional para usos adicionales de la encuesta como la ubicación de la muestra de salidas para el IPC.

E. Ingresos de todas las fuentes

(1) Ingreso del trabajo remunerado - salarios y pagas; cantidades especiales recibidas del empresario, del trabajo autónomo y de otros trabajos.

(2) Ingreso de la pensión.

(3) Ingreso de capital - de viviendas; vehículos; alquileres; intereses y dividendos; etc.

(4) Ingreso de beneficios sociales y ayudas - de la Seguridad Social; bienestar social; de otras familias; etc.

(5) Ingreso de todas las demás fuentes.

F. Gastos no de consumo

- (1) Impuesto sobre la Renta y otros impuestos directos - esto a veces no se investiga directamente y se imputa según las normas de los impuestos.
- (2) Contribuciones a pensiones y a la Seguridad Social - a veces se consideran como ahorros.
- (3) Envíos, regalos y transferencias a otras familias (regalos que no son dinero se pueden considerar como consumo).

G. Ahorros de todo tipo

- (1) Ahorros en metálico.
- (2) Ahorros inmobiliarios.
- (3) Ahorros comerciales.
- (4) Recibos no regulares - cantidades importantes de un seguro; venta de posesiones y propiedades; etc.
- (5) Otros ahorros - préstamos obtenidos; pagos de préstamos recibidos; etc.

Apenas existen otras encuestas sobre la familia que cubren tan amplia gama de variables de investigación. Muchos problemas temáticos están relacionados a las definiciones de varias variables de investigación, así que, la sintaxis de las preguntas es muy importante y en consecuencia, las instrucciones para los entrevistadores son complicadas. Aunque todos estos temas no entran en el ámbito de esta publicación, los temas metodológicos de las EGF, que conciernen a esta publicación, también están afectados por la complejidad de la encuesta debido a la amplia gama de las variables de investigación.

## CAPITULO 2. LA RECOGIDA DE DATOS

### 2.1. Cuestionarios y Diarios

En la mayoría de las encuestas familiares, los datos requeridos se recogen de las unidades de la muestra por medio de un cuestionario o de un diario, o a veces, de las dos cosas.

Con los cuestionarios, los datos requeridos se recogen por medio de un conjunto de preguntas que puede referirse a varios tipos de información:

- (a) Características que no varían con el tiempo, como sexo, año de nacimiento, etc.
- (b) Datos relacionados con el período de la investigación, como el último colegio, número de habitaciones en la vivienda, posesión de bienes duraderos, etc.
- (c) Información retrospectiva sobre acontecimientos que sucedieron en un período de referencia predeterminado en el pasado, por ejemplo, número de visitas al médico durante el último mes, gastos en varios artículos durante los últimos tres meses, visitas al extranjero en el último año, etc.
- (d) Información retrospectiva relacionada con el último acontecimiento bajo investigación, cuando no hay un período de referencia predeterminado, por ejemplo, información sobre la última visita al médico.

Los diarios se usan para registrar datos detallados de acontecimientos específicos que ocurren durante cierto período y pueden ser de dos tipos principales:

- 1) Un diario de "recuerdo" para registrar detalles de acontecimientos que ocurrieron en el pasado reciente. Por ejemplo, en Israel, en la Encuesta sobre Hábitos de Viajar, en 1985, se utilizaba un diario de memoria de dos días para recoger detalles de todos los viajes de cada persona. Aunque la información recogida tanto por los diarios de recuerdo como por los cuestionarios retrospectivos se basan en el recuerdo, se consideran diferentes por la forma de recoger la información requerida y porque los diarios de recuerdo deben referirse a un período de tiempo muy corto.
- 2) Un diario "actual" donde se registra la información de forma continua durante el período de recogida de datos, a la vez que el acontecimiento sucede (o, al menos, dentro de un plazo de tiempo muy corto después del acontecimiento), como detalles de cada actividad de un día en la Encuesta sobre el Uso del Tiempo, o detalles de cada gasto en un período de dos semanas en la EGF. En la práctica, incluso cuando se emplea un diario actual, algún registro se hace por recuerdo.

La utilización de cuestionarios para recoger información retrospectiva, o el uso de los diarios de recuerdo, se justifica si se puede suponer que los encuestados son capaces de suministrar

información fiable de acontecimientos en el pasado. Sin embargo, puede pasar que el encuestado no recuerda con exactitud todos los detalles relacionados con un acontecimiento en particular, o incluso puede olvidar la ocurrencia de un acontecimiento dentro del período de referencia. También puede pasar que el encuestado informe erróneamente de acontecimientos que ocurrieron antes del período de referencia, o colocar mal el acontecimiento en el tiempo un fenómeno que se llama "telescoping".

Estos efectos de la mala memoria dependen de la duración del período que tienen que recordar y del tipo de acontecimiento. Así que, los detalles de los acontecimientos importantes, generalmente se recuerdan muy bien, por ejemplo, mudanzas de casa, un viaje alrededor del mundo, o un accidente de coche grave. Los detalles de los acontecimientos menos importantes se olvidan con más facilidad, por ejemplo, las fechas de haber comido en un restaurante durante el último año, o el gasto en artículos corrientes durante el último mes. Los efectos de la mala memoria también pueden depender de las características del encuestado, por ejemplo, la gente mayor tiene más dificultad para recordar ciertos acontecimientos.

Dado que los efectos de la mala memoria pueden causar graves sesgos en las estimaciones, en varias encuestas hay que recoger cierta información por medio de un diario actual. En teoría, los efectos de la mala memoria se deben evitar con el uso de un diario actual, pero en la práctica, porque algo de la registración se hace por recuerdo, incluso con este tipo del diario se pueden introducir algunos efectos de mala memoria, aunque en menor medida que con los diarios de recuerdo. Por otra parte, los datos recogidos por recuerdo no son afectados por posibles cambios de comportamiento de la unidad investigada.

Con los diarios actuales, el ratio de la no respuesta es generalmente más alto que con los diarios de memoria. Además, la diferencialidad entre los ratios de la no respuesta en los diferentes subgrupos de la población es normalmente mayor cuando se emplean diarios actuales. Sin embargo, el uso de los diarios de recuerdo puede causar una representación inferior de los días antes de un fin de semana, por ejemplo, los sábados, si el encuestado no trabaja los domingos, mientras que, con los diarios actuales se puede controlar mejor la distribución de los diarios sobre los días laborables.

En cuanto al coste, un diario de recuerdo de un solo día requiere solamente una visita del encuestador, mientras que para un diario actual de un día, se necesitan dos visitas, una para dejar el diario y otra para recoger el diario completada. Enviar por correo los diarios completadas ahorraría una visita, pero generalmente da como resultado un ratio de respuesta mucho más bajo. Se debería señalar que cuando se usa un diario actual, el encuestado necesita instrucciones, y de todas formas, el encuestador tiene que registrar algún material por recuerdo junto con el entrevistado. Cuando se requiere varios días de registro en un diario, por ejemplo, una o dos semanas, el diario de recuerdo es más costosa que el diario actual, porque el encuestador debe visitar a la unidad más a menudo para evitar serios efectos de mala memoria.

No hay ninguna preferencia clara para un tipo del diario más que el otro. Hay que tomar la decisión de qué tipo del diario a emplear separadamente para cada encuesta, según la duración del período de registro, el tipo de información requerida y las características de la población bajo investigación. En algunas encuestas se usan los dos tipos del diario. Por ejemplo, en la Encuesta de 1992 del Uso del Tiempo, se empleaban los dos tipos del diario: el encuestador completó un diario de recuerdo de un día para el día anterior a la entrevista (o de dos días de fin de semana), y los entrevistados también tenían que completar un diario de recuerdo de un día (o dos días de fin de semana) para un día predeterminado (normalmente el día después de la entrevista), para devolverla por correo.

## 2.2. Métodos de Recogida de Datos en las Encuestas de Gastos Familiares

Para casi todas las EGF, la recogida de los datos se basa en alguna combinación de cuestionarios y diarios. Se hace esto para obtener la información muy detallada requerida para la encuesta con el grado más alto posible de exactitud. El entrevistado no puede suministrar información fiable por memoria de la compra de pequeños artículos frecuentemente comprados y de pequeños gastos, incluso si no se compran a menudo, a menos que el período de referencia sea muy corto. Dado que un período del diario demasiado corto llevará a muchas varianzas muestrales, se usa un diario actual para registrar todos los gastos en detalle (o sólo para ciertos grupos de artículos como comida o ropa). Sin embargo, no es posible imponer una carga demasiado pesada a los encuestados, pidiéndoles a mantener un diario durante un largo período de tiempo, porque menos familias estarán dispuestas a participar en la encuesta, y más familias abandonarán antes del final del período de recogida de datos. Además, la calidad de los registros deteriora a lo largo del período de la recogida.

Por lo tanto, con un período de recogida razonable, los gastos menos frecuentes se registrarán en el diario rara vez y entonces no se pueden obtener estimaciones fiables de ellos porque habrá una alta varianza muestral. Por eso, se usan preguntas retrospectivas en un cuestionario para recoger información sobre gastos poco frecuentes. Esto se hace suponiendo que los efectos de la mala memoria no sean serios, dado que el período de recogida de datos no es demasiado largo. Se debería notar que parte de la información recogida en los cuestionarios se puede basar en documentos como facturas, nóminas, etc. Aunque al referirse a los documentos no habrá efectos de mala memoria, quizás el encuestado ya no disponga de ellos o no quiera enseñarlos durante la entrevista.

También se utilizan los cuestionarios para recoger datos sobre las características requeridas de la familia y sus miembros, además de alguna información tal como los hábitos de consumo. En algunas encuestas se utilizan los cuestionarios o los diarios o las dos cosas para cada miembro adulto de la familia, y en otras encuestas se emplean para la familia entera como unidad.

En la mayoría de las EGF, el encuestador completa los cuestionarios, mientras que se requiere que los entrevistados completen los diarios por sí solos. Los diarios rara vez están diseñadas para ser completadas por los encuestadores (excepto en algunos países

subdesarrollados donde todavía existe un alto porcentaje de analfabetismo). Los encuestadores completan los diarios sólo cuando no hay ningún miembro de la familia capaz de hacerlo, por ejemplo en familias de la tercera edad.

Algunas veces, se realiza una entrevista extensa, antes de dejar el diario, y otras veces sólo se mantiene una entrevista corta para empezar y la entrevista principal se realiza después de recoger el diario completada.

En algunos países, se lleva a cabo una encuesta donde cada unidad de la muestra tiene que mantener un diario además de tener una entrevista sobre toda la información requerida. En otros países, se llevan a cabo encuestas por separado en diferentes muestras, donde cada muestra se concentra en una forma de recogida de datos, entrevisto o diario. A veces se realizan encuestas del diario separadas con el objetivo de registrar todos los gastos (o todas las transacciones monetarias), y a veces tienen como objetivo principales grupos específicos de gastos, generalmente comida, pero también de otros grupos como ropa y calzado.

### 2.3. Ejemplos de Varios Países

- (1) En Israel, las EGF de 1986-87 y 1992-93 que se realiza actualmente, se han basado en un diario de dos semanas para registrar todos los gastos, mientras que en encuestas anteriores, se usaban diarios mensuales. Se lleva a cabo una entrevista de presentación para recoger datos sobre las características demográficas y socio-económicas de todos los miembros de la familia, así como sobre sus hábitos de consumo. Al final del período de recogida de datos el encuestador rellena un cuestionario extenso relacionado con los bienes duraderos y otros gastos importantes, así como con los ingresos y ahorros. El período de referencia es de los últimos tres meses para la mayoría de los artículos, y, para gastos poco frecuentes, doce meses. En las encuestas anteriores, el período de referencia era los últimos doce meses, aparte de algunos gastos regulares como teléfono, agua, etc., donde sólo se precisaba el último pago.
- (2) En el Reino Unido, (Matheson, 1991), se usa un diario personal de dos semanas para recoger datos de todos los gastos, para cada miembro de la familia de 16 años y más. Se emplea dos cuestionarios, el primero incluye preguntas sobre la composición de la familia, gastos de vivienda, etc., así como reembolsos y desgravaciones relacionados con la misma. Este cuestionario se relaciona con la familia entera como unidad, mientras que el segundo cuestionario, el programa de ingresos, pregunta a cada persona sobre los ingresos de todas las fuentes. El encuestador rellena estos cuestionarios, antes de entregar los diarios.
- (3) En Australia, (Harrison, 1991), en la EGF de 1988-89, se utilizaba un diario personal de todos los gastos durante un período de dos semanas, para cada miembro de la familia de 15 años y más. Además, se empleaban dos cuestionarios, uno para gastos y otro para ingresos. Se completaban ambos cuestionarios durante la entrevista inicial (aparte de la información relacionada con miembros de la familia no presentes durante la

visita inicial), antes de entregar los diarios personales. La primera parte del cuestionario sobre gastos incluía datos sobre las características de la familia, y sólo se pedía esta información de un miembro adulto de cada familia. La segunda parte recogía información sobre cierto tipo de gasto por parte de todos los entrevistados de la familia. Las preguntas se basaban en el recuerdo con un período de referencia de entre tres meses a dos años, según el tipo de artículo. Se completaba, para cada adulto, un cuestionario aparte para obtener información sobre todas las fuentes de ingresos, impuestos pagados, etc., así como alguna información socio-económica, por ejemplo, posición en el empleo, ocupación, etc.

- (4) En Dinamarca, (Moller, 1991), se usa un diario de dos semanas para registrar cada compra de comida, y un diario de cuatro semanas para todas las demás compras. El encuestado tiene que anotar en el diario una descripción detallada de cada compra, el gasto y la cantidad. También se incluye alguna información adicional para algunos artículos especiales, por ejemplo, sexo y edad (para clasificar mejor la compra de ropa) y si los artículos se compraban en el extranjero. Después de un año de duración de la encuesta con diario, al principio del año siguiente, se lleva a cabo una larga entrevista sobre pagos regulares (alquiler, seguros, etc.), gastos de bienes más caros, ingresos por fuente, etc., así como las características socio-económicas de la familia y sus miembros. Una fuente adicional de información son los registros administrativos, por ejemplo, el registro de impuestos personales y riqueza. Estas fuentes de datos son fáciles de obtener en Dinamarca, por vía del número personal de registración, y así se disminuye la carga tanto de los encuestados como del encuestador.
- (5) En Francia, (Glaude, 1982), en la EGF de 1978-79, se utilizaba un diario personal de diez días para cada miembro de la familia de 15 años ó más. Se usaban tres cuestionarios, en tres entrevistas separadas. Un cuestionario sobre la composición de la familia, vivienda, coches, lista de bienes duraderos, etc., y se completaba justo antes del período de recogida de datos. El segundo cuestionario se completaba durante el período de recogida, incluyendo preguntas sobre los hábitos de compra, consumo de productos propios y de negocios y alguna otra información. El tercer cuestionario, completado inmediatamente después del período de recogida, correspondía a los ingresos y gastos de los últimos doce meses, y a gasto en ropa, calzado y calefacción durante los dos meses anteriores.
- (6) En Hungría, (Elteto, 1991), hasta 1976, las familias que participaban en las EGF, anotaban sus gastos e ingresos durante todo el año, muchas de ellas incluso durante un período de varios años. A partir de 1976, la cantidad de recogidas de datos por medio de diarios ha sido reducida sensiblemente, y se piden a las familias anotar sus ingresos y gastos diarios con las cantidades en un diario de dos meses. Se dan diarios adicionales a los miembros adultos de la familia que quieren mantener un diario aparte. El entrevistador visita a la familia varias veces durante los dos meses para ayudar con cualquier problema que surge y para comprobar la información de los diarios. Al final

del año, el entrevistador visita una vez más a las familias para hacer preguntas sobre las varias fuentes de ingresos anuales (por ejemplo, salarios de segundos trabajos, impuestos personales, etc.), el gasto anual en algunos artículos de valor considerable (por ejemplo, bienes duraderos, vehículos, vacaciones al extranjero, etc.). Al mismo tiempo, se hace un inventario de los bienes duraderos.

- (7) En Canadá, (1990 Manual de EGF), El programa de la EGF de 1990 incluye dos encuestas diferentes. Se realiza una encuesta con diario durante un año del calendario que proporciona detalles de gastos y cantidades de sólo la comida. Se piden a las familias que completen un diario a diario durante dos semanas. Antes de entregar los diarios, se lleva a cabo una corta entrevista, incluyendo preguntas sobre las características socio-económicas de cada miembro de la familia, sobre las características de la vivienda ocupada por la familia, sobre los ingresos personales durante los últimos doce meses, sobre gastos en comida fuera de casa y sobre los hábitos de gastos de la familia. En febrero y marzo del año siguiente, se administra una encuesta anual de refuerzo que proporciona detalles sobre todas las compras de bienes y servicios de consumo durante el año civil anterior, y que también incluye el gasto anual en comida (como grupo) así como ingresos, impuestos, primas de seguros, regalos, etc. de las personas, además de los cambios en la posición económica de la familia durante el año anterior. También, se recoge información sobre la composición de la familia.
- (8) En los EE.UU. (Jacobs, Jacobs y Dippo, 1989), en el EGF de 1960-61 se usaba un procedimiento de recuerdo anual, pero se abandonó este método de recoger datos y se estableció una nueva metodología para las encuestas consiguientes. La EGF de 1972-73 y la encuesta actual que se inició en 1979, consta de dos encuestas separadas, cada una con una técnica diferente de recogida y muestreo: la Encuesta del diario y la Encuesta de Entrevista que es una encuesta de rotación de panel.

En la Encuesta del diario, se pide a cada familia que mantenga un diario de todos los gastos durante dos períodos consecutivos de una semana. Durante la primera visita del entrevistador a la familia, se recogen datos sobre la unidad familiar y los datos demográficos de cada miembro, antes de dejar el diario. Al completar las dos diarios semanales, el entrevistador recoge datos sobre las características laborales y de los ingresos, de forma personal.

En la Encuesta de Entrevista se pide a cada unidad investigada cinco entrevistas trimestrales. Durante la entrevista inicial, se recogen datos de un período de sólo un mes de memoria, que se usan para propósitos de delimitación. También esta primera entrevista sirve para recoger datos de las características demográficas, así como para proporcionar un inventario de bienes duraderos principales. Durante las otras cuatro entrevistas se recoge la información por medio de cuestionarios uniformes con un período de referencia de tres meses. Algunos detalles sobre los ingresos sólo se recogen en la segunda y quinta entrevista.

Aunque los pocos ejemplos descritos aquí no reflejan todos los métodos de recogida de datos que se usan actualmente, parece ser que la recogida de datos en casi todas las EGF se basa en registros del diarios con una o más entrevistas suplementarias. Las principales diferencias entre los países son:

- La duración del período de recogida de datos del diario.
- La duración del período de referencia para los cuestionarios retrospectivos y el número de entrevistas que se realizan para cubrir el período.
- El momento de las entrevistas y del período de la recogida de datos.
- El uso de un período fijo de referencia para el cuestionario retrospectivo para la muestra entera en comparación con un período de referencia móvil.
- Los datos a anotar en los diarios y el contenido del cuestionario.
- La recogida de los datos de cada miembro de la familia o de la familia entera o de ambas cosas.
- El uso de una encuesta extensa en vez de dos (ó más) encuestas separadas para los diferentes grupos principales de gasto.

## CAPITULO 3. EL DIARIO

Los diarios Actuales o Continuos se usan en las EGF para recoger una parte de la información requerida porque se cree que los encuestados proporcionan datos más precisos para esta información en los diarios que en una entrevista retrospectiva. Sin embargo, esto no significa que los datos recogidos en los diarios están libres de errores de varios tipos, y su calidad depende del tipo y el contenido del diario, y por supuesto, de la duración del período del diario. Se discuten en las siguientes secciones de este capítulo algunos temas metodológicos sobre el uso de diarios en las EGF y se presentan los resultados de varios estudios relacionados.

### 3.1. Duración del Período de Recogida de Datos

La duración del período del registro puede afectar la fiabilidad de las estimaciones de la encuesta respecto a lo siguiente:

- (a) Errores en el registro - Cuanto más largo sea el período de recogida de datos, más sesgo habrá en el registro debido al cansancio. Sin embargo, si el período de recogida es corto, los efectos telescopicos en los diarios, es decir, la inclusión de artículos comprados antes del período del diario, durante los primeros días del registro, pueden causar más sesgo en las estimaciones de la encuesta.
- (b) Cambios en el comportamiento de gasto - Se pueden esperar cambios en el comportamiento de la familia a causa de la registración de los gastos, después de estar implicada en la encuesta al menos algunos días. Estos cambios pueden resultar en sesgos que incrementan según la duración del período de la recogida de datos.
- (c) Errores muestrales - Para un número dado de familias en la muestra, los errores muestrales disminuyen según la duración del período del recogida. No obstante, esto no es siempre verdad para un coste dado de investigación. El coste por familia incrementa según la duración del período de recogida, aunque no necesariamente a la par, dado que se invierte algún gasto corriente fijo en cada familia de la muestra, que no depende de la duración del período de la recogida de datos.
- (d) La no respuesta - Cuanto más larga sea la duración del período de la recogida, más difícil será persuadir a la familia que participe en la encuesta. Así que, se espera un ratio más alto de la no respuesta. Además, durante un período de recogida largo, hay un grave problema de abandonos de los participantes. Si se requiere un período muy largo, habrá posibles traslados de casa de las familias, y así, una pérdida de unidades durante la encuesta.
- (e) Cambios en la composición de la familia - Cuanto más largo sea el período de recogida, se esperan más cambios. Así que, la información adicional, no siempre fácil de obtener, se requiere para ajustar los datos de la familia.

(f) Estimaciones de la distribución y de los parámetros de variabilidad de la población - Con un período corto del recogida de datos, estas estimaciones por gasto son limitadas, o incluso imposibles de obtener, cuando existe una alta variabilidad en gasto entre períodos cortos en las familias.

Uno de los primeros estudios extensos sobre los efectos del uso de diferentes métodos del diario para las EGF se realizó en Israel en la EGF de 1956-57. La muestra de familias se dividió aleatoriamente en tres submuestras, aproximadamente iguales. Todas las familias en las tres submuestras tenían que registrar todos los gastos no de comida durante un mes entero, pero las familias en la submuestra (A) tenían que registrar gastos en comida durante todo el mes, en la submuestra (B) sólo durante las primeras dos semanas, y en la submuestra (C) sólo durante las últimas dos semanas del mes.

Los resultados principales, basados en una muestra total de 6.600 familias, se presentan en la Tabla 3.1.

**Tabla 3.1.  
Gastos Medios de las Submuestras (B) y (C) Comparados con la  
Submuestra (A)  
EGF de Israel 1956-57**

| Grupo de gasto | Submuestra |       |       |         |
|----------------|------------|-------|-------|---------|
|                | (A)        | (B)   | (C)   | (B)+(C) |
| GASTO TOTAL    | 1.000      | 1.099 | 1.053 | 1.076   |
| Comida         | 1.000      | 1.111 | 0.982 | 1.046   |
| No comida      | 1.000      | 1.091 | 1.103 | 1.097   |

Fuente: EGF, Israel C.B.S., Serie Especial nº 148, Jerusalem (1963)

Para Comida, la comparación entre (B)+(C) y (A) corresponde al efecto de la duración del período de la recogida, mostrando un exceso de 4,6% del registro de medio mes sobre el mes entero. La comparación entre (B) y (C) para Comida, puede expresar diferencias reales entre los gastos de las dos mitades del mes debido a la fecha de recibimiento del salario, generalmente el primer día del mes.

En cuanto a la No comida, las familias en las tres submuestras registraron sus gastos para exactamente el mismo período de un mes civil. Los diferentes resultados sólo pueden ser la consecuencia de las diferentes exigencias para el registro de gastos de comida. Los resultados muestran que el gasto medio no de comida es alto (+10,3%) para (C) y algo menos (+9,1%) para (B) al compararse con (A). Estos resultados indican que la registración es menor cuando se impone una carga más pesada a los encuestados.

Entre otros resultados que se obtenían de este estudio (que no se presentan aquí), una comparación del gasto medio en Comida en la segunda mitad del mes entre (A) y (C) mostraba que se obtenían unos medios más altos en (C): 4% para Total Comida, y para algunos

artículos de comida detallados las diferencias eran mucho mayor, por ejemplo, +14% para Carne y +23% para Aceite. Podría ser un efecto del cansancio en la recogida de datos. Una comparación del gasto medio de Total Comida en la primera mitad del mes entre (A) y (B), mostraba que en (B) se obtenía un medio más alto (+11%). Este resultado no fue fácil de explicar.

Los resultados de este estudio pionero indican la gravedad de los efectos de los diferentes métodos del diario sobre las estimaciones, y que estos efectos no son los mismos para diferentes grupos de artículos. Así que las ponderaciones para los IPC podrían estar incluso más seriamente sesgadas.

### **3.2. Variaciones durante el Período de Recogida de Datos**

La variación en los niveles del gasto medio como se recoge en los diarios los diferentes días del período de recogida podría ser debido a diferencias reales en los mismos gastos, o el resultado de los efectos de la recogida, o una interacción de las dos cosas.

Las diferencias reales pueden ser debido a los efectos de:

- (1) El día de la semana
- (2) Vacaciones y las vísperas (parecido a los fines de semana)
- (3) El período de preparación de las vacaciones, cuando se gasta más de lo normal.
- (4) El día de recibir el salario (por ejemplo, el primer día del mes cuando los salarios son mensuales, o los jueves cuando el salario es semanal).
- (5) Días cuando el comportamiento del comprador puede estar afectado por varios otros factores, por ejemplo, los primeros días de las rebajas de temporada, o períodos de cambios en los precios.

Los efectos de la recogida de datos pueden ser causados por:

- (1) El efecto del cansancio o falta de interés en la recogida de datos, que aumenta durante el período de la recogida, resultando en una caída de entradas en el diario según la duración de la misma.
- (2) Efectos telescopicos, es decir, la inclusión de artículos comprados anteriormente al período del diario, durante el primer día o días del registro.
- (3) Omisión de anotar cada gasto el mismo día de la compra, por ejemplo, los encuestados olvidan de empezar una nueva página del diario cuando ésta está estructurada para empezar una página nueva cada día, u olvidan cambiar la fecha cuando se requiere para distinguir entre los días.
- (4) Efectos de mala memoria, es decir, cuando los encuestados retrasan las anotaciones en el diario y luego intentan reconstruir las compras. Los efectos de memoria también suceden cuando los entrevistadores, durante sus visitas a la familia,

reconstruyen compras al comprobar los registros de los encuestados, o cuando tienen que completar ellos mismos los diarios.

- (5) Cambios temporales en el comportamiento de gasto normal, debido a la implicación en la encuesta. Las familias pueden comprar más de lo habitual durante los primeros días de la encuesta para impresionar al entrevistador, y así, más tarde compran menos para compensar. A veces los encuestados hacen compras durante el período del diario, que normalmente hubieron retrasado, para mostrar gastos "típicos" y ser "representativo". Mantener un registro también puede perturbar los hábitos de gasto cuando las anotaciones al principio del diario reflejan el comportamiento del encuestado, y si la imagen no le favorece, se abstiene de hacer ciertas compras. Es inevitable que, cuando las personas saben que son observadas, no se comportan con naturalidad, aunque nunca lo admitirían.
- (6) El momento de la visita inicial del entrevistador a la familia para dejar el diario y las visitas posteriores durante el período del registro pueden influir en los niveles de los registros, es decir, los niveles más altos de las anotaciones son motivados por las visitas, con efectos de disminución después de la visita.

Los efectos de la recogida de datos pueden causar sesgos en las estimaciones de la encuesta, pero unos podrían cancelar a otros, por ejemplo, el efecto "telescopico" podría causar una sobre-estimación, mientras que el cansancio podría resultar en una sub-estimación. Generalmente, se cree que los registros en los diarios tienden a subestimar los valores reales, aunque nunca ha sido probado.

El fenómeno de las variaciones en los niveles de gasto durante el período de recogida en el diario en las EGF también salió en muchas encuestas del diario. Este tema discute Prais (1958), Kemsley y Nicholson (1960), Kemsley (1961), Turner (1961) Sudman y Ferber (1971), McWhinney y Champion (1974), Pearl (1979), Nevraumont (1991) y otros. Se encontraba variación en el numero medio de anotaciones diarias, o en el gasto diario por anotación o en las dos cosas. La extensión de las variaciones no es la misma en las diferentes encuestas, y tampoco en los diferentes grupos de gasto en la misma encuesta.

Kantorowitz (1981) realizó un estudio sobre la variación durante el período del diario para la EGF 1975-76 de Israel. Este estudio se basaba en una muestra de 2.270 familias que tenían que registrar sus gastos en un diario durante un mes entero, en la mayoría de los casos, empezando el primer día del mes. La visita inicial del entrevistador a cada familia se llevó a cabo tan cerca como fuera posible al día de iniciar el diario.

Una comparación sencilla del gasto medio por los días del diario no separaría las diferencias reales de los efectos de registración. Los ratios del gasto medio diario al medio diario total se calcularon para cada día de la semana, como se presenta en la Tabla 3.2 y estos efectos reales de los días de la semana resultaron ser sustanciales. Así que, estos efectos así como los efectos de las fiestas y sus

vísperas (los sábados y viernes, respectivamente), fueron eliminados. Sin embargo, los otros posibles efectos reales, antes mencionados, no se podían eliminar.

**Tabla 3.2**  
**Ratio del Gasto Medio Diario al Medio Diario Total**  
**EGF de Israel 1975-76**

| Grupo seleccionado de artículos     | Domingo     | Lunes       | Martes      | Miercoles   | Jueves      | Viernes     | Sábado      |
|-------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| <b>GASTOS TOTALES DEL DIARIO</b>    | <b>1.00</b> | <b>1.07</b> | <b>1.10</b> | <b>1.25</b> | <b>1.28</b> | <b>1.00</b> | <b>0.21</b> |
| <b>Comida (excl. Verd. y Fruta)</b> | <b>0.80</b> | <b>1.00</b> | <b>1.10</b> | <b>1.32</b> | <b>1.43</b> | <b>1.19</b> | <b>0.16</b> |
| Pan                                 | 0.99        | 0.98        | 1.00        | 0.95        | 1.08        | 1.19        | 0.09        |
| Leche                               | 1.09        | 1.05        | 1.08        | 1.02        | 1.06        | 1.66        | 0.04        |
| Carne y Aves                        | 0.49        | 1.03        | 1.22        | 1.82        | 1.82        | 0.52        | 0.10        |
| Verduras y Fruta                    | 0.67        | 1.04        | 1.07        | 1.53        | 1.57        | 1.03        | 0.09        |
| Ropa y Calzado                      | 1.31        | 1.20        | 1.18        | 1.23        | 1.13        | 0.82        | 0.13        |
| Transporte y Comunicación           | 1.20        | 1.06        | 1.08        | 1.09        | 1.01        | 1.12        | 0.44        |

\*) Sábado es día de fiesta en Israel.

Fuente: Kantorowitz (1981)

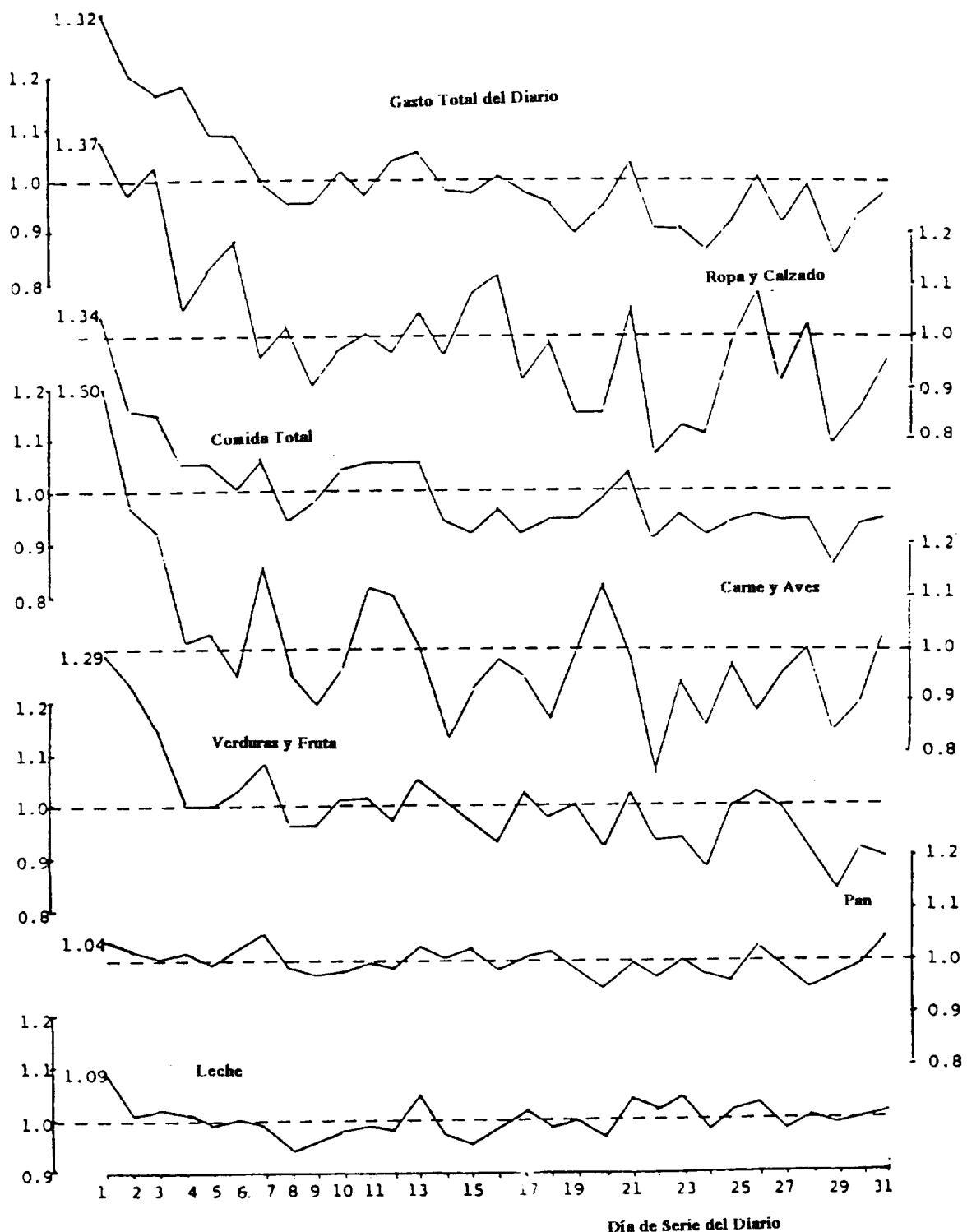
La diferencia entre los días de serie del diario (casi siempre coinciden con los días calendarios del mes), se expresaban por medio de los ratios del gasto medio para cada día, relativo a medio diario total (después de ajustes para días de labor y días de fiesta). Los ratios se presentan en la Figura 3.1 para grupos seleccionados de artículos: compras frecuentes (Leche, Pan); menos frecuentes (Verduras y Fruta, Carne y Aves); comparativamente infrecuentes (Ropa y Calzado), y también de Total de Comida y Gasto Total del diario.

El resultado más destacable fue el ratio sustancialmente alto de gasto en el primer día del registro para la mayoría de los artículos: 1.50 para Carne y Aves, 1.37 para Ropa y Calzado; 1.34 para Total Comida; 1.32 para Gasto Total del diario; 1.29 para Verduras y Fruta, pero menos para los artículos comprados con frecuencia, Pan y Leche, donde estos ratios fueron de 1.04 y 1.09 respectivamente. Además para la mayoría de los artículos, exceptuando los frecuentemente comprados, los primeros días del mes de la recogida de datos también tenían niveles relativamente altos de gasto, pero no tan altos que aquellos del primer día. Se debería notar que las diferencias entre los gastos medios diarios, sin la eliminación de los efectos reales para días de labor y de fiesta, fueron incluso mayores, pero siguiendo una pauta similar. Por ejemplo, los ratios del primer día para Carne y Aves fueron 1.50 y 1.74, respectivamente con y sin esta eliminación.

El efecto del primer día, especialmente para compras poco frecuentes, se podría explicar en parte en este encuesta por el efecto real de que los salarios a menudo se pagan el primer día del mes. Sin embargo, efectos similares del primer día también se encuentran en otras encuestas, donde los efectos reales de este tipo

no existen. Así que, es más probable que estos resultados sean debido al efecto telescopico. Este efecto podría tener un efecto sustancial sobre la recogida de datos en la encuesta israelí, como en otras similares, porque la visita inicial del entrevistador y la entrega del diario se realiza uno o dos días antes que el primer día del diario. Entonces, los encuestados, por error, pueden entender que tienen que comenzar la recogida inmediatamente. A consecuencia de estos descubrimientos, desde la EGF de 1986-87 en Israel, se recoge el primer día para propósitos de delimitación, pero no se incluye en las estimaciones de la encuesta.

**Figura 3.1**  
**Ratio del Gasto Medio(\*) para Día Calendario del Mes Relativo al**  
**Medio Total Diario**  
**EGF 1(\*) 975-76 Israel**



(\*) Después de hacer ajustes para días de labor y de fiesta.

Las diferencias entre las semanas se median por los ratios del gasto medio para cada una de las cuatro semanas en relación al medio semanal total (se ignoraban los últimos 2 ó 3 días del mes). Se calculaban estos ratios después de haber eliminado la mayoría de los efectos de días de fiesta, excluyendo los datos de dos meses (Octubre y Abril) que tienen las fiestas más importantes en Israel. Esto redujó ligeramente las diferencias entre semanas. La Tabla 3.3 presenta las diferencias entre las cuatro semanas, comenzando con el primer día del mes y con el segundo día.

**Tabla 3.3.**  
**Gasto Medio para Cada Semana Relativo al Medio Total Semanal**  
**EGF 1975-76 de Israel**

| Semana  | Grupos de Gasto Seleccionados |              |                  |              |      |       |                |
|---|-------------------------------|--------------|------------------|--------------|------|-------|----------------|
|   | Total del Diario              | Comida Total | Verduras y Fruta | Carne y Aves | Pan  | Leche | Ropa y Calzado |
| <b>Cuatro semanas comenzando el primer día del mes</b>  |                               |              |                  |              |      |       |                |
| Semana 1  | 1.14                          | 1.10         | 1.10             | 1.12         | 1.01 | 1.01  | 1.21           |
| Semana 2  | 0.98                          | 0.99         | 0.98             | 0.99         | 1.00 | 0.99  | 0.94           |
| Semana 3  | 0.95                          | 0.96         | 0.97             | 0.97         | 0.99 | 0.99  | 0.91           |
| Semana 4  | 0.93                          | 0.95         | 0.95             | 0.92         | 1.00 | 1.01  | 0.94           |
| <b>Cuatro semanas comenzando el segundo día del mes</b> |                               |              |                  |              |      |       |                |
| Semana 1  | 1.09                          | 1.07         | 1.07             | 1.08         | 1.01 | 0.99  | 1.17           |
| Semana 2  | 0.98                          | 0.99         | 1.00             | 0.99         | 0.99 | 0.99  | 0.94           |
| Semana 3  | 0.98                          | 0.98         | 0.99             | 0.98         | 1.00 | 1.01  | 0.94           |
| Semana 4  | 0.95                          | 0.96         | 0.94             | 0.95         | 1.00 | 1.01  | 0.95           |

Fuente: Kantorowitz (1981)

No se podían apreciar diferencias semanales significativas para Pan y Leche. Para otros grupos de gasto, los medios de la primera semana fueron significativamente más altos (1.10 y más), mientras que la cuarta semana era significativamente más baja (0.92-0.95). Para Ropa y Calzado el efecto de la primera semana era relativamente alto (1.21), probablemente debido en parte al efecto real del principio del mes. Los resultados muestran que incluso después de excluir los gastos del primer día, la primera semana aún tenía niveles más altos de gasto, aunque no tan altos como cuando se incluía el primer día.

Las diferencias entre semanas podría ser el resultado de una posible interacción entre el nivel de gasto de la familia y la forma en que se recogió su gasto. Así que, también se calcularon los ratios de cada semana usando el porcentaje semanal del total gasto de cada familia, en lugar de usar los gastos totales semanales. La mayoría de estos ratios eran iguales o con una ligera diferencia (aprox. 1%) de aquellos en la Tabla 3.3. Ropa y Calzado mostraron diferencias algo mayores, indicando la existencia de una ligera interacción.

También era interesante comprobar si la mayoría de las familias tenían una forma similar de recogida de datos en el diario, o si los resultados obtenidos caracterizaron una minoría de las familias con

variación extrema en sus registros durante el período de la agenda. Con este fin, se clasificaban las semanas dentro de cada familia según el nivel de gasto, es decir, la semana de mayor gasto se clasificaba como rango 1, etc., en total había 24 posibles combinaciones (o más si había empates). Se agruparon las familias según qué semana tenía el rango de 1, 2 etc., y según los rangos asignados a la primera semana, segunda semana, etc. La Tabla 3.4 presenta las proporciones de estos grupos de población y el gasto medio de cada semana en relación al medio semanal del grupo para el Gasto Total de Agenda.

**Tabla 3.4**  
**Gasto Medio Semanal Relativo al Medio Total Semanal**  
**de Familias con Diferentes Formas de Registro**  
**EGF 1975-76 de Israel**

| Comportamiento de Recogida |                |                |                | Porcentaje de Familias | Gasto medio semanal relativo a la media total semanal |                |                |                |                |
|----------------------------|----------------|----------------|----------------|------------------------|---|----------------|----------------|----------------|----------------|
| Rango de gasto             | Semana del mes |                |                |                        | Todas semanas   | 1 <sup>a</sup> | 2 <sup>a</sup> | 3 <sup>a</sup> | 4 <sup>a</sup> |
|                            | 1 <sup>a</sup> | 2 <sup>a</sup> | 3 <sup>a</sup> | 4 <sup>a</sup>         |   |                |                |                |                |
| <b>TODAS FAMILIAS</b>      |                |                |                | 100.0                  | 1.00  | 1.14           | 0.98           | 0.95           | 0.93           |
| 1                          | 1 - - -        | 37.3           | 1.00           | <u>1.56</u>            | 0.85  | 0.81           | 0.78           |                |                |
|                            | - 1 - -        | 22.5           | 1.00           | 0.89                   | <u>1.48</u>   | 0.83           | 0.80           |                |                |
|                            | - - 1 -        | 20.2           | 1.00           | 0.88                   | 0.82  | <u>1.47</u>    | 0.83           |                |                |
|                            | - - - 1        | 20.0           | 1.00           | 0.89                   | 0.83  | 0.82           | <u>1.46</u>    |                |                |
| 2                          | 2 - - -        | 25.7           | 1.00           | <u>1.07</u>            | 1.01  | 0.96           | 0.96           |                |                |
|                            | - 2 - -        | 26.6           | 1.00           | 1.22                   | <u>1.04</u>   | 0.87           | 0.87           |                |                |
|                            | - - 2 -        | 24.2           | 1.00           | 1.17                   | 0.94  | <u>1.04</u>    | 0.85           |                |                |
|                            | - - - 2        | 23.5           | 1.00           | 1.08                   | 0.94  | 0.93           | <u>1.05</u>    |                |                |
| 3                          | 3 - - -        | 19.5           | 1.00           | <u>0.85</u>            | 1.09  | 1.04           | 1.02           |                |                |
|                            | - 3 - -        | 26.0           | 1.00           | 1.19                   | <u>0.82</u>   | 1.02           | 0.97           |                |                |
|                            | - - 3 -        | 28.5           | 1.00           | 1.21                   | 1.03  | <u>0.81</u>    | 0.95           |                |                |
|                            | - - - 3        | 26.0           | 1.00           | 1.22                   | 1.01  | 0.96           | <u>0.81</u>    |                |                |
| 4                          | 4 - - -        | 17.5           | 1.00           | <u>0.65</u>            | 1.13  | 1.12           | 1.10           |                |                |
|                            | - 4 - -        | 24.9           | 1.00           | 1.22                   | <u>0.63</u>   | 1.07           | 1.08           |                |                |
|                            | - - 4 -        | 27.1           | 1.00           | 1.23                   | 1.08  | <u>0.62</u>    | 1.07           |                |                |
|                            | - - - 4        | 30.5           | 1.00           | 1.27                   | 1.09  | 1.04           | <u>0.60</u>    |                |                |

Source: Kantorowitz (1981).

Estos resultados muestran que 37% de todas las familias registraron sus gastos más altos durante la primera semana, mientras que 20% registraron su gasto más alto durante la cuarta semana. Esta distribución de las familias era significativamente diferente a la distribución uniforme esperada. Aunque muchas familias registraron sus gastos más altos durante la primera semana, no constituyeron una mayoría.

Para el 37% de las familias que tenían el gasto más alto durante la primera semana, el nivel de la semana con el gasto más alto era 56% mayor que el medio total semanal. Los gastos de las familias que tenían el gasto más alto durante la segunda, tercera y cuarta semana, superaron su medio semanal por sólo 48%, 47% y 46% respectivamente. Además, cuando las familias agrupadas por la semana con el intervalo 2 (así como los intervalos 3 y 4), la primera semana tenía un nivel de gasto un poco más alto: 1.07 contra 1.04-1.05 para el intervalo 2, 0.85 contra 0.82-0.81 para el intervalo 3, y 0.65 contra 0.63-0.60 para el intervalo 4. Así que, el fenómeno de los niveles más altos de gasto durante la primera semana fue debido, en parte, a que muchas familias tenían sus gastos más altos durante la primera semana del mes (37% superó el 25% esperado), pero también se debía a un gasto relativamente más alto durante la primera semana para todas las semanas del mismo intervalo.

En la encuesta israelí, los días en que el entrevistador visitó las familias durante el período del diario no fueron fijos, sino dependía de la capacidad de cada familia a mantener un diario, como medio cada 3 ó 4 días. Entonces, no es probable que el momento de las visitas afectó la variación diaria en general, aunque podría haber afectado cada familia por separado.

Sin embargo, en otras encuestas con días fijos para visitas, el efecto de las visitas repetidas fue sustancial. Por ejemplo, los resultados de la EGF de EE.UU. en 1972-73 (Pearl, 1979) donde se visitaron a la mayoría de las familias el octavo día de los registros en diarios, se observó un incremento significativo en el gasto medio para aquel día, como se puede ver en la Tabla 3.5.

**Tabla 3.5  
Gasto Relativo(\*) para Cada Diario de 14 Días  
EE.UU. Encuesta de 1972-73**

| Gasto     | Todos<br>days | Día de registro |     |     |     |     |     |     |                |     |     |     |     |     |     |
|-----------|---------------|-----------------|-----|-----|-----|-----|-----|-----|----------------|-----|-----|-----|-----|-----|-----|
|           |               | Primera semana  |     |     |     |     |     |     | Segunda semana |     |     |     |     |     |     |
|           |               | 1st             | 2nd | 3rd | 4th | 5th | 6th | 7th | 1st            | 2nd | 3rd | 4th | 5th | 6th | 7th |
| Comida    | 100           | 148             | 106 | 102 | 94  | 101 | 92  | 87  | 116            | 96  | 91  | 98  | 91  | 87  | 90  |
| No comida | 100           | 139             | 108 | 106 | 99  | 98  | 93  | 92  | 108            | 98  | 95  | 91  | 90  | 90  | 94  |

(\*) Ajustado para los desequilibrios en la entrega del diarios durante los días de labor.

Fuente: Pearl (1979)

Incluso si las diferencias significativas entre los diferentes períodos del diario que se encontraron en varios estudios fueron todas debidas a los efectos de la recogida de datos (aunque algunas podrían ser causadas por efectos reales que no se eliminaron por completo), no fue posible decidir si tales diferencias introdujeron sesgos en las estimaciones de la encuesta ni en qué dirección. Así que, las conclusiones derivadas de estos resultados no son claras. Incluso si las explicaciones más probables -el efecto telescopico en el primer día y algunos abandonos debido al cansancio más adelante- son aceptables, no se puede hallar una solución sencilla. Una aproximación es no utilizar los registros del primer día para la estimación, o recoger información "delimitadora" por memoria durante varios días anteriores, que tampoco se usa en la estimación. Sin embargo, siempre existe un riesgo de incrementar el sesgo general al intentar reducir una sola fuente de sesgo. Por lo tanto, quizás un método mejor sería retener los días de gasto alto con la justificación de que cualquier exageración de compra de esa fuente, compensaría las omisiones de compra en los días siguientes. Hace falta más investigación para llegar a unos resultados más conclusivos, respecto a la existencia de tales sesgos en las estimaciones de la encuesta y para determinar los cambios necesarios en el diseño de la encuesta para superar los sesgos.

### 3.3. Varianzas de Muestreo de Diarios de Diferente Duración

Aunque los errores muestrales sólo constituyen una parte del total de errores de las estimaciones de la encuesta, su papel puede incrementar, con estimaciones más detalladas y tamaños de muestra más pequeños. Cuando se estudien las diferencias entre los grupos o entre dos encuestas subsiguientes, los errores muestrales podrán ser dominantes, dado que los sesgos de magnitudes similares podrían anular los unos a los otros. Así que, en determinar la duración más eficiente del período de la recogida de datos en los diarios, hay que considerar los efectos sobre la varianza de muestra comparados con el coste.

A este propósito, vamos a examinar la eficiencia de las diferentes duraciones de los períodos del diario, cada una considerada como un período continuo de registración. Sin perder la generalidad, vamos a suponer para sencillez, que la muestra sea una muestra simple aleatoria y que se distribuya igualmente entre períodos del diario no solapados.

Vamos a denotar por  $p$  el período del diario del submuestra  $p$  ( $p=1,2,\dots,T$ ) donde  $T$  es el número de diferentes períodos del diario de misma duración durante el período de la encuesta, digamos un año. Así que, por ejemplo,  $T=12$  para un diario mensual,  $T=24$  para un diario de medio mes,  $T=26$  para un diario de dos semanas, etc. Deja que  $n(T)$  denote el tamaño total de la muestra, y  $n_p(T)=n(T)/T=n(T)$  el tamaño de cada submuestra, para un período del diario que corresponde a  $T$ .

Deja que  $X_{ip}$  sea el valor de la variable a investigar (por ejemplo, gasto) de la unidad  $i$  en el período  $p$ .

La estimación sencilla de la media por unidad en el período p correspondiente a T es

$$\bar{x}_p(T) = \frac{1}{n} \sum_{t=1}^T x_{tp}$$

y la estimación sencilla de la media por unidad durante el período entero es

$$\bar{x}(T) = \frac{1}{N} \sum_{p=1}^P \bar{x}_p(T).$$

La varianza de  $\bar{x}(T)$  bajo las asunciones anteriores es

$$\text{Var}[\bar{x}(T)] = \frac{1}{N} \sum_p \text{Var}[\bar{x}_p(T)] = \frac{1}{N} \sum_p \sigma_p^2(T) / n(T), \text{ when } 1/n(T)/N = 1.$$

Vamos a denotar por  $c(T)$  el coste de investigación por unidad de la muestra para un período de agenda que corresponde a T. Así que,  $n(T)=C/c(T)$  para un total fijo de coste de investigación, C. Así que, la duración óptima para la agenda para un coste total fijo es la que

$$\text{Var}[\bar{x}(T)] = T c(T) \sum_p \sigma_p^2(T) \text{ es mínima}$$

La eficiencia de la duración del período de una agenda basada en  $T^*$  comparada con una basada en T es

$$E(T^*/T) = \frac{\text{Var}[\bar{x}(T^*)]}{\text{Var}[\bar{x}(T)]} = \frac{T^*}{T} \left[ \frac{c(T)/c(T^*)}{\sum_{p=1}^T \sigma_p^2(T) / \sum_{p=1}^{T^*} \sigma_p^2(T^*)} \right].$$

Como ejemplo, vamos a considerar dos duraciones de un período de uso de agenda:

- (1) Una agenda de mes entero, donde  $T=12$  y p denota el mes de calendario
- (2) Una agenda de medio mes, donde  $T^*=24$  y  $p(1)$  y  $p(2)$  denotan la primera y segunda mitad del mes p, respectivamente.

En este caso  $E(T^*/T) = \frac{1}{12} [c(12)/c(24)] [1+\delta]$  donde,

$$1+\delta = \frac{\sum (\sigma_{p(1)}^2 + \sigma_{p(2)}^2) + 2\text{Cov}[p(1), p(2)]}{\sum (\sigma_{p(1)}^2 + \sigma_{p(2)}^2)}$$

y  $\delta$  es aproximadamente la correlación entre los gastos de las dos mitades de todos los meses (asumiendo que  $\sigma_{p(1)}^2 = \sigma_{p(2)}^2 = \sigma^2$  for all p). para todo p).

Aunque el coste por unidad para una agenda de medio mes es menor que para una agenda de mes entero, es más que la mitad. Por otra parte, la varianza entre los meses cuando se usa una agenda de mes entero será mayor que cuando se usa dos agendas de medio mes cada una para las dos submuestras independientes, sólo si la covarianza entre las dos medio meses es positiva. Se supone que sea el caso para los gastos frecuentes, como artículos de comida corrientes. Sin embargo, puede ser diferente en el caso de gastos poco frecuentes, donde se podría esperar una covarianza negativa.

Si realizó en Israel un análisis de la relativa eficiencia de tales dos períodos de agenda diferentes, basado en una muestra de 2.250 familias en la EGF de 1975-76 (Kantorowitz, 1981). Se compararon una agenda de medio mes con una de mes entero que se usaba en la encuesta. Aunque hubiera sido mejor haber comparado una agenda de

dos semanas (por varias razones, es preferible que una de medio mes), los datos disponibles no permitían la disimulación requerida para la estimación de las varianzas en la muestra.

Las varianzas muestrales para las dos alternativas se estimaban según las siguientes asunciones: (1) los efectos de diseño de la muestra son similares para los diferentes grupos de artículos para cada método; (2) la muestra neta está igualmente distribuida sobre todos los períodos del diario; (3) los posibles cambios en la forma de registro en los diarios por parte de los encuestados entre las dos mitades del mes no afectan a las varianzas estimadas; (4) los ratios y características de la no respuesta son iguales para las dos alternativas; y (5) no hay otras diferencias en el método de investigar.

Se llevó a cabo un análisis detallado del coste de los dos métodos, teniendo en cuenta muchos componentes, como el coste de la investigación en las varias etapas, viajes y dietas, edición, codificación, etc. Algunos de los componentes del coste para el período de medio mes sólo se podrían asesorar, porque nunca se había llegado a usarlos. A partir de este análisis de coste-función, se estimaba que el ratio del coste por encuestado para un diario de mes entero a una de medio mes era 1.45 a 1 (incluyendo los costes de las entrevistas)

La eficacia del diario de medio mes relativa a el diario de mes entero para estimaciones del gasto medio por familia se estimaba como el ratio inverso de las respectivas varianzas de la muestra para muestras del mismo coste. Esto se hizo tanto para grupos principales como subgrupos de artículos, para el total de la población, así como para algunos subgrupos principales de la población.

El método de medio mes sería preferible cuando la eficiencia sea mayor que 1. Las eficiencias calculadas, sólo usando los datos de los diarios,  $E(k)$ , para cada grupo  $k$  son relevantes a las estimaciones de gasto derivadas sólo de los datos de los diarios. Para otros grupos de gasto, hay que considerar una eficiencia general que tiene en cuenta la contribución de los datos de los cuestionarios también. Para las estimaciones que se basan tanto en el diario como en el cuestionario, la eficiencia general,  $\epsilon(k)$ , también se calculó, teniendo en cuenta el porcentaje de la estimación que se basa en el diario ( $k$ ). se debería notar que esto se hizo para el período de memoria que se usaba en la encuesta, es decir, 12 meses para la mayoría de los datos recogidos en el cuestionario. Por supuesto, un período diferente de memoria, podría causar resultados diferentes. Los resultados de este análisis para los principales grupos de artículos para el total de la población se presentan en la Tabla 3.6.

**Tabla 3.6.**  
**Eficiencia del diario de Medio Mes Relativa**  
**a el diario de Mes Entero**  
**EGF 1975-76 de Israel**

| Principal grupo de gasto (k) | % de estimación basado en diario $\pi(k)$ | Correlación entre dos medio meses $\delta(k)$ | Eficiencia de estimación del diario E(k) | Eficiencia de la estimación total $\epsilon(k)$ |
|------------------------------|---|---|--|---|
| <b>GASTO TOTAL</b>           | <b>48.5</b>                               | <b>0.49</b>                                   | <b>1.18</b>                              | <b>1.36</b>                                     |
| Comida (excl. verd.)         | 100.0                                     | 0.56  | 1.26                                     | 1.26  |
| Verduras y Fruta             | 100.0                                     | 0.55  | 1.25                                     | 1.25  |
| Ropa y Calzado               | 100.0                                     | 0.17  | 0.86                                     | 0.86  |
| Varios Bienes y Srv.         | 50.3                                      | 0.32  | 1.00                                     | 1.28  |
| Mantenimiento hogar          | 41.5                                      | 0.25  | 0.93                                     | 1.12  |
| Transporte y comun.          | 40.7                                      | 0.28  | 0.96                                     | 1.31  |
| Educación, Cultura, Ocio     | 34.3                                      | 0.31  | 0.98                                     | 1.35  |
| Muebles y Equip. hogar       | 20.7                                      | 0.09  | 0.80                                     | 1.35  |
| Salud                        | 10.2                                      | 0.25  | 0.93                                     | 1.39  |
| Vivienda                     | 0.0                                       | -   | -  | 1.45  |

Se encontró una correlación positiva  $\delta(k)$ , entre los gastos en el diario durante los dos medio meses para el Gasto Total, para todos los grupos principales y para casi todos los subgrupos. Sin embargo, para muchos subgrupos, la correlación fue bastante cerca al cero y se espera que para grupos de gasto más detallado la correlación podría ser negativa.

La eficiencia de  $E(k)$ , cuando sólo se usaban los datos del diario para las estimaciones (que es el caso para Comida, Verduras y Fruta, Ropa y Calzado, así como para algunos subgrupos), resultó ser mayor que 1 para el Gasto Total (1.18), pero sólo para dos grupos principales: Comida (1.26) y Verduras y Fruta (1.25). Excepto para la mayoría de los subgrupos dentro de estos dos grupos principales, sólo para unos pocos subgrupos  $E(k)>1$ , por ejemplo, Tabaco ( $E(k)=1.31$ ), ó para Transporte Público ( $E(k)=1.09$ ).

Como hemos mencionado antes, el ratio del coste por encuestado en el mes entero del diario al medio mes era 1.45. Así que, al reducir el período del diario de un mes a medio mes, se puede incrementar el tamaño de la muestra en 1.45 para un coste dado. Esto significa que para estimaciones de gasto que se basan sólo en los cuestionarios,  $\epsilon(k)=1.45$ . Es de considerable importancia, dado que la mayoría de las estimaciones de este tipo normalmente tienen una varianza muestral mayor que aquellas basadas en el diario.

La eficiencia general para Total gasto era  $\epsilon(k)=1.36$  y para todos los grupos principales  $\epsilon(k)>1$ , exceptuando Ropa y Calzado donde  $\epsilon(k)=0.86$ . La eficiencia para los subgrupos de gasto dentro de los grupos principales (no presentado aquí) era normalmente menor, exceptuando 10 subgrupos que se basaban casi por completo en el cuestionario con  $\epsilon(k)=1.45$ . Por ejemplo, mientras que para Comida o para Verduras y Fruta  $\epsilon(k)=1.26$ , para los 18 subgrupos dentro de

estos dos grupos  $0.91 \leq \epsilon(k) \leq 1.28$ ; para Transporte y Comunicaciones  $\epsilon(k)=1.31$ , mientras que para sus subgrupos  $1.07 \leq \epsilon(k) \leq 1.41$ .

En total, en los 62 subgrupos, para 10 subgrupos  $\epsilon(k)=1.45$ , para 10 subgrupos  $1.30 \leq \epsilon(k) < 1.45$ , para 15 subgrupos  $1.10 \leq \epsilon(k) < 1.30$ , para 7 subgrupos  $1.00 \leq \epsilon(k) < 1.10$ , y para 22 subgrupos  $\epsilon(k) < 1.00$  (10 de ellos de Ropa y Calzado).

Las estimaciones para el grupo principal de Ropa y Calzado y sus subgrupos tienen unas varianzas muestrales relativamente altas, por lo tanto, reducir el período del diario sólo las haría incluso menos eficientes. Por lo tanto, los datos sobre gasto en Ropa y Calzado se deben recoger durante un período de referencia más largo, usando un diario separada, ó bien por memoria.

Se podría calcular una medida generalizada para la eficacia de los dos períodos de los diarios, para todos los grupos de gasto. Se pueden usar tres tipos de criterios:

(1)  $\phi^2_1 = \sum_k \sigma^2(k)$ , la suma de todas las varianzas, que depende de las cantidades absolutas gastadas en cada grupo de artículos  $k$ , así dando más importancia a los grupos con varianzas más altas y probablemente gastos más altos.

(2)  $\phi^2_2 = \sum_k \sigma^2(k)/x^2(k)$ , la suma de todas las varianzas relativas, lo que atribuiría a cada grupo igual importancia, irrespectivo de su tamaño.

(3)  $\phi^2_3 = \sum_k \sigma^2(k)/\bar{x}(k)$ , que es un término medio entre los otros dos criterios.

Basado en cada una de estas tres funciones para cada uno de los períodos del diario, se calculaba una eficiencia media general  $\epsilon^*$  y se obtenían los siguientes resultados:

Para los 10 grupos principales:  $\epsilon^*_1=1.27$   $\epsilon^*_2=1.24$   $\epsilon^*_3=1.22$   
Para los 62 subgrupos:  $\epsilon^*_1=1.27$   $\epsilon^*_2=1.15$   $\epsilon^*_3=1.13$

A pesar de la ineficacia de reducir el período del diario para Ropa y calzado, se logró una eficiencia general, cualesquiera de los tres criterios se usara. Así que, basándose en consideraciones de varianzas muestrales, el diario de medio mes resultó ser más eficaz que una de mes entero.

Se apoyó la conclusión de que el período del diario se debería reducir por temor de los sesgos que se esperaban de varias fuentes con un período demasiado largo. Sin embargo, el problema con Ropa y Calzado entonces será aún más serio.

Se debería notar que el comportamiento diferente de gasto de las familias puede dar como resultado una imagen totalmente diferente. Por ejemplo, mientras que en el pasado reciente, la mayoría de la comida fresca como pan, leche y productos lácteos, etc. se compraban muy a menudo (incluso a diario), hoy en día, con el uso de frigoríficos y congeladores, este tipo de comida no se compra con tanta frecuencia. Además, las formas de gasto pueden diferir entre los países, y así la duración óptima del diario puede variar también.

entre países. Sin embargo, es interesante señalar que se recomendó el uso de un diario de dos semanas en muchos países, basándose en estudios similares.

### 3.4. Tipos de Diario

Se pueden usar dos tipos principales del diario en las EGF:

- (a) Un diario de tipo abierto
- (b) Un diario de tipo estructurado

Esta distinción se relaciona principalmente con el orden de los registros en el diario de varios gastos, y no con la forma de registrar los detalles de cada gasto, dado que la forma siempre está estructurada, es decir, la información requerida para cada gasto se registra en columnas preespecificadas en el diario (descripción del artículo, cantidad, pagos, etc.). En un diario de tipo abierto, se registran las compras en el orden en que se hicieron. Normalmente, una lista de productos y servicios (no muy completa) se adjunta de alguna manera a el diario, por ejemplo, impreso en la primera página del diario. Los diarios de tipo estructurado pueden estar estructuradas en varios niveles y según las varias clasificaciones. De éstas, lo que más se usa es un diario de productos, aunque también se puede considerar un diario de salidas. Los diarios de tipo abierto pueden utilizar páginas separadas para cada día, pero las páginas diarias podrían ser ineficientes si el diario está estructurada muy específicamente.

Los registros en los diarios de salidas se realizan según el tipo de almacén donde se hizo la compra, por ejemplo, alimentación, gasolinera, drojería, restaurante, etc. Así que, el diario está estructurada para que en cada página del diario, o partes de páginas, se usan líneas preespecificadas para el registro de los diferentes artículos en cada tipo de salida.

Un diario de productos está estructurada de una forma parecida, pero el diario está organizada por grupos de productos, tales como productos lácteos, carne, pescado, fruta, pan, ropa, suministros para la casa, etc. Se pueden emplear diferentes niveles de estructura, desde unos pocos grupos amplios a muchos grupos detallados. Para cada grupo de productos, hay líneas en blanco debajo de cada cabecera de sección para registrar las compras. El diario más estructurada es el tipo cerrado, donde para cada grupo de productos hay una línea para un artículo específico con una descripción del mismo ya escrita. Dado que el diario no puede incluir tantas líneas como productos existentes, se dejan líneas en blanco adicionales en cada grupo de productos.

El diario estructurada ayuda al encuestado a no omitir ningún gasto (o recibo) y puede ayudar a distinguir mejor entre los tipos de producto. También, con un diario de este tipo, es más fácil estructurarla para incluir información adicional sobre artículos especiales, por ejemplo, la edad y sexo para artículos de ropa para poder clasificar mejor los gastos.

Cuando se uso un diario de productos, se puede simplificar la codificación empleando códigos pre-impresos. Por otra parte, un gran

número de cabeceras puede confundir al encuestado, y puede tardar más buscando la línea apropiada. Los errores por parte de la familia en la línea de producto pueden llevar a una codificación incorrecta. Además, los encuestados pueden estar influenciados por las cabeceras de las listas y omitir de registrar gastos no especificados en el diario. El diario de salidas también puede llevar a errores de clasificación, sobre todo para los almacenes que venden comida, drojería y ropa a la vez.

Se realizaron experimentos preliminares en el Reino Unido en los años cincuenta con diferentes tipos de diarios (Kemsley y Nicholson, 1960). Los resultados de estos experimentos no mostraron diferencias significativas en el ratio de la respuesta, ni en el gasto medio registrado. Por lo tanto, dado que implica menos trabajo de codificación si los registros en el diario estén agrupados bajo cabeceras generales de gasto, se recomendó el uso de los diarios estructurados antes que los de tipo abierto.

Un estudio piloto realizado para la Oficina de Estadísticas Laborales de los Estados Unidos (Sudman y Ferber, 1971), incluyó un intento de explorar la efectividad de los tres tipos de diario:

- (a) Diario de tipo abierto;
- (b) Diario de tipo salidas;
- (c) Diario de tipo grupos de productos.

Una muestra de 525 familias elegibles se dividió entre los tres tipos, la mitad de ellas para el tipo salidas, y una cuarta parte para cada uno de los otros dos tipos. Cada una de estas submuestras se dividió, además, en cuatro submuestras, de aproximadamente el mismo tamaño, y se pedía a familias en cada grupo que se mantuvieran un diario durante 1 semana, 2 semanas, 3 semanas ó 4 semanas. Los resultados principales de este estudio se presentan en la Tabla 3.7.

**Tabla 3.7**  
**Efecto del Tipo del diario sobre Porcentajes**  
**de Cooperación(\*)**

| Grado de cooperación        | Total | Tipo de Diario |        |          |
|-----------------------------|-------|----------------|--------|----------|
|                             |       | Abierto        | Salida | Producto |
| Mantenía al menos 1 diario  | 87    | 79             | 81     | 84       |
| Mantenía al menos 2 diarios | 81    | 55             | 69     | 76       |
| Mantenía al menos 3 diarios | 66    | 46             | 56     | 68       |
| Mantenía al menos 4 diarios | 43    | 24             | 47     | 55       |

(\*) Porcentajes de todas las familias que mantenían un diario durante al menos ese período de tiempo.

Fuente: Sudman y Ferber (1971)

Aunque la muestra era relativamente pequeña, los resultados indicaron que al usar un diario de tipo abierto se obtenía menos cooperación y parece que los encuestados encuentran la tarea más fácil cuando el diario tiene algún tipo de estructura, con una preferencia para el tipo producto antes que para el tipo salidas.

Como resultado de este estudio, desde la EGF de 1972-73 en los EE.UU., se ha empleado un diario estructurada de tipo producto con un período de registro de dos semanas.

A continuación hay ejemplos de tipos del diario en uso en otros países. En el Reino Unido, una página doble para cada día se divide en 10 categorías principales y dentro de cada categoría, el mismo encuestado tiene que describir lo que compró. Un diario de productos con una estructura parecida se usa para la Encuesta de Gasto en Comida de Canadá. En Italia, una página doble para cada día se divide en cuatro categorías de 62 líneas, de las cuales 45 contienen cabeceras de gasto pre-impresas y no se puede añadir más descripción de los artículos. En Grecia, se usa un diario cronológica de tipo abierto para la familia entera, mientras que un diario personal individual (también empleada) se divide en 9 grupos principales de pequeños gastos (restaurantes, cafeterías, alcohol, tabaco, teatro, etc.) con un número diferente de líneas bajo cada categoría. En Israel, se usa un diario de tipo abierto (con una lista de productos impresa en la tapa posterior del diario), con páginas separadas para cada día. Cada página se divide en dos partes: la primera para todos los gastos comprados por la familia y la segunda para todos los servicios y bienes recibidos por la familia.

## CAPITULO 4. EL CUESTIONARIO

Como se mencionaba en el Capítulo 2, las estimaciones de las EGF casi siempre se basan tanto en los datos de los diarios como en los datos suplementarios que se recogen por medio de uno o más cuestionarios retrospectivos, relacionados con un período de referencia dado.

### 4.1. El Período de Referencia

En algunas encuestas, el período de referencia es el mismo para todas las familias de la muestra, por ejemplo, se entrevistan a todas las familias poco después del final de la encuesta anual del diario y el período de referencia es el año calendario anterior (como en Suecia, Canadá, y otros países). Otro diseño de encuesta se basa en períodos de referencia "móviles", donde para cada submuestra se usa un período de referencia diferente durante el año de la encuesta, digamos, los últimos tres meses o los últimos doce meses.

La duración del período de referencia del cuestionario (no necesariamente igual para todos los grupos de artículos), como la del diario, puede afectar la fiabilidad de las estimaciones de la encuesta. Las varianzas muestrales aumentarán para un tamaño neto fijo de muestra si el período de la muestra es corto y si la variabilidad es alta entre períodos parciales de las mismas familias. Dado que para la mayoría de los artículos del cuestionario esta variabilidad es mayor que aquella de los artículos del diario, la duración del período de referencia del cuestionario tiene un efecto mayor sobre las varianzas muestrales. También es el caso con la fiabilidad de las estimaciones de distribuciones y parámetros de variabilidad de la población. Cuanto más corto sea el período de referencia, probablemente más alto será la variabilidad estimada, y la caracterización de las familias por su nivel anual de ingreso y gasto será propensa a mayores discrepancias. Por otra parte, mientras que con un período de referencia más corto del diario la ganancia en el coste de la investigación es sustancial, con el cuestionario es marginal. Así que sólo se puede incrementar ligeramente el tamaño de la muestra con un período de referencia más corto, para el mismo coste general.

Con un período de referencia más largo, se espera un incremento en el ratio de la no respuesta debido a la mayor carga que se impone a los encuestados, pero no tanto como con un período más largo de registro en un diario.

Igual que con todos los datos recogidos en un cuestionario retrospectivo, los efectos de memoria pueden causar sesgos en las estimaciones de la encuesta y el ratio de la no respuesta puede incrementarse debido a los problemas de mala memoria, lo que requiere más imputaciones. Los efectos de memoria, a causa de la mala información suministrada por los encuestados, o a causa de la omisión de registrar algunos artículos, o el registro de acontecimientos que no ocurrieron durante el período de referencia (es decir, efecto telescopico), dependen principalmente de la duración del período de hacer memoria.

De todos estos factores afectados por la duración del período de memoria, la mayoría de los efectos de memoria (y, hasta cierto punto, consideraciones de la no respuesta y del coste), favorecen una reducción de la duración del período, mientras que otros factores (y, a veces, ciertos efectos telescopicos), favorecen lo contrario.

Dado que el período de referencia no tiene que ser igual que el período de recuerdo, es decir, se puede dividir en varios períodos de recuerdo que cubren el período entero de referencia, la solución es reducir el período de memoria, pero manteniendo un período más largo de referencia, digamos un año, llevando a cabo varias entrevistas durante el período de la encuesta. Este tipo de diseño de encuesta se usa en EE.UU. en la EGF actual, donde se realizan cuatro entrevistas de recuerdo de períodos de tres meses para cada familia. No obstante, con esta solución se incrementa significativamente el coste con el número de entrevistas por familia y también se espera un incremento en el ratio de la no respuesta. Además, con tal diseño de encuesta, pueden ocurrir cambios en el comportamiento de consumo de los encuestados, así introduciendo sesgos en las estimaciones de la encuesta.

Para poder decidir cuál es la duración óptima del período de hacer memoria y, para estudiar la eficacia (coste comparado con los efectos sobre las estimaciones de la encuesta) de realizar varias entrevistas en vez de una sola, se requiere la evaluación del efecto mutuo de todos los factores mencionados. No es siempre sencillo, porque no estén disponibles todos los datos pertinentes para tal evaluación, o no se puedan obtener a un coste razonable. Se presentan en las siguientes secciones varios estudios de evaluación.

#### 4.2. Efectos de Memoria

Los efectos de memoria dependen de varios factores, entre los cuales hay el tipo de variable investigada, las características del encuestado y los términos del cuestionario, pero el más significante es la duración del período de memoria. Cuanto más largo sea el período de recuerdo, menos fiable será la memoria del encuestado y serán menos exactos los resultados de la encuesta.

Se pueden distinguir dos tipos de errores de memoria:

- a) Efectos de memoria - Omitir registrar acontecimientos que el encuestado ha olvidado que lleva a un sesgo al menos en las estimaciones de la encuesta. Además, si el encuestado recuerda la ocurrencia de los acontecimientos, puede omitir los detalles exactos, que lleva a sesgos en cualquier dirección.
- b) Efectos telescopicos - Cuando el encuestado recuerda los acontecimientos, pero los coloca erróneamente en el tiempo. Pueden ser de dos tipos, un efecto telescopico hacia el futuro cuando se registran los acontecimientos más tarde que cuando ocurrieron, o hacia el pasado cuando se registran los acontecimientos como si hubieran ocurrido antes. El efecto telescopico afectará principalmente a la exactitud de las estimaciones de la encuesta cuando los acontecimientos que pasan fuera del período de memoria se incluyen dentro del período por

error, o vice versa, pero también cuando dentro de un período de memoria se confundelmomento cuando ocurrieron. Los efectos telescopicos pueden llevar tanto a demasiados registros como a pocos, según la dirección de los efectos telescopicos.

Probablemente los efectos telescopicos sean dominantes cuando se use un período de recuerdo corto, especialmente para acontecimientos poco frecuentes de considerable importancia. Los efectos de memoria probablemente sean más significativos cuando se use un período de memoria más largo, especialmente para acontecimientos frecuentes y menos importantes.

Así que, para reducir ambos tipos de efectos de memoria: (1) el período de recuerdo debería ser tan corto como sea posible; (2) se debería llevar a cabo una entrevista adicional para razones de delimitación antes del período de referencia. Los datos de delimitación no se usan para las estimaciones de la encuesta.

Sin embargo, esta solución (adoptada en la encuesta de los EE.UU.) no es tan simple en la práctica por ser demasiado costosa, porque se requiere una entrevista adicional para cada familia. Además, un período de referencia más corto puede llevar a un incremento significativo en las varianzas de la muestra, a menos que se incremente la muestra o el número de entrevistas por familia, es decir, la encuesta llega a ser más costosa. La solución de diseño de la encuesta, para recursos limitados, está en la obtención de estimaciones de los efectos de memoria para varias duraciones de períodos de recuerdo y, al mismo tiempo, el asesoramiento de los efectos sobre otros factores antes mencionados.

Un estudio sobre los efectos del recuerdo se realiza actualmente en Israel, sobre los datos recogidos en las entrevistas de la EGF de 1986-87 (basado en una muestra de 5.000 familias). A continuación se presentan algunos resultados preliminares. En esta encuesta, se usaba un período de memoria de 3 meses para la mayoría de los gastos, y un período de 12 meses para algunos gastos poco frecuentes. No había ninguna entrevista de delimitación antes del período de memoria. Sin embargo, se pedía a cada familia de la muestra que informasen sobre los gastos desde el principio de los últimos 3 (ó 12) meses de calendario hasta el momento de la entrevista. Los datos referidos al tiempo después del final del último mes no se incluían en la encuesta, se usaban sólo para información de delimitación.

El período "movible" relacionado a 15 meses del calendario desde marzo 1986 hasta mayo 1987 (aunque no todos los meses se representan igualmente) como se ilustra en la Figura 4.1. Para cada uno de los 11 meses del calendario, mayo 86 - marzo 87, se disponen de datos para tres meses de memoria: (1) el mes más "reciente"; (2) el mes de "en medio"; y (3) el mes más "lejano". Para cada mes del calendario, los datos de los tres meses de recuerdo corresponden a diferentes submuestras de familias. Los datos de los cuatro meses del calendario a los dos extremos se excluían del estudio para evitar los posibles efectos de los momentos de las compras.

Figura 4.1.  
Los Meses de Memoria por el Mes de Entrevista  
para los Períodos de Memoria de 3 Meses

| Mes de Entrevista | Mes de recuerdo |      |      |      |      |      |      |       |       |       |      |      |      |      |
|-------------------|-----------------|------|------|------|------|------|------|-------|-------|-------|------|------|------|------|
|                   | 3/86            | 4/86 | 5/86 | 6/86 | 7/86 | 8/86 | 9/86 | 10/86 | 11/86 | 12/86 | 1/87 | 2/87 | 3/87 | 4/87 |
| 6/86(*)           | [3]             | [2]  | 1    |      |      |      |      |       |       |       |      |      |      |      |
| 7/86              |                 | [3]  | 2    | 1    |      |      |      |       |       |       |      |      |      |      |
| 8/86              |                 |      | 3    | 2    | 1    |      |      |       |       |       |      |      |      |      |
| 9/86              |                 |      |      | 3    | 2    | 1    |      |       |       |       |      |      |      |      |
| 10/86             |                 |      |      |      | 3    | 2    | 1    |       |       |       |      |      |      |      |
| 11/86             |                 |      |      |      |      | 3    | 2    | 1     |       |       |      |      |      |      |
| 12/86             |                 |      |      |      |      |      | 3    | 2     | 1     |       |      |      |      |      |
| 1/87              |                 |      |      |      |      |      |      | 3     | 2     | 1     |      |      |      |      |
| 2/87              |                 |      |      |      |      |      |      |       | 3     | 2     | 1    |      |      |      |
| 3/87              |                 |      |      |      |      |      |      |       |       | 3     | 2    | 1    |      |      |
| 4/87              |                 |      |      |      |      |      |      |       |       |       | 3    | 2    | 1    |      |
| 5/87              |                 |      |      |      |      |      |      |       |       |       | 3    | 2    | [1]  |      |
| 6/87(*)           |                 |      |      |      |      |      |      |       |       |       |      | 3    | [2]  | [1]  |

(\*) El tamaño de la muestra de los meses al principio y al final era aprox. la mitad de los otros meses.

Para el análisis de tres meses de los efectos de memoria denotamos por  $y_{im}(r)$  el gasto registrado en un artículo por la familia  $i$  ( $i=1,2,\dots,n$ ) en el mes del calendario  $m$  ( $m=1,2,\dots,11$ ) que es el mes de memoria  $r$ -th de esa familia ( $r=1,2,3$  para los meses "reciente", "en medio" y "lejano", respectivamente).

Deja que  $x_{im}(r) = d_{im}(r) y_{im}(r)$ , donde

$d_{im}(r) = 1$  cuando  $y_{im}(r) > 0$

0 cuando  $y_{im}(r) = 0$

Para simplificar, vamos a suponer que una muestra sencilla aleatoria de tamaño  $n$  participe en cada mes y que se usen estimaciones sencillas, aunque de hecho, no era esto el caso en el análisis de los datos.

Las estimaciones del gasto medio por familia, la proporción de familias que registró gasto en el artículo y el gasto medio por familia de registro, en el mes  $m$  para el mes de memoria  $r$ , son:

$$\bar{x}_m(r) = \sum x_{im}(r)/n, \quad p_m(r) = \sum d_{im}(r)/n \quad \text{and} \quad \bar{y}_m(r) = \sum y_{im}(r)/n = \bar{x}_m(r)p_m(r).$$

Entonces los Efectos de Memoria (RE) sobre todos los 11 meses del calendario de los meses de memoria "en medio" y "lejano", comparados con el mes "reciente", se pueden estimar como:

$$RE(2) = \bar{z}_m(2) / \bar{z}_m(1) \quad \text{para el mes de memoria "en medio", y}$$

$$RE(3) = \bar{z}_m(3) / \bar{z}_m(1) \quad \text{para el mes de memoria "lejano"}$$

donde  $\bar{z}_m(r)$  puede ser  $\bar{x}_m(r)$ , or  $\bar{y}_m(r)$ , or  $p_m(r)$ .

También hay que estimar las varianzas de los efectos estimados de memoria, tomando en cuenta el modelo de estudio, para que se pueda aplicar un test-Z.

Algunos resultados del estudio israelí sobre los efectos de memoria por el mes de recuerdo, para un período de tres meses, se muestran en la Tabla 4.1. El análisis de los efectos de memoria sólo se hizo para artículos seleccionados del cuestionario, cuyas estimaciones tendían a una varianza relativamente baja.

**Tabla 4.1**  
**Efectos de Memoria por Mes de Recuerdo - Período de Tres Meses**  
**EGF 1986-87 de Israel**

| Grupo de gasto            | Recogida media mensual<br>Porcent. | Efecto de memoria mensual relativo al mes reciente |        |                          |        |             |        |
|---------------------------|------------------------------------|--|--------|--------------------------|--------|-------------|--------|
|                           |                                    | De % de Familias con gasto no cero                 |        | De Gasto medio           |        |             |        |
|                           |                                    |  |        | Por familia de respuesta |        | Por familia |        |
|                           |                                    | Mediano  | Lejano | Mediano                  | Lejano | Mediano     | Lejano |
| Todos artículos incluidos | 31.7                               | 0.92**   | 0.78** | 1.07                     | 1.30** | 0.98        | 1.00   |
| Joyería                   | 2.7                                | 0.61**   | 0.59** | 1.57**                   | 1.57** | 0.90        | 0.90   |
| Relojes                   | 2.1                                | 0.74**   | 0.59** | 1.04                     | 1.20   | 0.78*       | 0.77   |
| Edredones y mantas        | 1.1                                | 0.77*  | 0.69*  | 1.66*                    | 1.23   | 1.03        | 0.88   |
| Sabanas y toallas         | 2.3                                | 0.86   | 0.70** | 1.13                     | 1.23   | 0.98        | 0.79   |
| Reparaciones vivienda     | 3.3                                | 0.88*  | 0.54** | 1.27                     | 1.00   | 0.99        | 0.71*  |
| Servicio doméstico        | 4.2                                | 0.97   | 0.69** | 1.04                     | 0.71** | 1.02        | 0.66** |
| Muebles                   | 5.8                                | 0.99   | 1.01   | 0.92                     | 1.03   | 0.94        | 1.07   |
| Electrodomésticos         | 8.1                                | 1.00   | 0.80** | 0.98                     | 1.08   | 1.00        | 0.91   |
| Duraderos cultura y ocio  | 2.8                                | 0.92   | 0.82*  | 1.09                     | 1.29*  | 0.95        | 1.01   |
| Cuadros y marcos          | 0.8                                | 0.68*  | 0.56** | 1.13                     | 1.76*  | 0.74        | 0.92   |
| Odontología               | 8.8                                | 0.88**   | 0.50** | 1.01                     | 0.79*  | 0.88        | 0.56** |
| Médico particular         | 4.9                                | 0.99   | 0.52** | 1.29**                   | 1.45   | 1.28*       | 0.95   |
| Óptico                    | 3.7                                | 0.80**   | 0.62** | 1.27**                   | 0.93   | 0.98        | 0.80*  |
| Licencia vehículo         | 2.7                                | 0.77**   | 0.63** | 0.83*                    | 0.66** | 0.75*       | 0.52** |
| Mant. Rep. vehículo       | 6.6                                | 0.91   | 0.87*  | 1.06                     | 1.17   | 0.97        | 1.02   |

Nota: los efectos de memoria significativamente diferentes de 1 se denotan: \*\* para  $\alpha \leq 1\%$  y \* para  $1\% < \alpha \leq 5\%$ .

Para todos los artículos incluidos en el estudio, 35.2% de todas las familias informaron de al menos un artículo en el mes reciente, 32.4% en el mes de en medio y sólo 27.5% en el mes lejano, es decir, había efectos de memoria de 0.92 y 0.78, respectivamente. Se observaron ratios en disminución (no del mismo grado) del porcentaje de familias con gastos no de cero para casi todos los artículos, aunque no todos eran estadísticamente significativos. Para algunos artículos el efecto de memoria del mes de en medio es moderado, pero llega a ser sustancial en el mes lejano, por ejemplo, para

Odontología RE(2)=0.88 y RE(3)=0.50, para Reparación y renovación de la vivienda RE(2)=0.88 y RE(3)=0.54. Para otros artículos los ratios de registro disminuyen más gradualmente, por ejemplo, para Ropa de cama y toallas RE(2)=0.86 y RE(3)=0.70. Para algunos artículos la disminución principal en los ratios de registro se observa en los meses de en medio, por ejemplo, para Joyería RE(2)=0.61 y RE(3)=0.59. Para algunos artículos como Muebles, no se encontraron efectos de memoria.

Los resultados del gasto medio por familia indican efectos de memoria en la dirección opuesta para muchos artículos. Para el total de gastos recogidos de todos los artículos incluidos en el estudio, la media del mes lejano es significativamente más alta que en el mes reciente RE(3)=1.30. Para Joyería, la media de los meses en medio y lejano son aprox. 1.6 más que la media del mes reciente. Para Honorarios de médico particular, las medias de los meses de en medio y lejano son mayores que del mes reciente por 1.3 y 1.4, respectivamente. Un resultado al contrario ocurre con Licencias de vehículos, donde las medias son menores en el mes de en medio, 0.83, y en el mes lejano, 0.66, comparado con el mes reciente.

Mientras que la mayoría de los efectos de memoria en el ratio de registro fueron estadísticamente significativos, especialmente para el mes lejano, no es el caso con los efectos de memoria para el gasto medio por familia de registro. Esto es a causa de la alta variabilidad entre los gastos en el mismo tipo de artículo, o en diferentes artículos que pertenecen al mismo grupo de gasto. En consecuencia, las estimaciones para el efecto de memoria general correspondientes al gasto medio por familia también padecen de altas varianzas. Sólo para algunos artículos existe evidencia que la media más alta es en el mes reciente y la más baja en el mes lejano. Para la mayoría de los otros artículos, los efectos de memoria no son significativamente diferentes a 1, aunque se indica algún efecto de memoria.

Se debería notar que cuando se calculan las estimaciones de los efectos de memoria para los gastos medios por familia, hay que tomar en cuenta la diferencialidad de los cambios en precio de los diferentes artículos durante el período de memoria. Aunque todos los gastos durante el período entero de la encuesta se ajustaban según los cambios en el IPC general para que todos se relacionasen al mismo nivel de precio, no se ajustaban los cambios de precio para cada artículo por separado. Sin ajustar los datos según el IPC específico para cada artículo, los efectos de memoria medidos no eliminan la diferencialidad entre los cambios de precio de artículo. Sin embargo, durante un período corto de tres meses, el grado de inflación durante el período de la encuesta tenía menor importancia.

Se usaba un período de memoria de 12 meses para algunos artículos, como hemos mencionado anteriormente. Los períodos de memoria de las diferentes submuestras corresponden a 24 meses del calendario diferentes (junio 1985 a mayo 1987), como se ilustra en la Figura 4.2.

**Figura 4.2**  
**Meses de Memoria por Submuestras para**  
**Períodos de Memoria de 12 Meses**

| Mes de Entrevista | Mes de Memoria |         |         |         |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
|-------------------|----------------|---------|---------|---------|----|----|------|---|---|---|---|---|---|---|---|---|----|----|------|---|---|---|---|---|--|
|                   | 1985           |         |         |         |    |    | 1986 |   |   |   |   |   |   |   |   |   |    |    | 1987 |   |   |   |   |   |  |
|                   | 6              | 7       | 8       | 9       | 10 | 11 | 12   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12   | 1 | 2 | 3 | 4 | 5 |  |
| 6/86*)            | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 7/86              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 8/86              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 9/86              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 10/86             | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 11/86             | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 12/86             | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 1/87              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 2/87              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 3/87              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 4/87              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 5/87              | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |
| 6/87*)            | 12-11-10       | 9--8--7 | 6--5--4 | 3--2--1 |    |    |      |   |   |   |   |   |   |   |   |   |    |    |      |   |   |   |   |   |  |

\*) El tamaño de la muestra para estos dos meses era la mitad que de los otros meses.

Sólo dos meses del calendario (mayo y junio 1986) tienen 12 meses de recuerdo diferentes, dos meses del calendario tienen 11 meses de recuerdo, etc. Por lo tanto, una simple eliminación de una posible temporalidad como se hizo con el período de recuerdo de tres meses, no es posible para un período de recuerdo de doce meses. Incluso si se agrupan los meses de recuerdo en trimestres, sólo cuatro trimestres tienen los cuatro diferentes trimestres de recuerdo. Así que, hay que aplicar un modelo más sofisticado (todavía no adoptado) para analizar los efectos de memoria, utilizando la mayoría de los datos disponibles.

No obstante, una comparación preliminar de los datos entre los cuatro trimestres de recuerdo, con media de todas las submuestras, podría proporcionar alguna indicación de la existencia de los efectos de memoria, aunque habría que interpretar los resultados con precaución dado que no se eliminan los efectos de temporalidad. Para el gasto medio, se calcularon las comparaciones con y sin ajustes para la diferencialidad en los cambios de precio antes mencionados. Los resultados fueron afectados por este ajuste, pero no todos del mismo grado. La tabla 4.2. muestra los resultados para dos ejemplos, después del ajuste.

**Tabla 4.2**  
**Efectos de Memoria por Trimestre de Memoria**  
**Período de 12 Meses**  
**EGF 1986-87 de Israel**

| Grupo de gasto       | Recogida media mensual porcent. | Efecto trimestral de memoria relativo al trimestre reciente |         |        |                         |         |        |             |         |        |
|----------------------|---------------------------------|---|---------|--------|-------------------------|---------|--------|-------------|---------|--------|
|                      |                                 | De % de familias con gasto no cero                          |         |        | De gasto medio(*)       |         |        |             |         |        |
|                      |                                 |   |         |        | Por familia en recogida |         |        | Por familia |         |        |
|                      |                                 | Segundo   | Tercero | Cuarto | Segundo                 | Tercero | Cuarto | Segundo     | Tercero | Cuarto |
| Total                | 4.6                             | 0.80**  | 0.65**  | 0.47** | 1.12                    | 1.45**  | 1.41** | 0.89        | 0.95    | 0.61** |
| Ocio y excursiones   | 4.4                             | 0.80**  | 0.66**  | 0.47** | 1.13*                   | 1.39**  | 1.23** | 0.89        | 0.94    | 0.53** |
| Viajes al extranjero | 1.1                             | 0.86*   | 0.84*   | 0.62** | 1.05                    | 1.14*   | 1.18** | 0.90        | 0.96    | 0.73   |

(\*) Después de ajustes para la diferencialidad en los cambios de precio de los artículos.

Nota: efectos de memoria significativamente diferentes a 1 se denotan: \*\* para  $\alpha \leq 1\%$  y \* para  $1\% < \alpha \leq 5\%$ .

Se observaron efectos de memoria sustanciales (y estadísticamente significativos) para el porcentaje de familias que recogieron los gastos. Se informó del gasto total en Viajes al extranjero y Ocio y excursiones un 6.3% de las familias en el trimestre reciente, 5.0% en el segundo, 4.1% en el tercer trimestre y sólo un 3.0% en el trimestre más lejano. Los efectos de memoria son menos para los Viajes al extranjero que para ocio y excursiones, por ejemplo, los efectos de memoria son 0.62 y 0.47, respectivamente. Una parte del ratio relativamente alto de la recogida de estos artículos en el trimestre reciente se puede explicar por sesgos debidos a demoras de investigación en las familias que estaban ausentes en el momento que se designó originalmente para su investigación.

Los efectos de memoria del gasto medio por familia para el agregado de los dos grupos de gasto tienen un comportamiento similar a aquello de muchos de los artículos con un período de recuerdo de tres meses, es decir, cuanto más lejano sea el trimestre, más alto será la media: 1.12, 1.45, y 1.41 para el segundo, tercero y cuarto trimestre, respectivamente. También, casi es el caso para cada grupo por separado: para Viajes al extranjero los efectos son 1.05, 1.14 y 1.18 y para Ocio y excursiones 1.13, 1.39 y 1.23 para estos tres períodos de recuerdo. El efecto de memoria general sobre la media por familia muestra medias disminuyentes para los períodos de recuerdo más lejanos.

Los resultados de este estudio, incluso aunque no miden los efectos de memoria "puros", proporcionan evidencia de la existencia de graves efectos de memoria en los datos, para ambas duraciones del período de memoria. Por lo tanto, se requiere más análisis para

poder comprender mejor la naturaleza de los efectos de memoria sobre las EGF para intentar encontrar las formas de reducirlos en futuras encuestas.

También se analizaron los efectos de memoria en la Encuesta de Entrevistas sobre Gasto en Consumo de los EE.UU. en 1984, por tipo de encuestado, por ejemplo, tipo de familia, edad del encuestado, el uso de los registros, etc. (Silberstein, 1989). La encuesta de los EE.UU. se realizó con un diseño diferente que la encuesta israelí. La diferencia principal relativa a los efectos de memoria es el uso de una entrevista de delimitación anterior al período de memoria de tres meses, y, para cada uno de los períodos de memoria consecutivos de tres meses la entrevista de antemano sirve como una entrevista de delimitación. El período de la recogida incluye la parte del mes actual hasta el día de la entrevista y los gastos registrados para ese mes están disponibles al entrevistador durante la siguiente tanda para minimizar las duplicaciones. Este estudio incluyó unidades que participaron en el tercer trimestre de 1984, cubriendo los gastos entre abril y agosto y el análisis correspondió a las dos categorías de gasto, Ropa y Muebles de casa. Se estimaron los efectos de memoria haciendo comparaciones entre los tres meses de recuerdo de cada unidad, ignorando los posibles efectos de temporada para compras. Sólo se incluyeron los casos donde se había registrado al menos un artículo de estos categorías bajo análisis, y esto también podría causar sesgos en las estimaciones de los efectos de memoria, sin embargo, con varianzas reducidas.

El porcentaje de casos que registraron algunos gastos se muestran en la Tabla 4.3 por el mes de recuerdo, y los resultados muestran los efectos de memoria de una magnitud similar para ambas categorías, aprox. 15% menos de registros en el mes en medio y aprox. 25% menos en el mes lejano, comparados con el mes reciente.

**Tabla 4.3**  
**Porcentaje de Unidades de Consumo con**  
**Gastos Registrados por Mes de Recuerdo**  
**Encuesta de Entrevistas sobre Gasto en Consumo**  
**EE.UU., 1984**

| Categoría<br>de gasto | % de registros por<br>mes de recuerdo |         |        | % de registros relativo<br>al mes reciente |         |        |
|-----------------------|---------------------------------------|---------|--------|--|---------|--------|
|                       | Reciente                              | Mediano | Lejano | Reciente                                   | Mediano | Lejano |
| Ropa                  | 83                                    | 69      | 61     | 1.00                                       | 0.83    | 0.73   |
| Muebles               | 71                                    | 60      | 55     | 1.00                                       | 0.85    | 0.77   |

Fuente: Silberstein (1989)

Para más análisis de los datos, se asignaba cada unidad a uno de los tres grupos de efectos de memoria según el valor de los gastos recogidos en el mes reciente como un porcentaje del gasto total registrado para los tres meses de recuerdo. Se definían tres niveles del efecto de memoria: "ninguno" con 0% a 35%, "moderado" con 35% a 75%, y "grande" con 75% a 100%. Las distribuciones y los gastos medios por estos tres niveles se presentan en la Tabla 4.4.

**Tabla 4.4**  
**Gastos Medios por Niveles del Efecto de Memoria Individual**  
**Encuesta de Entrevistas sobre Gasto en Consumo**  
**EE.UU., 1984**

| Nivel del efecto de memoria | Ropa            |                     | Muebles         |                     |
|-----------------------------|-----------------|---------------------|-----------------|---------------------|
|                             | P % de unidades | Gasto medio mensual | P % de unidades | Gasto medio mensual |
| TOTAL                       | 100             | 102                 | 100             | 141                 |
| Ninguno                     | 44              | 106                 | 53              | 146                 |
| Moderado                    | 29              | 125                 | 18              | 163                 |
| Grande (*)                  | 27              | 71                  | 29              | 119                 |

\*) Aprox. la mitad registró todo el gasto en el mes reciente.

Fuente: Silberstein (1989)

El gasto medio en muebles es aprox. 1.4 veces más que en ropa, y esto, como se espera de los artículos de mayor gasto, puede llevar menos efectos de memoria: 53% sin efectos de memoria ("ninguno") contra 44%. La media más alta para gastos registrados para el nivel "moderado" de efectos de memoria en ambas categorías es bastante sorprendente.

Basado en los tres niveles de categorías de efectos de memoria para cada unidad, se pusieron modelos de log-lineales para medir el tipo de interacción de cada variable independiente, es decir, las características del encuestado, con la variable dependiente, es decir, el nivel del efecto de memoria, después de eliminar las interacciones entre las variables independientes. No se incluían más que tres variables independientes en cualquier de los modelos debido a limitaciones en el tamaño de la muestra (para más información, ver Silberstein, 1989).

Los resultados de este estudio (para ropa) muestran que los efectos de memoria son afectados por las características de los encuestados y por el rendimiento de la entrevista. A continuación describimos algunos ejemplos específicos de estos resultados. Se encontraron relaciones significativas de efectos de memoria para edad y educación. Los encuestados con un bajo nivel de educación son más propensos a tener grandes efectos de memoria. Los encuestados de edades 25-44 años muestran efectos moderados, mientras que los encuestados mayores (65 años y más) exhiben tanto ningún efecto como

grandes efectos. Sin embargo, no hay efectos de memoria evidentes para gente mayor cuando se controla el nivel de educación.

Es razonable el argumento que el tamaño de la familia influirá en los efectos de memoria, y de verdad, los resultados muestran que no había efectos de memoria en familias de 1 ó 2 miembros, había efectos moderados en familias con 3 ó 4 miembros, y se observaban efectos moderados a grandes en las familias de mayor tamaño. Los efectos de memoria difieren también según los ingresos de la familia, donde las familias con ingresos más bajos muestran mayores efectos de memoria.

#### 4.3. Duración del Período de Referencia y Varianzas del Muestreo

Muchas de las variables que son investigadas para cada familia de la muestra por el cuestionario de la EGF varían en el tiempo. Los gastos regulares, como pagos de electricidad, agua, alquiler, etc., y probablemente el ingreso varían menos que los gastos importantes poco frecuentes, como bienes duraderos y vacaciones. Para un período dado de encuesta, se puede considerar a cada familia de la muestra como un cluster de subperiodos, por ejemplo, meses o trimestres. El período de referencia del cuestionario para cada familia en la muestra puede ser el período entero de la encuesta, generalmente 12 meses, o una muestra de uno o dos subperiodos de al menos un mes, y no tiene que ser de la misma duración para los diferentes artículos.

Las diferentes duraciones del período de referencia para un tamaño fijo de muestra de familias tienen diferentes implicaciones en la varianza muestral, y estas dependen de la correlación intra-cluster entre los gastos de los subperiodos dentro del período de un año.

Vamos a denotar por  $X_{ij}$  el valor de una cierta variable investigada en la familia  $i$  ( $i=1,2,\dots,N$ ) en el mes  $j$  ( $j=1,2,\dots,M$ ).

Deja que  $S^2_b$  sea la varianza de los valores medios mensuales entre las familias  $N$ , y  $S^2_w$  la varianza entre los valores mensuales en las familias:

$$S^2_b = \sum (\bar{X}_i - \bar{\bar{X}})^2 / (N-1) \quad \text{and} \quad S^2_w = \sum \sum (X_{ij} - \bar{X}_i)^2 / N(M-1)$$

donde  $\bar{X}_i = \sum X_{ij} / M$  es la media mensual para la familia  $i$ , y  $\bar{\bar{X}} = \sum \sum X_{ij} / MN$  es la media mensual general por familia.

La varianza total entre las familias y los meses es

$$S^2 = \sum \sum (X_{ij} - \bar{\bar{X}})^2 / (NM-1) \approx (1/M)S^2_b + (1-1/M)S^2_w.$$

Por lo tanto, para una muestra simple aleatoria de  $n$  familias y  $(N-n)/N=1$ , con un período de referencia de todos los meses  $M$ , (muestreo de una etapa) la varianza muestral es

$$\text{Var}_M(\bar{\bar{X}}) \approx S^2_b / n.$$

Cuando para cada familia de la muestra, se selecciona una muestra de  $m < M$  meses aleatoriamente para el período de referencia (muestreo de dos etapas), la varianza muestral es

$$\text{Var}_m(\bar{\bar{X}}) \approx S^2_b / n + (1/m - 1/M) S^2_w / n.$$

Entonces, para una muestra de familias de tamaño fijo, con un muestreo de dos etapas, la varianza muestral será mayor, cuanto más alto sea la variabilidad entre meses en las familias y cuanto más pequeño sea  $m$ .

El incremento relativo en la varianza muestral al reducir el período de referencia de  $M$  a  $m$  será:

$$\alpha = \text{Var}_m(\bar{X}) / \text{Var}_M(\bar{X}) \approx 1 + (1/m - 1/M) S^2_t / S^2_b.$$

Sin embargo, en las EGF se usa un período de referencia de  $M=12$  meses, o un subperiodo  $t$  constando de  $m$  meses consecutivos de subperiodos de  $T=M/m$ , por ejemplo, un trimestre de cuatro, o un período de dos meses de seis meses. En este caso:

$$\alpha = 1 + (1-1/T) S^2_w / S^2_b, \text{ donde } S^2_w \text{ es la varianza interior entre los subperiodos } T.$$

Un extremo es cuando los gastos se distribuyen igualmente sobre los subperiodos de la encuesta. Entonces  $S^2_w = 0$  y así  $\alpha=1$ , por ejemplo, la duración del período de referencia se pudo reducir sin incrementar la varianza muestral. Sin embargo, rara vez es el caso con las variables de investigación en el cuestionario de las EGF.

El otro extremo son los artículos que se compran sólo una vez en cada familia, y que no se compran en el período de un año, como la mayoría de los bienes duraderos caros. Se puede mostrar que el incremento relativo en la varianza debido a la reducción de la duración del período de referencia para estas variables bajo la asunción de un gasto igual por artículo, será  $\alpha=1+(T-1)/(1-\pi)$ , donde  $\pi > 0$  es la probabilidad de que una familia compre el artículo en el año de la encuesta. Por ejemplo, el incremento en la varianza muestral al reducir el período de referencia de 12 meses a 3 meses para las variables de este tipo  $\alpha=1+3/(1-\pi)$ . Por lo tanto, como se puede esperar, la varianza para los artículos con este comportamiento de compra para un período de tres meses sería al menos cuatro veces mayor que en un período de 12 meses, y el incremento relativo de la varianza incrementa con  $\pi$ .

Sin embargo, en la práctica las variables de investigación del cuestionario de la EGF están entre estos dos extremos. A continuación se presentan algunas comparaciones empíricas de las varianzas para varias duraciones de períodos de referencia.

Las varianzas muestrales relativas de las EGF de Israel, 1979-80 y 1986-87, se compararon para varios grupos importantes de gasto, con datos obtenidos, en su mayor parte, de los cuestionarios, y para ingresos. En la EGF de 1979-80, el período de referencia para estos datos era de 12 meses, mientras que en la encuesta de 1986-87 se reducía a 3 meses. Los resultados de esta comparación (después de ajustes para el tamaño de muestra diferente en las dos encuestas) se presentan en la Tabla 4.5.

Estos resultados pueden proporcionar sólo una indicación del efecto de reducción del período de referencia sobre las varianzas muestrales. Porque las dos encuestas no corresponden al mismo tiempo, cambios en el comportamiento de consumo pueden afectar a los

resultados, y también porque estas encuestas no se realizaron bajo las mismas condiciones con el mismo diseño de encuesta, la comparación tiene una validez limitada.

**Tabla 4.5**  
**Comparación de Varianzas Muestrales Relativas para**  
**Períodos de Referencia de 12 meses y de 3 meses**  
**EGF de 1979-80 y 1986-87, Israel**

| Grupo principal de gasto               | % de familias en recogida 1986/87 | % de gasto del cuestionario | Varianza relativa muestral - 3 meses /12 meses ( $\alpha$ ) |
|--|-----------------------------------|-----------------------------|---|
| <b>Ingreso Total</b>                   | 100                               | 100                         | 1.2   |
| <b>Ingreso del trabajo</b>             | 76                                | 100                         | 1.6   |
| <b>Ingreso de pensiones</b>            | 16                                | 100                         | 1.1   |
| <b>Gasto Total (*)</b>                 | 100                               | (*) 55                      | 1.3   |
| <b>Seguridad Social</b>                | 94                                | 100                         | 1.1   |
| <b>Joyería y relojes</b>               | 22                                | 72                          | 1.3   |
| <b>Electrodomésticos</b>               | 41                                | 91                          | 1.9   |
| <b>Duraderos de cultura y ocio</b>     | 83                                | 81                          | 2.1   |
| <b>Sabanas y toallas</b>               | 17                                | 78                          | 2.2   |
| <b>Muebles</b>                         | 17                                | 99                          | 2.9   |
| <b>Odontología</b>                     | 28                                | 100                         | 2.9   |
| <b>Reparaciones vivienda</b>           | 67                                | 83                          | 4.1   |
| <b>Cuotas y donaciones</b>             | 80                                | 78                          | 7.3   |
| <b>Servicios legales y financieros</b> | 16                                | 78                          | 8.6   |

(\*) También incluye los gastos de un período de 12 meses en las dos encuestas.

A pesar de estas limitaciones, los resultados muestran que para los artículos con menos variación entre trimestres, el incremento en la varianza muestral es moderado, por ejemplo, para Ingreso Total  $\alpha = 1.2$ , para pensiones y para Seguridad Social  $\alpha = 1.1$ . Para los artículos con una mayor variación sobre el tiempo, el uso de un período de referencia más corto lleva a un incremento sustancial en la varianza, por ejemplo, para Muebles y Odontología  $\alpha = 3$ , para Reparaciones y renovación de la vivienda  $\alpha = 4$ , y para Servicios legales y financieros  $\alpha = 9$ . Esta comparación de gastos en los artículos más detallados entre las dos encuestas puede llevar a equivocaciones, pero es razonable suponer que la reducción del período de referencia resultará en un incremento considerable en las varianzas muestrales para artículos específicos.

Para algunos grupos de gasto, se hizo una comparación de las varianzas muestrales para diferentes duraciones de los períodos de referencia, empleando sólo los datos de la EGF de 1986-87. Para las mismas familias de la muestra, se calculaban las varianzas de subperiodos de los períodos de referencia. Para algunos artículos, para los cuales se usaba un período de 12 meses, se calculaban las varianzas para cuatro períodos de referencia de 3, 6, 9 y 12 meses, y para otros artículos con un período de referencia de 3 meses, para 1, 2 y 3 meses. Algunos de los resultados se presentan en las Tablas 4.6 y 4.7.

Los resultados en la Tabla 4.6 muestran que si el período de referencia para estos artículos hubiera sido reducido a 3 meses, las varianzas muestrales hubieran incrementado por factores en el intervalo de 2.5-3.3. Se eligió un período de referencia de 12 meses para estos artículos porque se esperaba un gran incremento en la varianza, bajo la asunción de que habría efectos de memoria moderados (que resultó ser falso en un estudio posterior). Aunque, como muestran los resultados en la Tabla 4.5 (a pesar de sus limitaciones), se asocian incluso mayores incrementos de varianza con otros grupos de artículos. No obstante, se decidió usar un período de referencia más corto para ellos, debido a los graves efectos de memoria esperados.

**Tabla 4.6**  
**Varianza Muestral Relativa por Duración del Período de Referencia**  
**con Artículos de un Período de Referencia de 12 Meses**  
**EGF 1986-87 de Israel**

| Artículo de gasto   | Varianza muestral relativa al período de 12 meses |         |         |          |
|---------------------|---|---------|---------|----------|
|                     | 3 meses   | 6 meses | 9 meses | 12 meses |
| Total               | 2.74  | 1.52    | 1.12    | 1.00     |
| Ocio y Excusiones   | 2.64  | 1.58    | 1.18    | 1.00     |
| Hoteles             | 2.46  | 1.49    | 1.11    | 1.00     |
| Estancia extranjero | 3.34  | 1.79    | 1.17    | 1.00     |
| Viajes extranjero   | 3.34  | 1.67    | 1.11    | 1.00     |

Los resultados en la Tabla 4.7 muestran que para el gasto total en todos los artículos (en el estudio de los efectos de memoria), si se hubiera reducido el período de referencia de 3 meses a 2 meses, la varianza hubiera incrementado por 1.4, y a 1 mes, por 2.4. La duración del período de referencia, como se puede esperar, tiene diferentes efectos sobre las varianzas de los diferentes artículos, según la correlación intra-meses. Por ejemplo, con Servicio doméstico, la duración del período tiene casi ningún efecto sobre la varianza, mientras que para bienes duraderos y para Cultura y ocio el efecto es sustancial.

Tabla 4.7  
 Varianza Muestral Relativa por Duración del Período de Referencia  
 con Artículos de un Período de Referencia de 3 Meses  
 EGF 1986-87 de Israel

| Artículo de gasto                               | Varianza muestral relativa al período de 3 meses |         |         |
|---|--|---------|---------|
|   | 1 mes  | 2 meses | 3 meses |
| <b>Todos artículos incluidos en el estudio.</b> | 2.38   | 1.36    | 1.00    |
| <b>Artículos de gasto seleccionados</b>         |  |         |         |
| Joyería   | 2.47   | 1.34    | 1.00    |
| Relojes   | 1.74   | 1.00    | 1.00    |
| Edredones y mantas                              | 2.40   | 1.32    | 1.00    |
| Sabanas y toallas                               | 2.69   | 1.26    | 1.00    |
| Reparaciones vivienda                           | 1.75   | 1.08    | 1.00    |
| Servicio doméstico                              | 1.02   | 1.05    | 1.00    |
| Muebles   | 2.68   | 1.50    | 1.00    |
| Electrodomésticos                               | 2.86   | 1.40    | 1.00    |
| Duraderos de cultura y ocio                     | 3.25   | 1.43    | 1.00    |
| Cuadros y marcos                                | 1.51   | 1.15    | 1.00    |
| Odontología                                     | 2.34   | 1.18    | 1.00    |
| Médico particular                               | 0.99   | 0.64    | 1.00    |
| Óptico  | 2.16   | 1.26    | 1.00    |
| Licencia vehículo                               | 2.70   | 1.40    | 1.00    |
| Mant. y reparaciones vehículo                   | 2.31   | 1.45    | 1.00    |

#### 4.4. Duración Optima del Período de Referencia

Cuando se dispone de varias duraciones alternativas del período de memoria, una consideración importante al elegir entre ellas es una comparación del MSE ("mean square error"- error de la media cuadrada) (que consta de la varianza muestral y el sesgo cuadrado) de las claves estadísticas para un coste dado. La varianza muestral depende del número de familias en la muestra neta, la duración del período de referencia para cada familia en la muestra, el diseño de la muestra, y del procedimiento de ponderación. Sin embargo, se debería notar que otros componentes de la varianza pueden afectar las estimaciones de la encuesta, como la varianza de respuesta.

Se pueden introducir sesgos en las estimaciones de la encuesta por los efectos de memoria, y también por los efectos procesales y por varios otros factores. Aunque no sólo los efectos de memoria se afectarán por la duración del período de memoria, es razonable suponer que se pueden ignorar las otras fuentes de sesgo al elegir entre las varias duraciones alternativas de memoria.

usar para determinar el período de recuerdo óptimo para cada una de las estimaciones encuestales,  $k$ . Esto se hace al comparar el  $MSE_k(1)=V^2_k(1)+B^2_k(1)$  para las duraciones del recuerdo alternativas. Sin embargo, generalmente, ninguna única duración es adecuada para todas las estimaciones requeridas. Además, incluso para una sola estimación, la mejor duración del recuerdo para el total de la población no siempre es la mejor para varios subgrupos.

Una posibilidad es la de elegir una duración del recuerdo "media mejor" para todas las variables investigadas. Esto se puede hacer teniendo en cuenta la importancia de las diferentes variables y los diferentes tamaños de sus MSE. Otra, es determinar las diferentes duraciones del recuerdo para los diferentes grupos de variables.

En cuanto a los costes, los costes de entrevistar son más o menos iguales para los diferentes períodos de recuerdo. No es el caso cuando el período de recuerdo es corto y el período de referencia es largo, es decir, se realizan varias entrevistas para cada familia en vez de una.

En la sección 4.3 discutimos las varianzas muestrales para varios períodos de recuerdo, y se presentaron resultados seleccionados de la EGF de Israel bajo la asunción de ningún cambio ocurriera en el coste. Se puede estimar el sesgo debido a los efectos de memoria, sólo si se puede suponer que el mes "reciente" esté libre de efectos de memoria. Si es así, se estiman los sesgos de memoria de diferentes duraciones del recuerdo, comparando las estimaciones obtenidas, usando los datos de cada período de recuerdo, con las estimaciones derivadas del período reciente. Se deben interpretar estos resultados con precaución, dado que pueden no distinguir entre los sesgos debidos a los dos tipos de efectos de memoria, efectos telescopicos y de memoria. Así que, por ejemplo, el sesgo estimado de un período de recuerdo de 3 meses puede incluir efectos telescopicos en el mes "lejano", mientras que para un período de recuerdo de 1 ó 2 meses se evitan los efectos telescopicos porque sus datos están delimitados por los meses anteriores. Además, no se puede distinguir el efecto telescopico intra-mensual de otros efectos de memoria. También se debería recordar que las estimaciones de los sesgos de memoria son propensas a unas varianzas muestrales muy altas.

A pesar de todas estas limitaciones, vamos a examinar algunos resultados que se obtenían de los datos de la EGF 1986-87 de Israel. Algunos ejemplos se muestran en la Tabla 4.8 para gastos de un período de recuerdo de 3 meses y en la Tabla 4.9 de un período de 12 meses.

Para la estimación del Gasto Total de todos los artículos incluidos en el estudio, la varianza muestral es el componente dominante del MSE, y el período de 3 meses es el mejor: el MSE de un período de 2 meses es 1.4 veces más alto, y de un período de 1 mes, 2.2 veces más alto. La situación era similar con Muebles y con Bienes duraderos para Cultura y ocio y con Reparaciones de vehículos. Sin embargo, para muchos otros grupos de artículos, el sesgo de memoria es el componente dominante del MSE para el período de 3 meses: para Odontología  $MSE(1)/MSE(3) = 0.2$ , para Servicio doméstico el ratio es

Odontología  $MSE(1)/MSE(3) = 0.2$ , para Servicio doméstico el ratio es 0.4, para Relojes el ratio es 0.5, y para Honorarios de médico particular, 0.6.

**Tabla 4.8**  
**Errores Relativos del Período de 3 Meses y la Comparación**  
**del MSE por Duración del Período de Referencia**  
**EGF 1986-87, Israel**

| Grupos de gasto                             | Período de 3 meses              |                                |                            | MSE relativo<br>relativo a 3 meses |         |       |
|---|---------------------------------|--------------------------------|----------------------------|------------------------------------|---------|-------|
|   | Error<br>muestral<br>relativo % | Sesgo<br>memoria<br>relativo % | $\sqrt{MSE}$<br>relativo % |                                    |         |       |
|   |                                 |                                |                            | 3 meses                            | 2 meses | 1 mes |
| Todos los artículos incluidos en el estudio | 2.8                             | - 0.7                          | 2.9                        | 1.00                               | 1.37    | 2.25  |
| <b>Artículos seleccionados de gasto</b>     |                                 |                                |                            |                                    |         |       |
| Joyería                                     | 8.1                             | - 6.7                          | 10.5                       | 1.00                               | 1.02    | 1.46  |
| Relojes                                     | 9.3                             | -15.0                          | 17.7                       | 1.00                               | 0.67    | 0.48  |
| Edredones y mantas                          | 12.3                            | - 3.2                          | 12.7                       | 1.00                               | 1.25    | 2.25  |
| Sabanas y toallas                           | 8.5                             | - 7.4                          | 11.3                       | 1.00                               | 0.72    | 1.51  |
| Reparaciones vivienda                       | 10.7                            | -10.0                          | 14.6                       | 1.00                               | 0.58    | 0.94  |
| Servicio doméstico                          | 8.2                             | -10.6                          | 13.4                       | 1.00                               | 0.40    | 0.38  |
| Muebles                                     | 6.1                             | + 0.1                          | 6.1                        | 1.00                               | 1.77    | 2.69  |
| Electrodomésticos                           | 4.7                             | - 3.2                          | 5.7                        | 1.00                               | 0.96    | 1.96  |
| Duraderos de cultura y ocio                 | 7.4                             | - 1.3                          | 7.5                        | 1.00                               | 1.51    | 3.19  |
| Cuadros y marcos                            | 14.9                            | -11.3                          | 18.7                       | 1.00                               | 1.19    | 0.96  |
| Odontología                                 | 6.1                             | -18.6                          | 19.5                       | 1.00                               | 0.20    | 0.23  |
| Médico particular                           | 9.5                             | + 7.8                          | 12.3                       | 1.00                               | 1.66    | 0.59  |
| Óptico                                      | 6.6                             | - 7.3                          | 9.8                        | 1.00                               | -0.58   | 0.98  |
| Licencia vehículo                           | 5.2                             | -24.3                          | 24.9                       | 1.00                               | 0.31    | 0.11  |
| Mant. y reparaciones vehículo               | 5.7                             | - 0.2                          | 5.7                        | 1.00                               | 1.49    | 2.30  |

Para Ocio y Excusiones, se encontró que el peor período de memoria era de 12 meses, mientras que los de 3, 2 ó 1 trimestres proporcionan MSE parecidos, por ejemplo,  $MSE(9)/MSE(12) = 0.2$ . Para Viajes al extranjero el mejor período de memoria era de 3 meses con un ratio de 0.4. Estos resultados no son faciles de explicar, a no ser que las estimaciones de sesgo de memoria sean sesgados en sí, como se explicaba antes en la sección 4.2.

**Tabla 4.9**  
**Errores Relativos del Período de 12 Meses y la Comparación**  
**del MSE por Duración del Período de Referencia**  
**EGF 1986-87, Israel**

| Grupo de gasto       | Período 12 meses                |                                |                            | MSE relativo<br>relativo a 12 meses |         |         |         |
|----------------------|---------------------------------|--------------------------------|----------------------------|-------------------------------------|---------|---------|---------|
|                      | Error<br>muestral<br>relativo % | Sesgo<br>memoria<br>relativo % | $\sqrt{MSE}$<br>relativo % |                                     |         |         |         |
|                      |                                 |                                |                            | 12 meses                            | 9 meses | 6 meses | 3 meses |
| Total                | 3.70                            | -13.6                          | 14.1                       | 1.00                                | 0.18    | 0.25    | 0.21    |
| Ocio y excusiones    | 3.71                            | -16.0                          | 16.4                       | 1.00                                | 0.13    | 0.19    | 0.17    |
| Viajes al extranjero | 4.84                            | -10.2                          | 11.3                       | 1.00                                | 0.56    | 0.49    | 0.37    |

Estos resultados muestran que ninguna duración del período de memoria es la mejor para todas las estimaciones objetivas de la encuesta. Sin embargo, se requiere más investigación para poder proporcionar estimaciones más fiables de los sesgos de memoria.

#### 4.5 Estimación de la Variabilidad

Las EGF, como varias otras encuestas, tienen como objetivo suministrar estimaciones de la distribución y de los parámetros de la variabilidad de la población durante un período dado, por ejemplo, mediciones de la desigualdad de la distribución de las familias por ingresos. Otro ejemplo es la distribución de familias por el porcentaje de gasto en comida del gasto total, dando alguna indicación del nivel de vida.

Cuando se calcula cualquier medición de la variabilidad de la población basándose en los datos recogidos durante un período de referencia que no cubre el objetivo del período entero para cada unidad, incluye tanto la variabilidad entre las unidades de población como la variabilidad entre los períodos parciales de cada unidad dentro de un período dado. Por lo tanto, cuanto más corto sea el período de referencia y más alto la variabilidad entre los períodos parciales, más amplia será la variación entre las unidades de población que se estiman.

En muchas EGF fue preferible reducir el período de referencia (tanto de la agenda como de la entrevista) para obtener mejores estimaciones de las medias y ponderaciones para poner al día el IPC, a costa de obtener las estimaciones de distribución de la población y las correspondientes mediciones del gasto.

No es el mismo caso con los ingresos, porque es importante clasificar a las familias según sus ingresos y proporcionar estimaciones de gasto por el nivel de ingreso.

Un estudio preliminar, experimental, se llevó a cabo en Israel para estudiar los efectos de reducción del período de referencia de 12 meses a 3 meses sobre las estimaciones de ingreso. Para este estudio se usaban los datos administrativos de las nóminas mensuales de varios grupos de empleados.

Las estimaciones se obtenían por simulación para períodos de referencia de  $T$  meses,  $T=1, 2, 3, 6, 12$ . Como se esperaba, la variabilidad estimada entre los empleados se incrementó cuando la duración del período de referencia era más corta. Los ratios de la variabilidad estimada relativa al período de 12 meses eran: 1.02 para 6 meses, 1.05 para 3 meses, 1.12 para 2 meses y 1.20 para 1 mes.

Este estudio experimental también proporcionó resultados sobre la cantidad de clasificaciones erróneas de unidades según las decimas de ingreso anual, para las diferentes duraciones del período de referencia. Para cada unidad se determinó la verdadera decila anual por medio de la información de los 12 meses. Entonces, esto se repetía según la información de los períodos de referencia más cortos. Se encontró que 25% estaba mal clasificado, usando sólo 6 meses, 40% estaba mal clasificado, usando sólo 3 meses, 47% estaba

mal clasificado, usando sólo 2 meses, y 53% estaba mal clasificado, usando sólo 1 mes. La mayoría de las clasificaciones erróneas estaban a menos distancia que 1 decila, pero algunas unidades estaban mal clasificadas a la distancia de 1 ó 2 decilas: 12% con un período de referencia de 1 mes, 7% con un período de 2 meses, y sólo 1% con 3 meses.

Aunque este estudio no representó la población entera ni todos los componentes de ingreso, los resultados indican la importancia de este tema.

## CAPITULO 5. MUESTREO

En la mayoría de los países, en general el diseño de muestreo de las EGF no es muy diferente que el de otras encuestas de familias. Sin embargo, puede haber algunas consideraciones específicas en el diseño de la muestra para las EGF, por ejemplo, la importancia de la muestra en el tiempo, el grado de agrupación, o la eficacia de diseñar una muestra auto-ponderada o una muestra con diferentes probabilidades.

### 5.1. Muestreo Bidimensional

En las EGF, como en otras encuestas donde las variables a investigar varían sobre el tiempo, la población de la encuesta consta de dos dimensiones. Una dimensión es el conjunto de todas las familias,  $\{N\}$ , que pertenece a la población objetivo de la encuesta. La segunda dimensión es el conjunto  $\{T\}$  de todas las "unidades de tiempo" básicas, por ejemplo, días, semanas, etc. que cubren el período entero de la encuesta, que normalmente es de un año entero. La muestra para una EGF anual se selecciona de las dos dimensiones, a no ser que cada familia seleccionada tenga que proporcionar todos los datos requeridos para el año entero. Hoy en día no se usan los diarios de año entero en casi ningún sitio. Sin embargo, una entrevista que se relaciona a un período de referencia de un año es bastante frecuente. En algunos países se relaciona al mismo año para todas las familias en la muestra, por ejemplo, en Canadá y en Suecia, pero no siempre es el caso. En otros países, las entrevistas se refieren a períodos de referencia "móviles" anteriores al período del diario, por ejemplo, los últimos 12 meses o los últimos tres meses.

En casi todas las EGF la muestra tiene que estar diseñada con respecto a las dos dimensiones y el primer tema es cómo asignar la muestra entre estas dos dimensiones. Para un coste dado, lo que hay que considerar es si es mejor diseñar una muestra con un gran número de familias y un período corto de referencia para cada familia, o extender el período de referencia a costa de reducir el número de familias en la muestra. Estas consideraciones se refieren principalmente a la duración del período del diario y se han discutido en detalle en el Capítulo 3.

Para una duración de período del diario predeterminada, el período entero de la encuesta se divide en  $T$  subperíodos  $p$  ( $p=1,2,\dots,T$ ), para cada familia  $i$  ( $i=1,2,\dots,N$ ). La población de la encuesta, entonces, consta del conjunto bidimensional de  $\{N*T\}$  unidades, y se debería seleccionar la muestra de este conjunto, para que la Última Unidad Muestral (UUM) es cierta familia  $i$  en cierto período de investigación  $p$ .

Teóricamente, son posibles las siguientes técnicas de muestreo:

- (1) Un muestreo sencillo aleatorio de  $\{N*T\}$ , que es el método de muestreo más ineficaz para las dos dimensiones, respecto tanto al coste como a la varianza muestral.
- (2) Una muestra de períodos  $t$  entre  $T$ , donde cada período se considera un cluster de todas las familias  $N$  y entonces, dentro

de cada período muestral, se selecciona una muestra de familias independientemente, por cualquier técnica de muestreo que se considera mejor. Con este método, aumentará la varianza muestral cuanto más alta sea la varianza entre los períodos. A parte de eso, el trabajo de campo no se puede llevar a cabo consecutivamente durante el período entero de la encuesta, lo que puede llevar a dificultades en la organización.

- (3) Una muestra de  $n$  familias entre  $N$ , donde cada familia es un cluster de  $T$  subperíodos, del cual se hace una muestra de un subperiodo aleatorio. Esta técnica llevará a una muestra ineficaz en la dimensión del tiempo.
- (4) Una muestra estratificada de  $\{N*T\}$ , donde los subperíodos  $T$ , sirven como estrata y se saca una muestra de  $n_p$  independientemente de cada período  $p$  por cualquier técnica de muestreo que se considera mejor. Las muestras  $n_p$  generalmente son del mismo tamaño, pero una muestra no proporcional es opcional, como en los Estados Unidos, donde para el período anterior al Año Nuevo, se saca una muestra mayor. Esta técnica tendrá como resultado varianzas muestrales más bajas, pero puede causar dificultades operacionales en el trabajo de campo.
- (5) Una muestra de  $n$  familias entre  $N$  y luego, distribuir la muestra entre todos los subperíodos  $T$ , no necesariamente de forma igual. Esta técnica es similar en sus características a la técnica de (4), aunque la muestra no se saca independientemente para cada período. Sin embargo, la muestra seleccionada de esta forma es más fácil para el trabajo de campo.

A continuación, se discuten las dos fases de la última técnica, que normalmente es la preferida.

## 5.2. Marcos y Diseños de Muestreo

Al igual que en otras encuestas, el diseño de la muestra depende de las características del marco de la muestra, es decir, el tipo de unidades en el marco y la información disponible para cada unidad.

Se pueden emplear cuatro tipos principales de marco:

- (a) Un marco de área, donde cada unidad es un área geográfica identificada por límites definidos y todas las unidades cubren el área entera de la población objetivo de la encuesta. Se disponen de las características geográficas de cada área (por ejemplo, región, tipo de localidad, etc.) y generalmente alguna otra información, por ejemplo, la que se obtenía en el último censo (por ejemplo, el número de viviendas, o detalles socio-económicos del área entera).
- (b) Un marco de unidades de vivienda, donde se relaciona cada unidad de vivienda por separado. Cada unidad de vivienda tiene sus detalles de identificación, (dirección y nombre de propietario o inquilino). Se pueden obtener las características geográficas de cada unidad a partir de los detalles de identificación, así como otras características como el número de habitaciones, tipo de vivienda (piso, casa, etc.). Algunas veces, se puede incluir

también más información sobre las familias ocupantes de las viviendas en el momento de compilar el marco (por ejemplo, del último censo).

- (c) Un marco de familia, que tiene características parecidas al marco de viviendas, pero se identifica cada unidad por los detalles de la familia. Una unidad de vivienda puede tener más de una familia, y una familia se puede asociar con más de una vivienda. Se podrían disponer de más características de la familia en este tipo de marco que en el marco de la vivienda.
- (d) Un marco de personas, donde cada persona (todas las personas o sólo las de cierta edad) se relaciona por separado con detalles de identificación, y normalmente alguna información adicional básica (sexo, edad, etc.).

Con todos los tipos de marco, casi siempre se usa una muestra de dos o más etapas, permitiendo varios niveles de cluster geográfico. Cuando se usa sólo un marco de área, primero se hace una muestra de área y luego se tiene que confeccionar una lista de todas las UUM, es decir, las familias, por medio de una investigación especial de campo. Se puede seleccionar la muestra del área en una o varias etapas, según el tamaño de las áreas en cada etapa, y el nivel de cluster para cada muestra. Entonces las listas preparadas para cada área se pueden usar para marcos de muestra para la etapa final de seleccionar la muestra de las UUM, o incluso como las listas de las UUM de la muestra. Un marco sólo de área se usa cuando no se dispone de una alternativa mejor, a causa del mayor coste y los errores implicados en la preparación de listas de familias dentro de las áreas de la muestra.

No obstante, casi siempre, al usar uno de los otros tipos de marco, primero se selecciona un área de muestra, en dos o más etapas. Por ejemplo, en la primera etapa, se seleccionan las Unidades de Muestra Primarias (UMP), y en la segunda etapa, se selecciona una muestra de distritos enumerados del último censo dentro de cada una de las UMP. Entonces, para cada una de las áreas seleccionadas, se usa el marco para elegir la muestra de UUM, es decir, viviendas, familias o personas, según el tipo de marco de la muestra.

Cuando la UUM es la vivienda, se incluyen en la muestra final todas las familias que ocupan la unidad en el momento de la encuesta. Esto puede aumentar la cantidad de clusters, según la frecuencia con que se encuentra que más de una familia ocupa una vivienda. Con un marco de vivienda, la mayoría de las UUM coinciden con las unidades investigadas en la encuesta.

Con un marco de personas, dado que las unidades de la población son familias, cada familia se representa en el marco por cada uno de sus miembros. Por lo tanto, porque la inclusión de cierta familia en la muestra es por medio de la inclusión de uno de sus miembros, las probabilidades muestrales de las familias variarán según el número de sus representantes en el marco. Con este diseño de muestra, la muestra de familias se hace con una probabilidad proporcional a su tamaño (PPT), por ejemplo, la probabilidad muestral de una familia de cuatro personas será el doble de la de las familias con dos personas, y cuatro veces más de la de una familia de una persona. Si

se usa un marco que incluye solamente las personas adultas (18 años y más) estas probabilidades variarán menos que cuando se incluyen todas las personas.

Sin embargo, si hay que hacer la muestra de familias con la misma probabilidad que en un marco de personas, es preciso definir de antemano sólo un representante de cada familia en el marco, algo que casi siempre es difícil de realizar. Alternativamente, se puede definir al miembro representante de la familia con unas normas predeterminadas para aplicar en el campo y sólo se incluyen en la muestra aquellas familias cuyo representante cumple estas normas. Por supuesto, este diseño es muy costoso, y por lo tanto, no muy eficiente. Así que, si hay una elección, el uso de un marco de personas sólo se justifica si la muestra de familias con probabilidades proporcionales a su tamaño se demuestra ser más eficiente que una muestra con probabilidades iguales (ver la sección 5.4).

Un problema común a todos los marcos muestrales es que quizás no estén puestos al día en el período de la encuesta (aunque no es el caso, generalmente, con un marco de área). Cuanto más antiguos sean los datos del marco, más costosas y menos exactas serán las estimaciones de la encuesta. Si se usa un marco antiguo, podría haber una falta de cobertura que puede resultar en sesgos en las estimaciones de la encuesta.

El marco de unidades de vivienda puede estar sujeto a una falta de cobertura debido a las nuevas construcciones, sin embargo, normalmente se disponen de marcos especiales separados para hacer una muestra suplementaria. Con un marco de familias, puede haber una falta de cobertura debido a nuevas familias, por ejemplo, recién casados, divorciados, adultos solos que abandonan la casa de sus padres, etc., y normalmente no hay marcos adecuados para poder realizar muestras suplementarias para estos grupos.

Por otra parte, el uso de marcos antiguos puede resultar en casos inelegibles en la muestra: con un marco de viviendas, casas vacías, comercios, etc., y con un marco de familias y de personas, muertes, migración, etc.

Cuando se usa un marco de familias o de personas y las direcciones no están al día, incrementará el ratio de la no respuesta debido al no contacto, porque el entrevistador puede tener dificultades para encontrar la familia en su nueva dirección. Con un marco de viviendas no hay cambios de localización, pero si han cambiado los nombres de las calles pueden surgir dificultades para localizar las unidades de la muestra. Así que, si los ocupantes de una unidad de la muestra han cambiado, pueden surgir problemas de localización cuando se identifica la vivienda por su familia residente.

Tanto los no contactos como los casos inelegibles incrementan el coste de la encuesta. Pero, mientras que los casos inelegibles en la muestra sólo tienen un efecto menor en las varianzas muestrales de la mayoría de las estimaciones de la encuesta, la falta de cobertura y los no contactos pueden ser una fuente de sesgos graves. Por lo tanto, cuando es posible elegir, un marco de unidades de viviendas es normalmente preferible a un marco de familias, aunque el último

puede proporcionar datos más adecuados para una mejor estratificación.

### 5.3 Muestreo de Familias

Al igual que con todas las encuestas que se basan en visitar a las unidades investigadas, es decir, excluyendo las encuestas por correo o por teléfono, hay que considerar la extensión de los clusters geográficos de la muestra. Para un tamaño de muestra dado, hay que encontrar un equilibrio entre el efecto negativo del cluster sobre las varianzas muestrales (cuando existe una alta correlación intra-cluster positiva), y el efecto positivo de disminuir el coste de la encuesta. El diseño de la muestra tiene que tener en cuenta el tamaño de la asignación de los entrevistadores para un período, que depende de la organización de campo y del diseño de la encuesta, por ejemplo, la duración del período de la agenda, el número de visitas del entrevistador a la familia durante el período de registro, el tiempo medio requerido para la entrevista, etc.

El muestreo de familias de una etapa, sin cluster en la dimensión de la población, puede ser demasiado costoso, incluso cuando se introducen clusters en la dimensión de tiempo al agrupar la muestra en clusters geográficos para las asignaciones del entrevistador. Por lo tanto, a menudo se usa la muestra de dos o más etapas con clusters también en la dimensión de la familia. Sin embargo, en lo posible, se debería evitar los clusters, dado que a menudo se encuentra una alta correlación intra-cluster positiva en pequeñas áreas para los tipos de variables bajo investigación en las EGF.

Como ilustración de un diseño de muestra típico para las EGF, vamos a examinar, para simplificar, un muestreo de dos etapas con probabilidades muestrales finales uniformes.

Vamos a denotar por  $N_j$  el número de familias en UMP j ( $j=1,2,\dots,M$ ), donde  $N = \sum N_j$  es el número de familias en la población entera de la encuesta, y deja que  $n$  sea el tamaño de la muestra entera,  $f=n/N$  la probabilidad final de la muestra, y  $\tilde{n}$  el tamaño de asignación de cada entrevistador por período.

Primero, las UMP con  $N_j > \tilde{n}/f$  se incluyen con certeza en la muestra. Entonces, de cada una de estas UMP, se saca una muestra de las UUM independientemente con probabilidad  $f$ . Así que, con respecto a esa parte de la población,  $N(1)$ , incluida en este grupo de UMP,  $M(1)$ , de hecho, es una muestra estratificada de una etapa de tamaño  $n(1)=fN(1)$ , cada una de las grandes UMP siendo una estrata separada.

Luego se hace una muestra de dos etapas a partir de todas las demás UMP  $M(2)$  que consta de una población de tamaño  $N(2)=N-N(1)$ .

En la primera etapa, se seleccionan las UMP con probabilidad proporcional a su tamaño (PPT), es decir,  $f_{j1}=N_jm(2)/N(2)$ , donde  $m(2)=[n-n(1)]/\tilde{n}$  es el número de UMP en la muestra de este grupo.

En la segunda etapa, se selecciona una muestra de UUM dentro de cada una de las UMP de la muestra con la probabilidad  $f_{j2}=f/f_{j1}=fN(2)/m(2)N_j$ , para que la probabilidad final sea igual para todas las UUM.

La muestra de las UUM en cada UMP será al menos de tamaño  $n_j = N_j f > \bar{n}$  para UMP j seleccionada con certeza, y  $n_j = N_j f_{j2} = \bar{n}$  para cada una de las demás UMP. Esto será el caso si  $N_j$  son correctas, si no  $n_j$  variarán según el ratio del tamaño real de la UMP y su tamaño en el marco, y el tamaño real de la muestra total puede ser algo diferente al tamaño planificado.

Con esta técnica de muestreo, si el tamaño de la muestra dentro de las UMP que se incluían con certeza, es demasiado grande para la asignación de un entrevistador y demasiado pequeño para dividirse entre dos entrevistadores, puede resultar incómodo para la organización de campo. Para resolver este problema, se puede modificar el diseño de la muestra, agrupando estas UMP de tal forma que los tamaños de su población acumulados dará un tamaño de muestra de aproximadamente un múltiple de  $\bar{n}$ . Entonces, el número de clusters en la muestra en cada una de las UMP agrupadas se determina por un muestreo PPT y se ajusta la probabilidad muestral dentro de cada una de las UMP para que corresponda a la probabilidad final f. Por ejemplo, cuando una UMP es de tamaño  $N_1 = 1.4\bar{n}/f$  y una segunda UMP es de tamaño  $N_2 = 1.6\bar{n}/f$ , se juntan las dos UMP para que  $(N_1 + N_2)f = 3\bar{n}$ . Entonces, se selecciona una muestra PPT de tamaño 3, para que de una de estas dos UMP se seleccionarán dos clusters y sólo un cluster de la otra UMP.

Normalmente se aplica la estratificación en la muestra de las UMP, así como en las UUM, según la información disponible en el marco muestral.

A continuación se presentan unos ejemplos típicos de diseños de muestra para diferentes EGF:

- (a) En Israel se usa un diseño de muestra de dos etapas con probabilidades finales uniformes, pero la mayor parte de la muestra (aprox. 80% de la población) es una muestra estratificada de una etapa con una modificación de tamaño de cluster fijo. Se planifica cada una de las asignaciones de los entrevistadores para incluir unas 12 familias elegibles (de las cuales 2 ó 3 serán de no respuesta). Se incluye en la muestra con certeza cada localidad de tamaño que corresponde al menos a la asignación de un entrevistador. Las localidades pequeñas se seleccionan por PPT con una estratificación según el tipo, tamaño, región e índice socio-económico. Para cada localidad seleccionada, se ordena por direcciones el fichero municipal de impuestos que se usa en el marco muestral de unidades de vivienda. Aunque estos ficheros están continuamente puestos al día, se realiza una muestra suplementaria para cubrir nuevas construcciones desde el momento de la selección de la muestra primaria hasta el final del período de la encuesta. Se seleccionan muestras adicionales de marcos especiales de poblaciones semi-institucional, por ejemplo, residencias de estudiantes, nuevos centros de inmigrantes, etc., que no están cubiertos por el marco principal. El tamaño de la muestra es aprox. 7.500 unidades de vivienda.
- (b) En el Reino Unido (Elliot, 1991) se usa un diseño de muestra de dos etapas con probabilidades finales uniformes: en la primera etapa, un muestreo de áreas (cada área contiene aprox. 2.500

familias) y en la segunda etapa, un muestreo de direcciones dentro de las áreas de la muestra. La muestra de las UMP se estratifica por región, nivel de urbanización e indicadores del nivel socio-económico con información del censo más reciente de la población. Las direcciones se seleccionan del Fichero de Códigos Postales, es decir, una lista de puntos de entregas postales por todo el país. Cada una de las muestras anuales incluye aprox. 10.000 familias.

- (c) En Hungría (Elteto, 1991) la muestra de las EGF es una submuestra de una muestra principal que consta de distritos de enumeración censal. La muestra está diseñada en tres etapas. En el primer etapa hay un muestreo de localidades: con certeza se incluyen en la muestra ciudades con más de 25.000 habitantes, y las demás localidades se seleccionan con una probabilidad proporcional al número de viviendas, después de una estratificación por tamaño y estructura ocupacional. La población de pueblos pequeños tiene una mayor probabilidad de ser seleccionada que la de las ciudades. En la segunda etapa, se seleccionan los distritos de enumeración por PPT según su tamaño, después de una estratificación por estructura ocupacional. En la tercera etapa se seleccionan familias dentro de cada uno de los distritos de la muestra a partir de un marco de muestra estratificado por tamaño y grupo social. El marco de la muestra está puesto al día, corrigiendo la lista de viviendas a derribar y de las nuevas construcciones, en los distritos de la muestra antes de seleccionar la muestra. La muestra de la EGF cubre más de 1.300 distritos censales en casi 400 localidades. En cada distrito de enumeración se seleccionan 9 familias, así que el tamaño total de la muestra es de casi 12.000 familias.
- (d) En Italia, (Russo y Coccia, 1991) se usa un muestreo de dos etapas con familias para las UUM en vez de las viviendas. En la primera etapa, se usan las localidades para las UMP: se seleccionan con certeza todas las localidades de más de 50.000 habitantes, y las localidades más pequeñas se seleccionan con una probabilidad proporcional a su tamaño, después de estratificación por tamaño, zona y actividad económica predominante. En la segunda etapa, se seleccionan a las familias sistemáticamente a partir de listas de familias registradas dentro de las UMP. El uso de un marco de viviendas lleva a un alto ratio de no contacto y por lo tanto, se usa un proceso de sustitución. Cada mes se incluyen en la muestra 3.200 familias, 38.500 en un año.

#### 5.4 Muestreo No Proporcional de Familias

En general, cuando el objetivo primario de una encuesta es obtener una estimación para la población entera, o varias estimaciones altamente correlacionadas, una muestra no proporcional puede ser más eficiente que una proporcional. En ese caso, para un tamaño fijo de muestra,  $n$ , se pueden calcular las probabilidades óptimas de selección  $f_h(\text{opt})$ , dentro de la estrata  $h$ , según la asignación de Neyman:

$$f_h(\text{opt}) = \frac{n_h(\text{opt})}{N_h} = n \frac{s_h/\sqrt{c_h}}{\sum N_h s_h/\sqrt{c_h}}$$

donde, para cada estrata  $h$ ,  $N_h$ ,  $S_h$ , y  $c_h$  son el tamaño de la población, la desviación estándar y el coste por unidad, respectivamente y  $n_h(\text{opt})$  son los tamaños óptimos de la muestra en la estrata  $h$ . Esto significa que serán más altas las probabilidades de selección óptima, cuanto más alta sea la variabilidad dentro de la estrata y más barata sea la investigación por unidad de esa estrata. Si el coste por unidad no difiere mucho entre las estratas, y las  $s_h$  son de parecida magnitud, una muestra proporcional es preferible.

Las EGF son de multi-usos y normalmente no hay un solo conjunto óptimo de  $f_h$  para todas las estimaciones diferentes que hay que obtener de los datos de la encuesta. Por lo tanto, la mayoría de las muestras para las EGF se seleccionan con probabilidades iguales. Sin embargo, pueden haber otras consideraciones para diseñar una muestra no proporcional para las EGF.

Algunas veces, existe una especial necesidad a suministrar estimaciones separadas para ciertas sub-poblaciones que son bastante pequeñas y entonces, las muestras para estos grupos se seleccionan con probabilidades más altas para asegurar muestras de un tamaño suficiente. Por ejemplo, en varios países existen muestreos de las áreas rurales con probabilidades más altas porque se requieren estimaciones especiales para esta sub-población. Otra situación es cuando para ciertos grupos de la población se esperan ratios mucho más altos de la no respuesta, por ejemplo, en familias de tamaño 1. Por lo tanto, para obtener una muestra proporcional de los encuestados, se puede considerar una probabilidad de selección más alta para los grupos con ratios altos de no respuesta. En la práctica, sólo se puede hacer esto si el marco de la muestra incluye la información necesaria para esa estratificación.

Otra forma de muestreo no proporcional (que se usa en Suecia) es seleccionar la muestra entera de familias con una PPT según el tamaño de la familia. Esto es posible cuando se dispone de un marco muestral de familias con la información sobre el tamaño de la familia puesta al día, o cuando se usa un marco de personas. Con este último, se selecciona una muestra de personas individuales, y se incluyen en la muestra las familias a que pertenecen estas personas (como se explicaba antes en la sección 5.2)

Vamos a examinar la eficacia de este método de muestreo.

Deja que  $Z_i$  sea el tamaño de la familia  $i$ , que se puede definir por todos sus miembros, o sólo por sus miembros adultos, digamos de 18 años y más. Entonces, se selecciona la familia  $i$  con probabilidad  $\pi_i = Z_i/Z$ , donde  $Z$  es el número total de individuos en la población.

Deja que  $X_i$  sea el valor de la variable investigada (por ejemplo, gasto total) para la familia  $i$  de la muestra ( $i=1,2,\dots,n$ ) y deja que  $\bar{X} = \sum X_i/N$  sea el gasto medio por familia en la población.

Deja que  $A_i = X_i/Z_i$  sea la media por persona en la familia  $i$ , y  $\bar{A} = \bar{X}/Z$  la media por persona en la población total, donde  $Z = Z/N$  es el tamaño medio de familia.

La estimación sencilla de la media por familia es  $\bar{x}_{ep} = \frac{\sum x_i}{n}$  con probabilidades iguales, y  $\bar{x}_{pps} = \frac{(1/n)(1/N)\sum x_i/\pi_i}{\pi_i} = \frac{\bar{z}(1/n)\sum A_i}{\pi_i} = \bar{z}\bar{A}_{ep}$  con PPT, con tal que se sepa N.

La varianza muestral relativa (asumiendo para simplificar, un muestreo con sustitución) es

$$R.Var(\bar{x}_{ep}) = \frac{\frac{\sum (x_i - \bar{x})^2}{n}}{\frac{\sum x_i^2}{nN^2}} = \frac{C_x^2}{nN^2} \quad \text{con probabilidades iguales, y}$$

$$R.Var(\bar{x}_{pps}) = \frac{\frac{\sum \pi_i (x_i/\pi_i - \bar{x})^2}{nN^2}}{\frac{\sum \pi_i^2}{nN^2}} = \frac{\frac{\sum \pi_i (A_i/\bar{A} - \bar{A})^2}{nN^2}}{\frac{\sum \pi_i^2}{nN^2}} = \frac{C_A^2}{n} \quad \text{con PPT.}$$

Por lo tanto, el muestreo de PPT será mejor si  $C_A^2 < C_x^2$ , es decir, la variación relativa entre el gasto de las familias es mayor que entre el gasto por persona. Esto puede ser válido para algunas de las variables bajo investigación en las EGF, pero no para todas ellas, ni siquiera para las más importantes. Se puede esperar que  $C_A^2 < C_x^2$  para gasto en comida, donde la cantidad de gasto depende del número de personas a alimentar. Pero, no sería el mismo caso con gasto en salud, que será mayor en familias de una o dos personas, muchas de ellas personas mayores. El gasto familiar en Ocio puede disminuir con el tamaño de la familia, así como con Viajes al extranjero. La relación entre  $C_A^2$  y  $C_x^2$  depende, por supuesto, de la definición del tamaño de familia que se elige para determinar las posibilidades de selección - todos los miembros de la familia, o sólo los adultos.

Se examinaba en Israel la eficacia de una muestra PPT de familias, estudiando los datos de la EGF 1986-87. Las varianzas muestrales de las estimaciones de la encuesta para la población total se estimaban, primero con el diseño de muestra actual, es decir, con probabilidades de selección iguales. En segundo lugar, se calculaban las varianzas muestrales para las estimaciones de la encuesta para cada familia de tamaño h ( $h=1,2,\dots,6+$ ). Se utilizaban estas varianzas para estimar la varianza muestral para las estimaciones de la población entera si se hubiera usado una muestra estratificada por tamaño de familia con probabilidades de selección que corresponden a una muestra PPT de familias. Esto podría ser una buena aproximación, dado que se usaba una post-estratificación por tamaño de familia para la ponderación y se tenían en cuenta la mayoría de los efectos del proceso de ponderación y de los ratios de la no respuesta (que se discute en los siguientes capítulos) en el cálculo de las varianzas muestrales.

Vamos a denotar por  $\sigma_h^{(1)}$  y  $\sigma_h^{(2)}$  las varianzas muestrales de la estrata h, es decir, familias de tamaño h, con una muestra estratificada proporcional y con una muestra estratificada no proporcional, respectivamente. Se debería notar que para las familias de tamaño 7 y más, las probabilidades de selección se elegían para ser iguales que para una familia de tamaño 6, para evitar demasiada diferencialidad. Se puede implementar en la práctica, un diseño de muestra de este tipo, al incluir sólo 6/7 de las familias de tamaño 7, 6/8 de tamaño 8, etc.

Deja que  $n_h(1)$  y  $n_h(2)$  sean el tamaño de la muestra neta para una muestra proporcional y no proporcional, respectivamente, para un tamaño dado de la muestra general neta n, es decir,  $n=\sum n_h(1)=\sum n_h(2)$ .

Deja que  $N_h$  sea el numero de familias en la estrata h en la población, y  $W_h N_h / N$ .

Se puede demostrar que  $n_h(2)=n n_h(1)h/\sum n_h(1)h$  donde h=6 para 6+.

La Tabla 5.1 presenta la distribución de la muestra neta en la EGF de Israel de 1986-87, por tamaño de familia, a partir del muestreo proporcional de hecho empleado y la distribución que se hubiera obtenido si se hubiera usado un muestreo PPT. También se presentan en la Tabla 5.1 los ratios  $\beta_h$ , es decir, el ratio entre la varianza muestral de la muestra proporcional y de la muestra PPT para los subgrupos por tamaño de familia.

**Tabla 5.1**  
**Ponderaciones de Población y Tamaño Neto de la Muestra**  
**por Tamaño de Familia y Diseño de Muestra**  
**EGF de Israel 1986-87**

| Ponderación de población<br>y Tamaño neto de muestra                    | Tamaño de familia h |      |       |      |       |       |       |
|---|---------------------|------|-------|------|-------|-------|-------|
|   | Total               | 1    | 2     | 3    | 4     | 5     | 6+    |
| Ponderación   | 100.0               | 17.5 | 22.3  | 13.6 | 17.2  | 14.4  | 15.0  |
| $n_h(1)$ : Proporcional   | 5,000               | 659  | 1,081 | 698  | 963   | 764   | 835   |
| $n_h(2)$ : PPT  | 5,000               | 188  | 614   | 591  | 1,096 | 1,086 | 1,425 |
| $\beta_h = \frac{n_h(2)}{n_h(1)} = \frac{\sigma^2_h(1)}{\sigma^2_h(2)}$ | 1.00                | 0.28 | 0.57  | 0.85 | 1.14  | 1.42  | 1.71  |

Para las estimaciones de los subgrupos de la población por tamaño de familia (uno de los objetivos principales de las EGF), está claro que las varianzas muestrales de las familias de tamaño 1, 2 ó 3 incrementarán sustancialmente con un diseño PPT ( $\beta_1=0.28$ ,  $\beta_2=0.57$  and  $\beta_3=0.85$ ) mientras que para familias más grandes de tamaño 4, 5, y 6+, la varianza se disminuirá ( $\beta_4=1.14$ ,  $\beta_5=1.42$  and  $\beta_{6+}=1.71$ ).

Para la población total, el ratio de las varianzas muestrales serán

$$\beta = \frac{\sigma^2(1)}{\sigma^2(2)} = \frac{\sum W_h^2 \sigma^2_h(1)}{\sum [W_h^2 \sigma^2_h(1)/\beta_h]} \quad \text{y el diseño PPT sería más eficaz cuando}$$

La eficacia del muestreo PPT por tamaño de familia (valores de  $\beta$ ) y para el muestreo PPT por el número de miembros de la familia mayores de 18 años (valores de  $\beta^*$ ), se calculaban para grupos de gasto e ingreso. Algunos de estos resultados se presentan en la Tabla 5.2.

Los resultados demuestran claramente que para la mayoría de las estimaciones, el muestreo con probabilidades iguales es más eficiente que ambos tipos de muestreo PPT. Además, para casi todas

las estimaciones, para la población entera, PPT por el tamaño de la familia es menos eficiente que PPT por el número de adultos.

Para Gasto Total  $\beta=0.73$  y  $\beta^*=0.96$ , para Ingreso Total  $\beta=0.81$  y  $\beta^*=0.93$ . Incluso con Comida  $\beta=0.84$  y  $\beta^*=0.81$ , que se puede explicar por el consumo de comida más barata en las familias grandes, y comida más cara en las familias más pequeñas. Tal y como se esperaba,  $\beta$  es comparativamente pequeño para Salud ( $\beta=0.64$ ), mientras que  $\beta^*$  es mucho más alto ( $\beta^*=0.98$ ) lo que se explica en parte por el sistema de pagos a la seguridad social en Israel.

Se presentan aquí algunos subgrupos de gasto. Para Pan,  $\beta$  es cerca a 1 y también para carne y Aves, para Productos Lácteos y Huevos y para Agua. Con la mayoría de los otros subgrupos,  $\beta<1$  y también  $\beta^*<1$ . Por ejemplo, para Odontología  $\beta=0.66$ , para Viajes al extranjero  $\beta=0.66$ , para Ocio y excursiones  $\beta=0.65$  y para Comidas fuera de casa  $\beta=0.56$ . Se hace una excepción con Ropa de niño, para la cual  $\beta=1.25$ .

También, para las estimaciones de proporciones, por ejemplo, la posesión de bienes duraderos, el muestreo PPT lleva, obviamente, a varianzas muestrales más altas. Así que, para la mayoría de las estimaciones de la población total, así como los subgrupos, se encontró, sin ninguna duda, que el muestreo PPT está lejos de ser óptimo.

Tabla 5.2  
 Ratio de las Varianzas Muestrales del Muestreo Proporcional y de PPT  
 Población Total  
 EGF de Israel, 1986-87

| Ingreso/Grupo de gasto              | Ratio de varianzas muestrales de proporcional<br>y de PPT por nº de miembros de la familia |                    |
|-------------------------------------|--|--------------------|
|                                     | Todos miembros (B)   | Adultos (18+) (B*) |
| <b>INGRESO TOTAL</b>                | 0.81   | 0.93               |
| Ingreso neto                        | 0.74   | 0.96               |
| Ingreso de trabajo                  | 0.80   | 0.89               |
| Ingreso de capital                  | 0.61   | 0.97               |
| Ingreso de pensiones                | 0.51   | 0.72               |
| Ingreso de asistencia social        | 0.71   | 0.78               |
| <b>GASTO TOTAL</b>                  | 0.73   | 0.96               |
| <b>GRUPOS PRINCIPALES DE GASTO:</b> |  |                    |
| Comida (excl. verduras y fruta)     | 0.84   | 0.81               |
| Verduras y fruta                    | 0.83   | 0.93               |
| Vivienda                            | 0.67   | 0.71               |
| Mantenimiento vivienda              | 0.71   | 0.88               |
| Muebles y equipo hogar              | 0.77   | 0.88               |
| Ropa y calzado                      | 0.87   | 1.00               |
| Salud                               | 0.64   | 0.98               |
| Educación, cultura y ocio           | 0.75   | 0.89               |
| Transporte y comunicación           | 0.73   | 0.84               |
| <b>SUBGRUPOS DE GASTO:</b>          |  |                    |
| Productos de pan, cereales y pasta  | 1.01   | 0.93               |
| Carne y aves                        | 0.91   | 1.06               |
| Productos lácteos y huevos          | 0.93   | 1.02               |
| Comidas fuera de casa               | 0.56   | 0.71               |
| Agua                                | 0.97   | 0.88               |
| Ropa de niño                        | 1.25   | 0.99               |
| Calzado                             | 0.84   | 0.93               |
| Seguros médicos                     | 0.65   | 0.91               |
| Odontología                         | 0.66   | 0.93               |
| Ocio, cultura y deporte             | 0.76   | 0.87               |
| Ocio y excursiones                  | 0.65   | 0.81               |
| Duraderos de cultura y ocio         | 0.59   | 1.01               |
| Transporte público                  | 0.80   | 0.88               |
| Viajes al extranjero                | 0.66   | 0.71               |
| Correos y teléfono                  | 0.68   | 0.70               |

## 5.5 Asignación de la Muestra a un Período de Tiempo

En primer lugar, vamos a examinar el modelo de la asignación de la muestra entre los períodos de la encuesta. Lo que distingue a los varios modelos de asignación es la cantidad de tiempo solapado entre los períodos, que puede tener diferentes efectos sobre las estimaciones de la encuesta debido a un posible sesgo en la recogida de datos del diario (ver la sección 3.2).

Para un diario de dos semanas, son posibles los siguientes tres modelos:

- 1) El modelo no solapado - La muestra se divide en 26 submuestras, cada una de ellas se asigna un período diferente de dos semanas. Todas las familias de la misma submuestra empiezan sus diarios el mismo día.
- 2) El modelo solapado en una semana - La muestra se divide en 52 submuestras y cada semana, en el mismo día, comienza otra de las submuestras. Así que, cada semana (excepto la primera y la última), hay dos submuestras en la encuesta, una en su primera semana de recogida del diario y la otra en su segunda semana de registro.
- 3) El modelo solapado a diario - La muestra se divide en 365 submuestras y cada día una nueva submuestra comienza a registrar. Así que, todos los días, excepto el principio y el final del período de la encuesta, participan 14 submuestras en la encuesta, cada una con un día de recogida de serie diferente.

Por supuesto, pueden haber otros modelos de asignación con diferentes períodos de solapado. No se puede implantar ninguno de estos modelos en la práctica, sin alguna deformación, sobre todo en el modelo diario.

Vamos a examinar los efectos de estos modelos en las estimaciones de la encuesta bajo la asunción de que cada submuestra sea una muestra aleatoria y que no exista la no respuesta o si existe, es aleatoria.

Vamos a denotar por  $d$  ( $d=1,2,\dots,7$ ) y por  $d+7$  ( $d+7=8,9,\dots,14$ ) el día de serie de recogida en la primera semana y la segunda semana del diario, respectivamente. Así que, si por ejemplo, el día inicial del diario es martes,  $d=1,8$  se refiere a todos los martes,  $d=6,13$  se refiere a todos los domingos, etc.

Para cualquier período de dos semanas, deja que  $\bar{X}_d$  y  $\bar{X}_{d+7}$  denotan el verdadero gasto por familia el día  $d$  y día  $d+7$  de la recogida del diario.

Deja que  $\bar{X}^*_d$  y  $\bar{X}^*_{d+7}$  denotan los valores respectivos registrados. Entonces los gastos verdaderos y el gasto medio recogido para todo el período de la encuesta son:

$$\bar{x} = \sum_d (\bar{X}_d + \bar{X}_{d+7}) \text{ and } \bar{x}^* = \sum_d (\bar{X}^*_d + \bar{X}^*_{d+7})$$

Entonces los sesgos de registro relativos de  $\bar{X}^*_{d+7}$ ,  $\bar{X}^*_{d+7}$  y  $\bar{X}^*$ , son:

$$\beta_d = (\bar{X}^*_{d+7}/\bar{X}_d) - 1, \quad \beta_{d+7} = (\bar{X}^*_{d+7}/\bar{X}_{d+7}) - 1 \text{ and } \beta = (\bar{X}^*/\bar{X}) - 1.$$

Se puede demostrar que el sesgo relativo  $\beta$  no es igual para los tres modelos de asignación:

Con el modelo no solapado  $\beta(1) = \sum w_d (U_d \beta_d + U_{d+7} \beta_{d+7}) = \sum w_d \beta_w^*$ ,

con el modelo solapado en una semana  $\beta(2) = \sum w_d \frac{\beta_d + \beta_{d+7}}{2} = \sum w_d \beta_w^*$

y con el modelo solapado a diario  $\beta(3) = \frac{1}{7} \sum \frac{\beta_d + \beta_{d+7}}{2} = \frac{1}{7} \sum \beta_d$

donde  $w_d = (\bar{X}_d + \bar{X}_{d+7})/\bar{X}$  son las ponderaciones de los días de semana, y

$$U_d = \bar{X}_d / (\bar{X}_d + \bar{X}_{d+7}) \text{ and } U_{d+7} = \bar{X}_{d+7} / (\bar{X}_d + \bar{X}_{d+7})$$

son las ponderaciones de la primera y segunda semana para cada uno de los días de semana.

No habría ninguna diferencia entre los sesgos de los tres modelos si para todos los días  $w_d = 1/7$  y también  $U_d = U_{d+7} = 1/2$ . Sin embargo, como se demostraba en la sección 3.2, no se espera que el gasto semanal se distribuya igualmente entre los diferentes días de la semana. Así que, tanto en el modelo no solapado como en el modelo solapado en una semana, normalmente los sesgos serán diferentes que en el modelo solapado a diario. Los sesgos de los modelos no solapados y del modelo semanal no serán iguales si  $U_d \neq U_{d+7}$ , por ejemplo, si el comportamiento de gasto de las familias es afectado por el momento del pago del salario donde se basan en un ciclo de dos semanas. Excepto por el modelo a diario, el sesgo depende de la interacción entre los sesgos del registro diario y las ponderaciones diarias del gasto.

Un estudio empírico podría indicar qué es el modelo con menor sesgo. Sin embargo, una evaluación de  $\beta_d$  no es tan sencillo, aunque existe evidencia que el sesgo en los registros de las agendas varía según el día del registro. Si cuanto más altos sean los  $\beta_d$ , más altos serán los  $w_d$ . sería preferible el modelo solapado a diario. El modelo semanal es mejor que el modelo no solapado si  $\beta_d$  son más altos en esa semana con  $w_d$  más altos. Aunque no hay evidencia de tales interacciones, probablemente sí existen y por lo tanto el modelo solapado a diario parece ser el mejor. Sin embargo, este modelo es difícil, si no imposible a aplicar en la práctica.

Una aproximación del modelo a diario se puede lograr con el modelo semanal, si no hay un día inicial fijo para todas las familias, y se instruyen a los entrevistadores a distribuir la entrega de las agendas, tan igualmente como sea posible, entre los días de la semana. Este es el modelo que se usa en la EGF de Israel, 1992-93. Se debería notar que en la EGF anterior se usaba un modelo no solapado, y encontramos que los entrevistadores entregaban la mayoría de las agendas durante la primera semana de su asignación de dos semanas, por lo tanto cada segunda semana había muy pocos días iniciales. El uso del modelo solapado a diario, incluso si no se implementa totalmente en la práctica, probablemente proporcionará

datos para un estudio que podría llevar a una mejor comprensión de los sesgos del registro.

Para la asignación de la muestra sobre el período de tiempo, primero se determina el número de períodos por medio del modelo de asignación. Para cualquier número de períodos T, los clusters finales de la muestra se pueden asignar a submuestras T aleatoriamente, o por alguna estratificación, y cada submuestra se designa aleatoriamente a uno de los períodos T. Esta técnica puede ser difícil de aplicar porque no tiene en cuenta la organización del trabajo de campo durante el año. Por lo tanto, se requiere una forma más sofisticada de distribuir la muestra entre los períodos, con las limitaciones impuestas por el trabajo de campo. Aunque esto puede llevar a una desviación de una asignación puramente aleatoria, es preferible porque, de todas formas, es probable que ocurran muchas desviaciones de la asignación deseada cuando no se tengan en cuenta las limitaciones de campo.

Anteriormente a la asignación de la muestra entre los períodos de la encuesta, todas las familias de la muestra se agrupan desde el principio en los clusters finales, cada uno consta de la asignación de un entrevistador durante un período, o tienen que agruparse dentro de aquellos clusters que contienen más de una asignación. Algunas veces, para disminuir la cantidad de clusters de período dentro de los UMP muy grandes, primero se divide la muestra sistemáticamente en dos o más submuestras y se realiza la agrupación en asignaciones dentro de cada submuestra por separado.

Por ejemplo, vamos a examinar la forma en que se hizo la asignación en la EGF de Isreal, 1992-93, donde se emplea un diario de dos semanas. En cada asignación de un período, el entrevistador trabaja durante cuatro semanas, es decir, entregando los diarios durante las primeras dos semanas, luego visitando a las familias durante el tiempo que mantienen el registro, y llevando a cabo las entrevistas finales. Así que, cada entrevistador puede tener hasta un total de 13 asignaciones, aunque pueden ser menos, según cuándo comienza y finaliza su tarea.

## CAPITULO 6. NO RESPUESTA

### 6.1 Efectos sobre las Estimaciones de la Encuesta

En cada encuesta, se puede dividir la muestra en cinco grupos:

- (1) Encuestados - Unidades que pertenecen a la población de la encuesta y para las cuales toda la información requerida para todos los artículos bajo investigación que se obtenía era adecuada para ser incluida en el proceso de la encuesta, es decir, no fue rechazada durante el proceso de validación.
- (2) Encuestados con no respuesta de artículos - Unidades que pertenecen a la población de la encuesta y para las cuales sólo faltan los datos requeridos para un número limitado de artículos. Para estos casos los datos que faltan se imputan o se dejan como "desconocido", y están incluidos en las estimaciones de la encuesta.
- (3) Encuestados de no respuesta - Unidades que pertenecen a la población de la encuesta, pero para las cuales no se obtenía ninguna información (a causa de negación, abandono durante el periodo de investigación, enfermedad, ausencia, etc.) o para las cuales los datos recogidos no fueron suficientes para emplear, o eran de mala calidad y eran rechazados durante el proceso de validación.
- (4) Casos inelegibles - Unidades fuera del ámbito de la población de la encuesta, de una muestra de grupo que no se podría excluir del marco de la muestra, por ejemplo, unidades de casas vacías. Las unidades inelegibles de la muestra tienen que omitirse y no se deberían considerar como no respuesta.
- (5) No contacto - Las unidades para las cuales no se obtenía información en absoluto a causa de la imposibilidad de localizarlas por parte del entrevistador, o incluso que ni intentó localizarlas (por ejemplo, no tenía acceso fácil, entrevistador sobrecargado de asignaciones, etc.). Algunos de los no contactos podrían ser unidades inelegibles y por lo tanto, este grupo en su conjunto no se puede considerar como no respuesta. Normalmente se consideran a las unidades ausentes como no respuesta, aunque también pueden considerarse no contactos.

La distribución de la muestra en estos cinco grupos depende del marco de la muestra, de los procedimientos de la investigación, de las calificaciones de la plantilla de campo, así como del tipo de datos y de la población bajo investigación.

La no respuesta reduce el tamaño neto de la muestra y por lo tanto incrementa la varianza de la muestra. Pero, a menudo el tamaño inicial de la muestra,  $n^*$ , se planifica de antemano para tener en cuenta la respuesta esperada,  $\phi$ , para que  $n^*=n/\phi$ , donde  $n$  es el tamaño objetivo de la muestra neta. Sin embargo, esto incrementa el coste de campo porque hay que contactar con los encuestados de la no respuesta,  $n^*-n$ , y se hace todo tipo de esfuerzo para que participen en la encuesta. Además los ratios de la no respuesta normalmente no

se distribuyen igualmente entre las asignaciones del entrevistador y la eficiencia del trabajo de campo disminuye y por lo tanto, es incluso más costoso. El abordamiento alternativo de sustituir las no respuestas para obtener un tamaño de muestra fijo es más complicado y de este modo probablemente más caro.

Sin embargo, el problema más grave de la no respuesta está en los posibles sesgos que pueden introducirse en las estimaciones de la encuesta, y la sustitución de las no respuestas no resuelve este problema.

Si  $\mu_1$  y  $\mu_2$  son las medias de la variable investigada de los encuestados y de las no respuestas, respectivamente, y  $\phi$  y  $1-\phi$  son las proporciones de estos grupos, la media de la población entera es  $\mu = \phi\mu_1 + (1-\phi)\mu_2$ .

Por lo tanto, el sesgo de la estimación de la encuesta  $\hat{\mu}$ , que se basa sólo en las respuestas será

$$\text{Sesgo } (\hat{\mu}) = E(\hat{\mu}) - \mu = \mu_1 - [\phi\mu_1 + (1-\phi)\mu_2] = (1-\phi)(\mu_1 - \mu_2)$$

La cantidad de sesgo es un producto de la proporción del ratio de la no respuesta y la diferencia entre las medias de los encuestados y de las no respuestas. La encuesta proporciona estimaciones de  $\mu_1$  y de  $\phi$ , pero no proporciona información sobre  $\mu_2$ .

Muchos estudios en varias encuestas han demostrado que las de la no respuesta a menudo difieren de los encuestados, es decir,  $\mu_2$  difiere de  $\mu_1$ . Puede ser que  $\mu_2$  es parecido a  $\mu_1$  para la población entera, pero puede no ser el caso para los diferentes subgrupos de la población, por ejemplo, para un subgrupo  $\mu_2 > \mu$  y  $\mu_2 < \mu$  para el otro grupo, y así que, los sesgos para la población entera podrían anular los unos a los otros.

Se debería mencionar que la presencia de unidades inelegibles en la muestra también lleva a un incremento en el coste de la encuesta (porque, normalmente, el hecho de que no pertenecen a la población de la encuesta se descubre solamente después de haber hecho los contactos en el campo) y puede incrementar la varianza en las estimaciones de la encuesta. Sin embargo, las unidades inelegibles no causan sesgos en las estimaciones de la encuesta, con tal que se evite sustituirlas (para mantener un tamaño fijo de muestra).

## 6.2 Ratios de la No Respuesta

A causa de la carga pesada que impone las EGF a los encuestados, y a causa del tipo de variables a investigar, los ratios de la no respuesta en estas encuestas, casi en todas partes, son mucho más altos que en otras encuestas familiares. En las EGF, la no respuesta debido a negación y a abandono forma la mayoría de la no respuesta total, mientras que en otras encuestas familiares juegan un papel mucho menos importante. En las EGF algunas familias no son capaces de registrar gastos en un diario, y ni siquiera de proporcionar al entrevistador los datos requeridos.

Los ratios de la no respuesta debido a ausencia y no contacto suelen ser más bajos en las encuestas EGF que en otras encuestas

familiares. Esto es porque los entrevistadores generalmente intentan una y otra vez de hacer contacto con las unidades ausentes (mucho más que en otras encuestas), si han vuelto a casa dentro del período asignado para su investigación, o si se permiten algunas retrasos. También, se invierte mucho más esfuerzo en las EGF para reducir el número de no contactos.

Los ratios de la no respuesta no son iguales en las diferentes EGF, dependen del diseño de la encuesta, de la disposición de la población de revelar los detalles de la economía personal y de la familia, de la calidad del trabajo de campo y de otros factores. Una comparación de los ratios de la no respuesta de varios países puede ser engañosa, a causa de las diferencias en la definición de la no respuesta (por ejemplo, si se incluyen o no los de no contacto), y debido a las formas alternativas de calcular los ratios de la no respuesta (por ejemplo, de todas las unidades de la muestra o sólo las unidades elegibles). Además, en las EGF donde se permite más imputación para las unidades con mala calidad de información, se dejan menos no respuestas. En algunas EGF, las unidades que proporcionan datos para el diario o para el cuestionario se consideran como encuestados, mientras que en otras encuestas los encuestados tienen que completar ambas cosas. No obstante, es interesante señalar que muchos países desarrollados informan de ratios de la no respuesta parecidos, como en los siguientes ejemplos.

En el Reino Unido, los ratios de la no respuesta varían poco de año en año y están en el rango de 26% a 32%, la mayoría de los cuales es debido a negaciones iniciales y abandonos durante el período de dos semanas del diario, o inmediatamente después de la entrevista inicial (Elliot, 1991).

En Australia, donde se usa un diario personal de dos semanas, con información adicional recogida en entrevistas con cada miembro adulto de la familia y la familia entera, el ratio de la no respuesta en la encuesta de 1988-89 era 21%. Pero para otro 15.6% sólo se había obtenido parte de la información, imputando la que faltaba (Harrison, 1991).

En Hungría, en la encuesta de 1989, con un diario de dos meses y una entrevista suplementaria, el ratio de la no respuesta era de 25%, del que 70% eran negaciones (Elteto, 1991).

En Francia, el ratio de la no respuesta de la EGF de 1978-79, donde se recogían los datos en un diario personal de diez días y durante tres entrevistas, era aprox. 25%. En una encuesta piloto que se realizó en 1977, para un diario de siete días el ratio de la no respuesta era de 20%, para una de diez días, 25%, para un diario de catorce días, 30% y un poco más de 50% para un diario de uno o dos meses. En esta encuesta piloto, el ratio de abandono después de haber iniciado un diario de un mes era muy bajo y la mayoría de la no respuesta era de familias que se negaban a cooperar desde el principio, o se negaban cuando el entrevistador les pidió que mantuvieron diarios (Glaude, 1982).

En los Estados Unidos, donde se llevan a cabo dos encuestas por separado, el ratio de la no respuesta es de aprox. 15% para los

cuatro trimestres consecutivos de la Encuesta de Entrevistas. En la Encuesta del diario, aprox. 18% no completan ni siquiera uno de los dos diarios de una semana, y aprox. 9% completan sólo uno (Tucker, 1992).

Existen, por supuesto, excepciones, por ejemplo, en Holanda se informa de un ratio de la no respuesta altísimo de 80% (!) (de Heer, 1991).

En Israel, los ratios de la no respuesta incrementaron regularmente hasta 1979-80. En la EGF de 1986-87, se redujó el período del diario de un mes a dos semanas y el período de referencia de la entrevista de memoria también se redujo de 12 meses a 3 meses para la mayoría de los artículos. Estos cambios en el diseño de la encuesta junto con otras mejoras, daba como resultado una reducción sustancial en los ratios de la no respuesta. Los ratios de la no respuesta de las cuatro EGF realizadas recientemente en Israel desde 1968-69 se presentan en la Tabla 6.1, por grupo primario de la no respuesta.

**Tabla 6.1**  
**Ratios de Respuesta y No Respuesta por Razones Principales**  
**EGF de Israel 1968-69 a 1986-87**

| Grupo de No respuesta                    | 1968/69     | 1975/76     | 1979/80     | 1986/87     |
|--|-------------|-------------|-------------|-------------|
| <b>Total de familias elegibles (lds)</b> | 100.0       | 100.0       | 100.0       | 100.0       |
| <b>Total de familias no respondidas:</b> | <u>28.2</u> | <u>34.7</u> | <u>41.0</u> | <u>21.5</u> |
| <b>Negaciones</b>                        | 19.4        | 15.9        | 18.4        | 13.6        |
| <b>Abandonos</b>                         | 3.5         | 11.8        | 16.4        | 1.5         |
| <b>No capaces</b>                        | 4.7         | 5.5         | 4.8         | 3.6         |
| <b>Ausentes</b>                          | 0.6         | 1.5         | 1.4         | 2.8         |
| <b>Total encuestados:</b>                | <u>71.8</u> | <u>65.3</u> | <u>59.0</u> | <u>78.5</u> |
| <b>Inadecuados para procesar</b>         | 2.0         | 3.0         | 3.2         | 2.4         |
| <b>Adecuados para procesar</b>           | <u>69.8</u> | <u>62.3</u> | <u>55.8</u> | <u>76.1</u> |

En las tres primeras encuestas, el ratio total de la no respuesta incrementó de 28.2% en la encuesta de 1968-69 a 34.7% en la encuesta de 1975-76, y era muy alto, 41% en la encuesta de 1979-80. Aunque todas las encuestas tenían un diseño parecido, el incremento principal se debía a abandonos, 3.5%, 11.8% y 16.4%, respectivamente. Los abandonos ocurrieron cuando el encuestado no completó el diario de un mes (a menos que cuando sólo faltaban algunos días), o incluso cuando había completado el diario pero faltaba la entrevista final. El porcentaje de casos cuyos datos resultaron ser inadecuados para procesar también incrementó ligeramente.

El aumento del ratio de la no respuesta durante estas tres encuestas se podría explicar por el incremento de inflación de precios. El ratio anual de la inflación de los precios era menos que 3% en 1968-69, aprox. 30% en 1975-76 y aprox. 140% (!) en 1979-80. No cabe

duda de que cuanto más alta sea la inflación de precios, más difícil será mantener registros en un diario y proporcionar datos sobre gasto e ingreso por memoria para la entrevista. Además cuando hay un ratio muy alto de inflación, el público puede estar menos dispuesto a proporcionar información sobre actividades económicas a un departamento del gobierno que es, en su opinión, responsable de la situación económica.

En 1986-87, el ratio de la no respuesta se redujó a 21.5%, mayoritariamente debido a una reducción considerable de abandonos (sólo 1.5%). También había menos casos de unidades incapaces o inadecuadas para el procesamiento, 1% cada uno. Probablemente se debía al recorte del período de registro en los diarios y período de recuerdo, pero no era la única razón. Se incrementó el número de visitas del entrevistador a las familias durante el período del diario, así que, el entrevistador visitó a las familias como media, cinco veces, aparte de las entrevistas inicial y final. El número de visitas a cada familia se determinaba en el campo, según la ayuda que necesitaba la familia para registrar el gasto. Incluso en algunas familias, el entrevistador les visitaba a diario. También, para la encuesta de 1986-87 se usaba una estrategia diferente de incentivos, que resultó ser muy efectiva (ver sección 6.4). Además, esta encuesta se realizó bajo una inflación anual de precios moderada, aprox. 20%.

### 6.3 Características de la No Respuesta

Como se mencionaba antes, el sesgo debido a la no respuesta depende tanto del ratio de la no respuesta como de la diferencia entre las medias de las respuestas y de las no respuestas. Así que, el problema del alto ratio de la no respuesta en las EGF es de una importancia considerable cuando la no respuesta no se distribuye igualmente entre los subgrupos de la población que difieren con respecto a las variables bajo investigación.

Mientras que no hay ningún problema en determinar el ratio total de la no respuesta de la encuesta misma, no hay ninguna forma de obtener estimaciones sin sesgos de las variables de la encuesta para las no respuestas. Sin embargo, para reducir el sesgo de la no respuesta en lo posible, por ejemplo, ponderando los procedimientos, se intenta caracterizar a la no respuesta por características que se suponen estén correlacionadas con las variables de la encuesta. No es tarea fácil, porque para la mayoría de la no respuesta, la misma encuesta no proporciona tal información. Normalmente, sólo se dispone de una cantidad muy limitada de información del marco de la muestra para cada unidad de la muestra original. Los intentos para recoger datos básicos en las EGF (y en la mayoría de las otras encuestas) para la no respuesta no suelen tener mucho éxito.

Se puede llevar a cabo un estudio de los diferentes ratios de la no respuesta entre los diferentes subgrupos cuando se realiza la EGF en el mismo momento (o muy cerca) que un censo de la población, o cuando se dispone de un registro de la población con datos detallados para cada unidad. Este tipo de estudio se realizó en el Reino Unido para la EGF de 1971 y de 1981, encajando la muestra de la encuesta (las respuestas y las no respuestas) con el censo del mismo año. En ambas encuestas se encataban un 95% de las familias

elegibles. Los censos proporcionaron muchas características de cada unidad encajada y se obtenían dos conjuntos de distribuciones por cada variable del censo, uno para las respuestas y el segundo para la no respuesta.

Se hacían comparaciones univariadas (Kemsley, 1975 y Redpath, 1986) y también se realizó un análisis multivariado basado en los mismos datos, examinando el efecto conjunto de las características de las familias sobre el ratio de respuesta (Elliot, 1991). Algunos resultados de estos estudios para la encuesta de 1981 (basada en una muestra de 2.500 familias) se presentan en la Tabla 6.2.

La impresión de estos estudios es que los ratios de la no respuesta disminuyan con el tamaño de la familia, coincidiendo con la disminución del ratio del número de hijos dependientes.

Por otra parte, las familias sin hijos, particularmente aquellas familias de personas mayores parecían menos dispuestos a cooperar. Se asociaban diferencias sustanciales con la edad del cabeza de familia. Una regresión lineal de los ratios de respuesta y grupos de edad (para la encuesta de 1971) alcanzó un coeficiente de regresión de -2.22 con un coeficiente de correlación de 0.92. Esto es aceptable, dado que a menudo las personas mayores no están dispuestos a llevar a cabo la difícil tarea de mantener un diario, o son incapaces de entender la tarea, o tienen miedo de cometer errores. También había altos ratios de no respuesta entre las familias de adultos que conviven relativamente independientemente unos de los otros, y donde puede ser que no quieran comprometerse a cooperar con la contabilidad o a participar conjuntamente en la encuesta. La respuesta más baja de las familias de personas autónomas comparada con las familias de empleados puede deberse a una mayor desgana de revelar los detalles de los ingresos y gastos en una encuesta gubernamental.

Como ya hemos mencionado, la caracterización de la no respuesta por medio de una comparación con un censo es posible cuando se llevan a cabo las EGF al mismo tiempo que el censo. Esto no es el caso en muchos países donde las EGF sólo se realizan una vez en varios años y rara vez coinciden con los censos. Por lo tanto, la caracterización de la no respuesta sólo se puede hacer indirectamente, cuando se disponen de estimaciones externas fiables de la distribución de la población por características de la familia. Este tipo de análisis se hizo con las EGF de Israel.

Tabla 6.2  
Ratios de la No Respuesta por Características de la Familia  
EGF del Reino Unido, 1981

| Características de la familia       | Ratio de la no respuesta |                   |
|-------------------------------------|--------------------------|-------------------|
|                                     | %                        | Relativo al total |
| <b>TOTAL</b>                        | 26                       | 1.00              |
| <b>Tamaño de familia:</b>           |                          |                   |
| 1                                   | 29                       | 1.12              |
| 2                                   | 25                       | 0.96              |
| 3                                   | 29                       | 1.12              |
| 4+                                  | 23                       | 0.88              |
| <b>Edad del cabeza:</b>             |                          |                   |
| 16 - 35                             | 16                       | 0.62              |
| 36 - 45                             | 23                       | 0.88              |
| 46 - 55                             | 31                       | 1.19              |
| 56 - 65                             | 30                       | 1.15              |
| 66 - 70                             | 25                       | 0.96              |
| 71+                                 | 35                       | 1.35              |
| <b>Tipo de familia:</b>             |                          |                   |
| Padre solo con hijos dependientes   | 17                       | 0.65              |
| Matrimonio con hijos dependientes   | 22                       | 0.85              |
| Matrimonio sin hijos                | 27                       | 1.04              |
| Padre solo con hijos independientes | 31                       | 1.19              |
| Matrimonio con hijos independientes | 39                       | 1.50              |
| <b>Nivel de empleo:</b>             |                          |                   |
| Empleados                           | 24                       | 0.92              |
| Autónomos                           | 35                       | 1.35              |
| <b>Títulos Educacionales:</b>       |                          |                   |
| Alguno                              | 14                       | 0.54              |
| Ninguno                             | 25                       | 0.96              |

Fuentes: Kemsley (1975), Redpath (1986) y Elliot (1991).

El tamaño esperado de la muestra del grupo k es  $E(n_k)=N_k f$ , donde  $N_k$  es el tamaño conocido (o estimado) de la población del grupo k y f es la probabilidad de muestreo de la muestra de la encuesta. El ratio estimado de la no respuesta del grupo k es  $(1-o_k)=1-r_k/E(n_k)$ , donde  $r_k$  es el número de encuestados en el grupo k. Por supuesto, estas estimaciones están sujetas a una varianza muestral. Tales estimaciones de los ratios de la no respuesta, por ejemplo, por tamaño de familia, se presentan en la Tabla 6.3 para las tres EGF de Israel.

**Tabla 6.3**  
**Ratios de la No Respuesta por Tamaño de Familia**  
**EGF de Israel de 1975-76 a 1986-87**

| Tamaño de familia | 1975/76 |                   | 1979/80 |                   | 1986/87 |                   |
|-------------------|---------|-------------------|---------|-------------------|---------|-------------------|
|                   | %       | Relativo al total | %       | Relativo al total | %       | Relativo al total |
| TOTAL             | 37.7    | 1.00              | 44.2    | 1.00              | 23.9    | 1.00              |
| 1                 | 45.9    | 1.22              | 60.3    | 1.36              | 41.1    | 1.72              |
| 2                 | 28.2    | 0.75              | 48.4    | 1.10              | 26.0    | 1.09              |
| 3                 | 37.7    | 1.00              | 48.5    | 1.10              | 20.7    | 0.87              |
| 4                 | 11.1    | 0.29              | 27.8    | 0.63              | 14.2    | 0.59              |
| 5                 | 32.6    | 0.86              | 37.2    | 0.84              | 18.0    | 0.75              |
| 6+                | 50.0    | 1.33              | 36.3    | 0.82              | 15.0    | 0.63              |

Las familias de tamaño 1 tienen ratios altos de la no respuesta en las EGF tanto en Israel como en el Reino Unido. Sin embargo, en las encuestas israelíes para diferentes tamaños de familia, las diferencias en los ratios de respuesta son mucho más sustanciales. Estos ratios para las familias de tamaño 1, relativos al ratio total de la no respuesta fueron de 1.1 en el Reino Unido en las EGF de 1971 y 1981, mientras que en Israel fueron de 1.2 en la encuesta de 1975-76, 1.4 en la de 1979-80 y 1.7 para la encuesta de 1986-87. Este incremento se puede explicar en parte por el incremento en la proporción poblacional de familias de una sola persona mayor entre las familias de tamaño 1. Por el contrario, las familias de tamaño 4 cooperaron constantemente más que otras. Se puede ver que los ratios de la no respuesta por tamaño de familia no eran constantes sobre el tiempo, por ejemplo, los ratios de la no respuesta de las familias grandes de 6 y más miembros, eran más bajos que la media en las dos encuestas recientes, pero tenían el ratio más alto de la no respuesta en la encuesta de 1975-76. Quizás esto se podría explicar por la proporción de personas con un nivel más bajo de educación en las familias numerosas, proporción bastante alta en 1975-76 y que desde entonces ha disminuido dramáticamente.

Las estimaciones de los ratios de la no respuesta se obtenían de una forma similar para algunas otras características de las familias, altamente correlacionadas con el nivel de gasto de la familia. Por ejemplo, se presentan en la Tabla 6.4 el ratio estimado de la no respuesta de grupos de familias, definidos por tamaño de familia y número de habitaciones en la vivienda, en la EGF de Israel de 1986-87. Estos resultados muestran que dentro del tamaño de la familia, el ratio de la no respuesta disminuye con la densidad de personas por habitación. El hecho que las familias con bajo nivel socio-económico estaban menos dispuestas a cooperar en la EGF de 1986-87, es un buen indicador.

Los ejemplos anteriores de los ratios diferenciales de la no respuesta entre grupos que se espere que tengan diferentes niveles de gasto, indican la seriedad de los posibles sesgos en las

estimaciones de la encuesta debido a la no respuesta. Por lo tanto, se invierte mucho esfuerzo en intentar reducir los ratios de la no respuesta: se lleva a cabo una amplia campaña de publicidad antes y durante la encuesta; se envía una carta a cada familia seleccionada, justamente antes de la visita del entrevistador, explicando la importancia de la encuesta y la importancia de su participación en la encuesta; y, los entrevistadores entregan folletos con más explicaciones de los propósitos y usos de la encuesta.

Además, los entrevistadores son entrenados a persuadir a las familias que participen en la encuesta y de asegurar a los encuestados de la confidencialidad. Las familias que al principio no están dispuestas a cooperar reciben cartas especiales de la oficina central de estadística y la visita del supervisor local para animarlas a cooperar. En adición, a menudo se usan varios tipos de incentivos, y su efecto se discute en la siguiente sección.

**Tabla 6.4**  
**Estimaciones de los Ratios de la No Respuesta**  
**por Tamaño de Familia y Número de Habitaciones,**  
**EGF de Israel 1986-87**

| Tamaño de familia | Número de habitaciones | Ratio de la no respuesta |                |                 |
|-------------------|------------------------|--------------------------|----------------|-----------------|
|                   |                        | %                        | Relativo a     |                 |
|                   |                        |                          | Tamaño Familia | Población Total |
|                   | TOTAL                  | 24                       | ..             | 1.0             |
| Tamaño 1          | Total                  | <u>41</u>                | <u>1.0</u>     | <u>1.7</u>      |
|                   | 1.0-1.5                | 52                       | 1.3            | 2.2             |
|                   | 2.0-2.5                | 40                       | 1.0            | 1.7             |
|                   | 3.0+                   | 39                       | 0.9            | 1.6             |
| Tamaño 2          | Total                  | <u>26</u>                | <u>1.0</u>     | <u>1.1</u>      |
|                   | 1.0-2.0                | 32                       | 1.2            | 1.3             |
|                   | 2.5-3.0                | 26                       | 1.0            | 1.1             |
|                   | 3.5+                   | 14                       | 0.5            | 0.6             |
| Tamaño 3          | Total                  | <u>21</u>                | <u>1.0</u>     | <u>0.9</u>      |
|                   | 1.0-2.5                | 26                       | 1.3            | 1.1             |
|                   | 3.0                    | 18                       | 0.8            | 0.7             |
|                   | 3.5+                   | 21                       | 1.0            | 0.9             |
| Tamaño 4          | Total                  | <u>14</u>                | <u>1.0</u>     | <u>0.6</u>      |
|                   | 1.0-2.5                | 24                       | 1.7            | 1.0             |
|                   | 3.0                    | 17                       | 1.2            | 0.7             |
|                   | 3.5+                   | 10                       | 0.7            | 0.4             |
| Tamaño 5          | Total                  | <u>18</u>                | <u>1.0</u>     | <u>0.8</u>      |
|                   | 1.0-2.5                | 29                       | 1.6            | 1.2             |
|                   | 3.0-3.5                | 23                       | 1.3            | 1.0             |
|                   | 4+                     | 12                       | 0.7            | 0.5             |
| Tamaño 6+         | Total                  | <u>15</u>                | <u>1.0</u>     | <u>0.6</u>      |
|                   | 1.0-3.0                | 27                       | 1.8            | 1.1             |
|                   | 3.5-4.0                | 19                       | 1.3            | 0.8             |
|                   | 4.5+                   | 10                       | 0.7            | 0.4             |

#### 6.4 Efectos de los Incentivos

Normalmente, las organizaciones de estadísticas oficiales no dan regalos ni pagan a los encuestados por participar en las encuestas. Sin embargo, a causa del relativamente alto ratio de la no respuesta en las EGF y a causa de la carga pesada que imponen estas encuestas en los participantes, siempre se ha considerado el uso de incentivos, pagos en metálico o varios tipos de regalo. También se ha considerado el momento ideal de ofrecer el regalo, durante la primera visita para animarles a participar, o al final de la investigación como un señal de agradecimiento. Se realizaron varios experimentos para evaluar los efectos de los diferentes tipos de incentivos sobre una mejora en los ratios de la respuesta.

Uno de los primeros experimentos sobre los efectos de los incentivos en las EGF se llevó a cabo en el Reino Unido en 1951 (Kemsley y Nicholson, 1960), en dos encuestas, cada una con un tamaño de muestra de aprox. 1.500 direcciones:

- (1) En la primera encuesta (Enero, 1951), se pidió a cada adulto (16 años y más) en cada familia (excepto los colegiales) que se dejase entrevistar para dar detalles de compras regulares e infrecuentes, y luego les pidió que registrasen en un diario todas las compras de la siguiente semana. Al final de la semana, los entrevistadores recogían los diarios y pedían más información sobre los ingresos. La muestra se dividió en tres submuestras del mismo tamaño y se asignaba a cada entrevistador el mismo número de direcciones en cada una de las tres submuestras. Al final de la primera entrevista se ofrecían varios tipos de pago por su cooperación: nada a la primera submuestra; 5 chelines a cada persona que cooperaba en la segunda submuestra; y 10 chelines a cada persona que cooperaba en la tercera submuestra.
- (2) En la segunda encuesta (Julio-Septiembre 1951), la muestra se dividió en dos submuestras. En la primera submuestra, se pidió a todos los adultos que mantuviesen un diario durante una semana, mientras que en la segunda submuestra, el diario era de cuatro semanas. Se pidió a los adultos de ambas submuestras que completasen un cuestionario adicional con detalles completos de sus ingresos.

Recibieron los pagos bajo la condición de llenar tanto el programa de ingresos como de gastos. En la primera submuestra, el pago de cada persona que cooperaba era de 2.5 chelines y con 7.5 chelines adicionales por persona si cooperaban todos los adultos de la familia, sumando a un pago máximo de 10 chelines por persona. En la segunda submuestra el pago era 2.5 chelines por persona y semana con 2.5 chelines adicionales por persona si todos los adultos de la familia cooperaban en esa semana, y 5 chelines adicionales por persona si todos cooperaban durante las cuatro semanas, sumando un pago máximo de 25 chelines por persona.

Los resultados de estos experimentos se presentan en la Tabla 6.5 y muestran claramente que los pagos a las personas que cooperan en las

encuestas de este tipo causan un sustancial incremento en los ratios de la respuesta.

En la primera encuesta, el efecto de los primeros cinco chelines parece ser ligeramente mayor que el efecto de los cinco chelines adicionales, que es sustancial (y estadísticamente significativo). El pago de diez chelines aumentó el ratio de la plena cooperación de 36% a 53%. En esa encuesta, pagaron a cada individual que cooperaba, sin tener en cuenta la cooperación de otros miembros de la familia. En consecuencia, se hicieron pagos al 14% de las familias donde sólo se obtenía una cooperación parcial.

En la segunda encuesta, los pagos a cada persona eran de mayor cantidad sólo si se obtenía plena cooperación de toda la familia. Esto llevó a una reducción considerable en el ratio de la cooperación parcial a sólo 3% de todas las familias para un diario de una semana.

Dado que no se pueden usar los datos obtenidos de una cooperación parcial, no se pueden considerar a estos casos como respuestas. No obstante, los resultados indican que el pago de un máximo de 10 chelines por persona bajo la condición de la participación de los otros miembros de la familia da como resultado un ratio similar de respuesta plena que el pago incondicional de 10 chelines por persona, pero el último sale más caro. Un pago reducido de 5 chelines por persona resultó en un ratio más bajo de plena respuesta, pero es, por supuesto, menos caro.

En la segunda encuesta se evaluaba el efecto sobre el ratio de la respuesta de pedir a las familias que mantuvieron un diario durante más de una semana. Los pagos no eran proporcionales a la cantidad de trabajo implicada. Por lo tanto, no es sorprendente que se obtenía un ratio más bajo de plena cooperación con un pago máximo de 25 chelines, comparado con el diario de una semana con un máximo de 10 chelines. De todas formas, 46% de las familias cooperaban con el diario de cuatro semanas, cuando se ofrecían pagos, mientras que sólo 36% cooperaban con el diario de una semana cuando no se ofrecía ningún incentivo. La conclusión general de este experimento era que ofrecer pagos a los participantes en las EGF tiene, sin duda, efectos positivos en el ratio de la respuesta.

**Tabla 6.5**  
**Ratios de Respuesta con Incentivos de Dinero**  
**EGF Experimental 1951, del Reino Unido,**  
**Oficina Central de Estadística**

| Submuestra                                  | Total familias contactadas |       | Porcentaje cooperando |         |      |
|---|----------------------------|-------|-----------------------|---------|------|
|   | Número                     | %     | Pleno                 | Parcial | No   |
| <b>Primera encuesta:</b>                    |                            |       |                       |         |      |
| 1) Diario 1 sem. sin pago                   | 479                        | 100.0 | 35.7                  | 16.5    | 47.8 |
| 2) Diario 1 sem. 5ch./persona               | 490                        | 100.0 | 46.7                  | 14.7    | 38.6 |
| 3) Diario 1 sem. 10ch./persona              | 500                        | 100.0 | 53.4                  | 13.8    | 32.8 |
| <b>Segunda encuesta:</b>                    |                            |       |                       |         |      |
| 1) Diario 1 sem. con máx. 10ch./persona(*)  | 466                        | 100.0 | 55.6                  | 3.0     | 41.4 |
| 2) Diario 4 sem. con máx. 25ch./persona(**) | 898                        | 100.0 | 46.4                  | 5.8     | 47.8 |

(\*) 2.5ch./persona-semana + 7.5ch. de plus; (\*\*) 2.5ch./persona-semana + 5ch. de plus. Basado en Kemsley y Nicholson (1960)

Sudman y Ferber (1971) realizaron un estudio piloto para la encuesta BLS de los Estados Unidos con el objetivo de estudiar los efectos de ofrecer regalos, en vez de una compensación de dinero. Las cuatro submuestras aleatorias recibieron diferente tratamiento:

- (1) Ningún regalo, excepto una carpeta de plástico y un bolígrafo para usar en el registro (con valor de aprox. 1.00\$) que no se refiere a un regalo aunque las familias se quedaron con ellos.
- (2) Se ofrecían una bandera de los Estados Unidos con soporte, posteres o un libro a elegir entre 40 (valor aprox. 5.00\$).
- (3) Se ofrecía un informe de las compras de la familia resumido en principales categorías y comparado con otras familias con ingresos parecidos.
- (4) Se regaló una gran carpeta de escritorio con pluma (valor aprox. 5.00\$) antes de que la familia empezara a mantener el registro.

La Tabla 6.6 muestra que la cooperación era constantemente más baja en el grupo que no recibió ningún regalo, indiferente a la duración del período del diario y que la magnitud de las diferencias incrementó con una mayor duración del período.

Para el período de cuatro semanas, menos que mitad de encuestados en el grupo sin regalo completaba las cuatro diarios (23%), comparado con todos los grupos que recibieron algún regalo (49%). Entre los regalos, el regalo inmediato, (la carpeta de escritorio) era más efectivo que cualquier de los otros regalos más tardíos (excepto para el período de cuatro semanas). De los dos regalos tardíos, la oferta del informe era ligeramente mejor que la oferta de la bandera o un libro. Los resultados de este estudio sugieren que un regalo

inicial es más efectivo a corto plazo, y se requiere algún refuerzo para un período más largo.

Tabla 6.6  
Ratios de Cooperación por Tipo de Regalo  
y por Duración del Período del diario  
porcentajes (\*)

| Grado de cooperación       | Sin regalo       | Con regalo | Tipo de regalo  |                 |                    |
|----------------------------|------------------|------------|-----------------|-----------------|--------------------|
|                            |                  |            | Bandera o libro | Resumen informe | Carpeta escritorio |
| <b>Mantenido al menos:</b> | <b>1 semana</b>  | 77         | 85              | 79              | 86                 |
|                            | <b>2 semanas</b> | 62         | 75              | 71              | 71                 |
|                            | <b>3 semanas</b> | 54         | 67              | 64              | 67                 |
|                            | <b>4 semanas</b> | 23         | 49              | 51              | 55                 |
|                            |                  |            |                 |                 | 39                 |

(\*) de aquellos que tenían el diario durante al menos ese período de tiempo.

Basado en Sudman y Ferber (1971)

Recientemente se realizó un experimento con incentivos en Suecia (Lindstrom 1991), durante la EGF de 1988. Desde 1958, las EGF de Suecia han dado algún tipo de compensación a las familias que participan en la encuesta. Sin embargo, el valor de la compensación ha disminuido con cada encuesta sucesiva; una de las razones es que se encontró que lo que importa es la promesa de alguna compensación más que su valor. Sin embargo, mientras que las EGF anteriores obtenían un ratio de respuesta de aprox. 70%, el ratio de la respuesta de la encuesta de 1988 bajó considerablemente, sólo 55% en las primeras tres meses de la encuesta, e incluso menos en el mes subsiguiente (45%). Durante la realización de la encuesta, se adoptaba una nueva estrategia de incentivos, empleando incentivos de antemano y regalando una calculadora de bolsillo. Además se incluyó en la encuesta un experimento controlado para evaluar el efecto de tales incentivos en el ratio de la respuesta. Como conclusión, decidieron seguir con la nueva estrategia.

Los incentivos también pueden llevar a un registro mejor de los encuestados. Un experimento para evaluar los efectos de este tipo se llevó a cabo en la Oficina del Censo de los Estados Unidos con la EGF de 1972-73 (Cowan, 1977) para una encuesta de 8 semanas, con tres tratamientos asignados aleatoriamente a tres submuestras: ningún pago, 5\$ y 10\$. Cada entrevistador manejó sólo un tratamiento, así que no se controlaban completamente los efectos del entrevistador. Cada submuestra incluía unas 500 familias que completaban uno o dos diarios de una semana. El gasto medio por familia, la proporción de encuestados que informó sobre gasto y la media del gasto registrado de las tres submuestras se comparaban para cada una de las 72 categorías principales de gasto.

En la Tabla 6.7 se presentan ejemplos de los resultados obtenidos de este estudio para el cual se derivan las estimaciones primarias de los diarios para categorías de gasto frecuente. Para todas estas

categorías (menos una), la proporción de las personas que informaron sobre gasto incrementó con el incentivo, y para casi todas el incremento era mayor, cuanto más cantidad tuviera el incentivo. El gasto medio semanal para los artículos registrados también aumentó con la cantidad del incentivo en la mayoría de las categorías. En consecuencia, los efectos de los incentivos sobre el gasto medio semanal era significativo, por ejemplo, para Comida Total el incremento era de 1.13, para Tabaco de 1.22, para Productos Lácteos de 1.1, etc.

Sin embargo, esto no significa necesariamente que los incentivos reducen los sesgos de los registros. La asunción que "más es mejor" no es siempre correcta. Por ejemplo, si la no respuesta es más frecuente entre las familias de tamaño 1, y su gasto medio semanal es más bajo que en otras familias, los incentivos deberían resultar en una media más baja si se incrementa el ratio de la respuesta.

**Tabla 6.7**  
**Efectos de los Incentivos**  
**Encuesta de Gastos de Consumo, EE.UU., 1972-73**

| Categorías de gasto<br>seleccionadas                                       | Cantidad de incentivo |              |              | Ratio al no pago |             |             |
|--|-----------------------|--------------|--------------|------------------|-------------|-------------|
|  | 0\$                   | 5\$          | 10\$         | 0\$              | 5\$         | 10\$        |
| <b>1. Porcentajes de encuestados que informaron de categoría de gasto.</b> |                       |              |              |                  |             |             |
| <b>Comida total en casa:</b>   | <u>83.4</u>           | <u>89.1</u>  | <u>87.1</u>  | <u>1.00</u>      | <u>1.07</u> | <u>1.04</u> |
| Productos panadería  | 78.7                  | 82.3         | 81.6         | 1.00             | 1.05        | 1.04        |
| Productos lácteos  | 82.4                  | 85.3         | 84.4         | 1.00             | 1.04        | 1.02        |
| Carne  | 75.4                  | 77.1         | 78.9         | 1.00             | 1.02        | 1.05        |
| Verduras   | 67.3                  | 68.8         | 71.3         | 1.00             | 1.02        | 1.06        |
| Bebidas no alcohólicas   | 67.5                  | 71.5         | 73.5         | 1.00             | 1.06        | 1.09        |
| <b>Comida y bebida fuera de casa</b>                                       | <u>64.5</u>           | <u>69.0</u>  | <u>75.2</u>  | <u>1.00</u>      | <u>1.07</u> | <u>1.17</u> |
| Bebidas alcohólicas  | 28.5                  | 29.6         | 30.0         | 1.00             | 1.04        | 1.05        |
| Tabaco   | 42.6                  | 50.6         | 49.2         | 1.00             | 1.19        | 1.15        |
| Suministros hogar  | 70.7                  | 73.9         | 79.0         | 1.00             | 1.05        | 1.12        |
| Ropa y calzado   | 43.2                  | 45.3         | 49.0         | 1.00             | 1.05        | 1.13        |
| <b>2. Gasto medio semanal registrado</b>                                   |                       |              |              |                  |             |             |
| <b>Comida total en casa:</b>   | <u>21.39</u>          | <u>22.68</u> | <u>23.15</u> | <u>1.00</u>      | <u>1.06</u> | <u>1.08</u> |
| Productos panadería  | 2.21                  | 2.33         | 2.38         | 1.00             | 1.06        | 1.08        |
| Productos lácteos  | 3.24                  | 3.36         | 3.48         | 1.00             | 1.04        | 1.08        |
| Carne  | 7.16                  | 7.41         | 7.43         | 1.00             | 1.03        | 1.04        |
| Verduras   | 2.21                  | 2.24         | 2.31         | 1.00             | 1.01        | 1.05        |
| Bebidas no alcohólicas   | 2.41                  | 2.46         | 2.60         | 1.00             | 1.02        | 1.08        |
| <b>Comida y bebida fuera de casa</b>                                       | <u>10.05</u>          | <u>10.14</u> | <u>10.97</u> | <u>1.00</u>      | <u>1.01</u> | <u>1.09</u> |
| Bebidas alcohólicas  | 6.42                  | 6.79         | 8.30         | 1.00             | 1.06        | 1.29        |
| Tabaco   | 4.46                  | 4.27         | 4.72         | 1.00             | 0.96        | 1.06        |
| Suministros hogar  | 3.41                  | 3.55         | 3.52         | 1.00             | 1.04        | 1.03        |
| Ropa y calzado   | 20.69                 | 21.55        | 21.94        | 1.00             | 1.04        | 1.06        |
| <b>3. Gasto medio semanal</b>  |                       |              |              |                  |             |             |
| <b>Comida total en casa:</b>   | <u>17.84</u>          | <u>20.21</u> | <u>20.16</u> | <u>1.00</u>      | <u>1.13</u> | <u>1.13</u> |
| Productos panadería  | 1.74                  | 1.92         | 1.94         | 1.00             | 1.10        | 1.11        |
| Productos lácteos  | 2.67                  | 2.87         | 2.94         | 1.00             | 1.07        | 1.10        |
| Carne  | 5.40                  | 5.71         | 5.86         | 1.00             | 1.06        | 1.09        |
| Verduras   | 1.49                  | 1.54         | 1.65         | 1.00             | 1.03        | 1.11        |
| Bebidas no alcohólicas   | 1.63                  | 1.76         | 1.91         | 1.00             | 1.08        | 1.17        |
| <b>Comida y bebida fuera de casa</b>                                       | <u>6.84</u>           | <u>7.00</u>  | <u>8.25</u>  | <u>1.00</u>      | <u>1.08</u> | <u>1.27</u> |
| Bebidas alcohólicas  | 1.83                  | 2.01         | 2.49         | 1.00             | 1.10        | 1.36        |
| Tabaco   | 1.90                  | 2.16         | 2.32         | 1.00             | 1.14        | 1.22        |
| Suministros hogar  | 2.41                  | 2.62         | 2.78         | 1.00             | 1.09        | 1.15        |
| Ropa y calzado   | 8.94                  | 9.76         | 10.75        | 1.00             | 1.09        | 1.20        |

Los resultados de otros experimentos que se han llevado a cabo para medir la efectividad de los incentivos, indican que en general, los incentivos mejoran los ratios de la respuesta. Especialmente es el caso en las encuestas del diarios donde los encuestados tienen que comprometerse a una tarea continua, más que a una sola entrevista normal.

Por lo tanto, en muchos países se usan pagos de varias cantidades de dinero u otros tipos de incentivos, o las dos cosas. Por ejemplo, en el Reino Unido, se paga a cada persona; en Irlanda, la familia puede elegir entre un pago a cada miembro de la familia o la participación en un sorteo con grandes premios de dinero; en Dinamarca se paga a cada familia y también participan en un sorteo con 100 premios de dinero. En la encuesta de 1986-87 de Israel se regaló una balanza de cocina a cada familia en la muestra en la primera visita del entrevistador, incluso antes de que se asintieron a participar en la encuesta. No cabe duda que este tipo de incentivo era muy efectivo y, según los informes de los entrevistadores, ahorraba muchos esfuerzos en persuadir a las familias que se participasen en la encuesta. Aunque la balanza de cocina era un regalo impresionante, los entrevistadores se quejaban de su tamaño. Por lo tanto, para la encuesta de 1992-93, se eligió como regalo una calculadora de bolsillo, bastante sofisticada, de similar valor pero mucho más fácil de transportar. Se debería mencionar que en las encuestas anteriores se ofreció el regalo después de la entrevista final lo que no era muy efectivo. Por otra parte, hay países sin incentivos de ningún tipo, como los Estados Unidos, Italia, Grecia y otros.

En conclusión, aunque está claro que los incentivos reducen la no respuesta, y probablemente también mejoran la calidad de los registros, no hay una forma fácil de cuantificar los beneficios derivados del uso de incentivos comparado con su coste. Una posible alternativa es la de usar los incentivos sólo para ciertos subgrupos de la población, sujetos desde el principio a ratios más altos de la no respuesta (sin embargo, esto parece un poco injusto). Siempre podría haber dudas que fuera mejor invertir el gasto de incentivos en mejorar el trabajo de campo, por ejemplo, mejor entrenamiento de los entrevistadores, o incrementar el número de visitas del entrevistador a las familias durante el período del diario.

## CAPITULO 7. PONDERACION

### 7.1 La Ponderación en las Encuestas de Gastos Familiares

En las EGF, al igual que en todas las encuestas de muestreo, se determina una ponderación para cada unidad de la muestra que contesta para expresar el número de unidades en la población representadas por esa unidad. El procedimiento de ponderación tiene como objetivo minimizar el MSE (mean square error) de las estimaciones de la encuesta, comprendiendo la varianza muestral y los sesgos de la muestra neta, principalmente debidos a la no respuesta. A causa de los ratios relativamente altos de la no respuesta en las EGF y porque estas encuestas proporcionan información en detalle, la ponderación tiene una importancia especial.

Para la ponderación, primero, el inverso de la probabilidad de selección de cada familia determina una ponderación inicial. Luego, se ajustan las ponderaciones iniciales y se determinan las ponderaciones finales por medio de un procedimiento complejo de ponderación de multi-etapa.

Un ajuste que se usa a menudo es una corrección de los ratios diferenciales de la no respuesta dentro de las áreas geográficas. Puede no ser muy efectivo en las EGF dado que la no respuesta es casi totalmente debido a negaciones a cooperar y que el área residencial no caracteriza necesariamente su gasto.

Otro ajuste que se requiere a menudo es una corrección de la distribución desigual de la muestra neta sobre el período de la encuesta. En la práctica, incluso cuando inicialmente la muestra se distribuye igualmente sobre los subperiodos del período total de la encuesta, la muestra neta no se equilibra sobre el tiempo. Como las variables que se investigan en las EGF varían con el tiempo, esa desviación puede llevar a sesgos de naturaleza estacional en las estimaciones de la encuesta. Por ejemplo, una demora por razones de ausencias puede causar un sesgo mayor en las estimaciones de los gastos de vacaciones. Hay varias formas de reducir el sesgo en el tiempo por el proceso de ponderación, asignando diferentes ponderaciones a los diferentes períodos investigados para la misma familia, o teniendo una ponderación para cada familia.

Algunas veces, también se requieren ajustes adicionales para ciertos subgrupos que no se cubrieron completamente en la muestra. Un ejemplo es un ajuste para compensar no haber puesto al día la muestra de nuevas construcciones durante todo el período de la encuesta.

Unas etapas importantes en la ponderación son los ajustes para características claves correlacionadas con las variables investigadas. Se puede hacer esto, empleando los mismos datos de la encuesta, sólo cuando se realicen esfuerzos especiales para recoger estas características claves de la no respuesta, y que tengan éxito, o cuando el marco de la muestra puede proporcionar los datos necesarios. Luego, se aplican factores correctivos a los encuestados para compensar a los no encuestados similares.

Alternativamente, se emplea ampliamente en las EGF ajustes que aprovechan de una distribución externa de la población de la encuesta por sus características claves, es decir, post-estratificación y sus variantes. En las siguientes secciones se discuten algunos temas metodológicos relacionados con esos ajustes.

## 7.2 Efectos de la Post-Estratificación

En general, para una muestra simple aleatoria de tamaño  $n$  de una población de tamaño  $N$  con  $r$  encuestados en la encuesta ( $r \leq n$ ), la estimación simple es, de hecho, la estimación de la media de la población de encuestados:

$$\bar{y}_S = \frac{\sum y_i}{r}.$$

Se puede aplicar la post-estratificación si se puede clasificar a cada encuestado en una de las post-estratas,  $h=1, 2, \dots, H$ , con tal que se separen las ponderaciones de la población  $W_h = N_h/N$ . Entonces, la estimación de la post-estratificación de la media de la población es

$$\bar{y}_{PS} = \sum W_h \bar{y}(R)_h, \quad \text{donde } \bar{y}(R)_h \text{ son las estimaciones simples basadas en}$$

$r_h$  encuestados en la estrata  $h$ .

También se puede aplicar la post-estratificación usando estimaciones fiables de  $W_h$  que se puede derivar de otra encuesta de una escala mucho mayor, sustituyendo  $\hat{W}_h$  por  $W_h$ .

Aunque se puede aplicar la post-estratificación a diseños más complejos de muestreo y a otros tipos de estimaciones, la siguiente presentación es, para simplificar, limitada al caso sencillo sin perder la generalidad.

Si los ratios de respuesta son  $\phi_h = r_h/n_h$  y  $\phi = r/n$  para la estrata  $h$  y para la muestra total, se puede expresar la estimación simple como

$$\bar{y}_S = \sum w_h (\phi_h/\phi) \bar{y}(R)_h, \text{ where } w_h = n_h/n.$$

Mientras que con la estimación simple, todos los encuestados tienen implícitamente la misma ponderación, con la post-estratificación se pueden dar diferentes ponderaciones a las diferentes post-estratas. Las ponderaciones no son necesariamente iguales, en primer lugar dado que las  $n_h$  son variables aleatorias, y en segundo lugar, debido a la posible variación de los ratios de respuesta en las diferentes estratas.

Vamos a examinar los efectos de la post-estratificación, por una clasificación dada en post-estrata, en las estimaciones de la encuesta.

### a) Los Efectos sobre la Varianza Muestral

El ratio de las varianzas muestrales esperadas de la estimación de post-estratificación y de la estimación simple puede constar de tres componentes:

$$\frac{E[\text{Var}(\bar{y}_{PS})]}{E[\text{Var}(\bar{y}_S)]} = (1-R^2)(1+\delta)(1+D).$$

A continuación se supone que  $N_h$ , los tamaños de la población de la estrata sean bastante grandes, es decir,  $1/N_h$  son insignificantes,  $n$  también sea bastante grande, pero los  $fpc$  son insignificantes y  $S^2_h \approx S^2(R)_h$ , es decir, las varianzas de los encuestados y de la no respuesta son similares.

- (1) El primer componente,  $1-R^2$ , es la reducción de la varianza muestral que se hubiera obtenido por muestreo estratificado proporcional comparado con un muestreo simple aleatorio, bajo una respuesta completa.

$$1-R^2 = \frac{\overline{\text{Var}}(y_{ST} | \phi=1)}{\overline{\text{Var}}(y_S | \phi=1)} \approx S^2_w / S^2 = 1 - S^2_b / S^2$$

donde  $S^2_w$  es la varianza dentro de la estrata,  $S^2_b$  la varianza entre estratas y  $S^2$  la varianza general, así que  $0 \leq R^2 \leq 1$ . Una clasificación en post-estrata que produce un valor más alto de  $R^2$ , es decir, con una divergencia más alta entre las medias de las estratas, lo que resultará en una mayor reducción de la varianza.

- (2) El segundo componente,  $1+\delta$ , es el incremento en la varianza debido al hecho de que bajo una post-estratificación la muestra no se selecciona automáticamente como una estratificada.

$$1+\delta = \frac{\overline{\text{E}}[\text{Var}(y_{PS})]}{\overline{\text{E}}[\text{Var}(y_{ST})]} \approx \lambda / n,$$

donde  $\lambda$  es el tamaño media de la muestra por estrata y  $\lambda$  que puede ser no insignificante, es el ratio entre  $(1/\lambda) \sum S^2_h / \phi_h$  y  $\sum W_h S^2_h / \phi_h$ .  $\lambda=1$  cuando  $S^2_h$  y  $\phi_h$  son iguales para todos los  $h$ . Sin embargo, no es probable que se cumpla la segunda condición. Así que, para un tamaño dado de muestra,  $\lambda$  es minimizado, primero, por limitar las post-estratas, con tal que el tamaño medio de la muestra en la post-estrata final sea bastante grande. Segundo, por minimizar  $\lambda$ , o al menos por evitar  $\lambda > 1$  al construir las post-estratas con tamaños similares, por ejemplo, para  $n_h \geq 25$ , la varianza incrementará en comparación con una muestra estratificada proporcional, aproximadamente por 1.04.

- (3) El tercer componente,  $1+D$ , es el efecto sobre la varianza, resultado de la variación en los ratios de respuesta entre las estratas.

$$1+D = \left[ \frac{\overline{\text{Var}}(y_S | \phi=1)}{\overline{\text{E}}[\text{Var}(y_S)]} \right] \left[ \frac{\overline{\text{E}}[\text{Var}(y_{ST})]}{\overline{\text{Var}}(y_{ST} | \phi=1)} \right].$$

Se puede mostrar que

$$1+D \approx \phi \frac{\sum W_h S^2_h / \phi_h}{\sum W_h S^2_h} \approx \frac{\overline{\text{Var}}(y_{OPT}) [1 + \text{Relvar}(S_h G_h)]}{\overline{\text{Var}}(y_{ST})}$$

donde,  $G_h = N_h / E(r_h)$  son las ponderaciones esperadas y  $\text{Var}(y_{OPT}) = (\sum W_h S^2_h)^2 / n$  es la varianza muestral para un muestreo estratificado óptimo de tamaño  $n$ . Para el caso donde  $S^2_h$  son similares para todos  $h$ ,

$1+D \approx S^2[1+Relvar(G_i)]$ , y se usa ampliamente como una aproximación para la contribución de las diferentes ponderaciones al incremento de la varianza muestral. Por lo tanto, para evitar un incremento de la varianza es deseable evitar post-estratas con ratios de la no respuesta demasiado altos.

En un caso raro, donde hay una alta correlación positiva entre  $S_h$  y  $\phi_h$ , para que  $S_h/\phi_h$ , son casi iguales para todas las estratas, la asignación de la muestra entre las post-estratas, de hecho, puede llegar a ser óptima y  $1+D<1$ . Si es así, este componente puede incluso contribuir a una disminución de la varianza.

- (4) Un incremento adicional en la varianza muestral, sería cuando se usan las estimaciones de las ponderaciones de población,  $W^h$ . Este incremento adicional es aproximadamente  $S^2_{b/n'}$ , donde  $n'$  es el tamaño de la muestra de que se estiman  $W^h$ . Así que, la varianza muestral incrementará cuanto más pequeño sea  $n'$ , y cuanto más grande sea la varianza entre las post-estratas, es decir, cuando  $R^2$  sea mayor. Se debería señalar que para las estimaciones de los subgrupos este efecto adicional será menor, cuanto más pequeño sea el subgrupo.

Dado que el incremento en la varianza muestral por usar estimaciones de  $W_h$ , puede dominar sobre la disminución de la varianza que se logra con una mejor clasificación en post-estratas, se justifica el uso de la post-estratificación, sólo si es para un ratio dado de  $n/n'$ ,

$$R^2 > \frac{(1+\delta)(1+D) - 1}{(1+\delta)(1+D) - n/n'}$$

#### b) Los Efectos sobre el Sesgo de la No Respuesta

La reducción del sesgo de la no respuesta por post-estratificación depende de los siguientes componentes:

- (1) El primer componente es la divergencia de los ratios de respuesta,  $\phi_h$ , entre las post-estratas. En el caso raro de respuestas completas para todas las estratas ( $\phi_h=1$ ), tanto las estimaciones simples como las de post-estratificación no tienen sesgo. Cuando los ratios de respuesta son iguales para todas las estratas, ambas estimaciones tienen el mismo sesgo. Por lo tanto, el uso de la post-estratificación para reducir el sesgo de la no respuesta se debería considerar sólo cuando las post-estratas se construyen para que los ratios de respuesta no sean iguales para todas las estratas.
- (2) El segundo componente es la divergencia entre la media de las post-estratas. Cuanto más alto sea esta divergencia, es decir, cuanto más homogéneas sean las post-estratas, se logrará más reducción del sesgo por post-estratificación.
- (3) El tercer componente depende de las diferencias entre las respuestas y las no respuestas dentro de la post-estrata. Cuanto más similitud se espere entre los encuestados y la no respuesta

dentro de la post-estrata, más homogéneas serán las post-estratas, y entonces, menos sesgado será la estimación de post-estratificación. Sin embargo, estas diferencias pueden no ser insignificantes y la estimación post-estratificación permanecerá sesgada.

- (4) El cuarto componente depende de la relación entre los tres componentes anteriores. La post-estratificación reducirá el sesgo sólo bajo ciertas condiciones, si no, se pueden producir sesgos mayores que con la estimación simple.
- (5) Otra fuente de sesgo es cuando se usan  $W^h$  sesgados para la post-estratificación. Si el sesgo de esta fuente es sustancial, puede ser que no se justifique el uso de la post-estratificación.

Todos los componentes anteriores, tanto los de efectos sobre la varianza como sobre el sesgo, se deberían tener en cuenta cuando se construyen las post-estratas, y cuando se decide si se debería usar la post-estratificación o no. Tanto la varianza muestral como el sesgo de la no respuesta se reduciría sólo con  $R^2$  suficientemente alto y esto se puede lograr construyendo post-estratas homogéneas bajo las limitaciones descritas antes. Sin embargo, en general, hay que hacer concesiones mutuas entre los efectos sobre el sesgo y sobre la varianza.

### 7.3 Arboles de Clasificación de los Subgrupos Homogéneos

Normalmente, se determinan las post-estratas por una clasificación cruzada completa de las unidades por el número de variables clasificadoras, para la cual se disponen de datos externos o estimaciones fiables del tamaño de la población. Sin embargo, para un tamaño dado de muestra, el numero de variables clasificadoras y sus categorías es limitado, porque cada grupo final tiene que ser suficientemente grande. Otro método de ponderación que se llama "rastrillado", por el cual se controlan las distribuciones marginales de la población, permite el empleo de un gran número de variables, aunque sin una clasificación cruzada completa.

Alternativamente, la determinación de las post-estratas se puede basar en una clasificación jerárquica, construyendo un árbol de grupos homogéneos, donde cada rama del árbol se puede dividir de nuevo con una variable diferente y no necesariamente al mismo número de subramas. Esto hace posible el uso de un mayor número de variables clasificadoras y, potencialmente, proporciona una mejor clasificación de post-estratas homogéneas para un tamaño dado de muestra (Kantorowitz, 1990).

A continuación se describen dos procedimientos para construir un árbol de clasificación de grupos homogéneos para una sola variable de estimación.

- a) Un Procedimiento de Pasos Multirama - Este procedimiento de clasificación se realiza jerárquicamente según un conjunto de variables clasificadoras categóricas potenciales. En el primer paso se selecciona la variable clasificadora que da la reducción máxima para la varianza interior de la población entera, es decir, máxima  $R^2$ . El segundo paso es determinar para cada grupo definido por el primer paso, una variable clasificadora que dará máxima  $R^2$  dentro de ese grupo. Similarmente, el tercer paso es determinar las mejores variables clasificadoras, por separado para cada uno de los grupos determinado por el segundo paso, etc., paso a paso.

En cada paso, para cada grupo elegido para más clasificación, el número de categorías y los límites de las variables clasificadoras, no utilizados en los pasos anteriores, se redefinen. Esto se hace basándose en los tamaños de las muestras, en la similitud de sus medias, y según sus definiciones (por ejemplo, grupos de edades cercanos, ocupaciones similares, etc.).

Se aplica una norma de parada en cada paso para cada grupo, teniendo en cuenta los efectos acumulados y marginales, antes mencionados, para decidir si sería mejor parar la clasificación en el paso anterior, o combinar los grupos. Sin embargo, en la práctica, esto sólo se hace en las etapas más avanzadas.

Con este procedimiento, se construye un árbol cuyas ramas en las diferentes etapas no tienen necesariamente el mismo número de subramas, dado que esto depende del número de categorías de la variable clasificadora usada para cada sub-rama. Además no todas las ramas llegan hasta arriba porque la "grosura" de la rama no siempre permite el uso de todas las variables clasificadoras. Una variable clasificadora se puede usar sólo una vez en los diferentes nodos de la misma rama.

Esta clasificación no es única porque se implican muchas consideraciones judiciales. Por ejemplo, la elección de la variable clasificadora en cada paso depende claramente de la categorización de las diferentes variables clasificadoras. En cada paso también es posible considerar si clasificar según una variable con un grado considerable de categorización o limitar el número de categorías y así, hacer posible más clasificación con una variable clasificadora adicional en el siguiente paso. Además, al igual que en la situación de la regresión por pasos, tal clasificación no es necesariamente la clasificación óptima. Sin embargo, esto no disminuye las ventajas de usar esta clasificación incluso si siempre está la posibilidad de construir una mejor.

- b) Un Procedimiento de Arbol Binario, Basado en CART - Primero, el software CART (Breiman, Friedman, Olshen y Stone, 1984) se usa para obtener un árbol binario, es decir, cuando se divide una rama del árbol, siempre se divide en dos subramas. Así que, la misma variable clasificadora se puede utilizar más de una vez en los diferentes nodos de la misma rama. Las variables clasificadoras, inicialmente, pueden ser cuantitativas o categóricas, y para cada una de ellas, en cada nodo, se

determinan los dos mejores subgrupos y la mejor variable clasificadora para minimizar la varianza interior de ese nodo, es decir, maximizar  $R^2$ .

Porque un árbol demasiado detallado no es adecuado para usos prácticos, CART "poda" el árbol, según una norma predeterminada de poda que depende del equilibrio entre el beneficio marginal de reducir la varianza interior y una cierta medición de la complejidad del árbol. Se pueden poner otras limitaciones opcionales para normas de parada, tal como definir el tamaño máximo de la muestra para más clasificación.

El árbol CART proporciona una clasificación óptima (bajo ciertas limitaciones), sin embargo, sólo respecto a  $R^2$ . Por lo tanto, el árbol resultante no es siempre satisfactorio para la post-estratificación, porque no hay una forma de controlar los otros efectos, como hay con el árbol multirama de pasos. Así que hay que modificar el árbol CART. Por ejemplo, para obtener grupos finales de cierto rango de un tamaño de muestra, se necesita un árbol muy detallado y hay que podarlo manualmente.

Se puede aplicar el procedimiento CART sólo a una muestra auto-ponderada, mientras que el procedimiento multirama de pasos se puede aplicar a los datos de una muestra con ponderaciones desiguales.

Ambos procedimientos son adecuados para una sola variable objetivo. Sin embargo, la mayoría de las encuestas son de multiusos y suministran estimaciones para una variedad de variables. Generalmente, no habría un solo árbol de clasificación mejor para todas las variables objetivos, y claramente, una clasificación en grupos homogéneos con referencia a una sola variable será beneficioso para esa variable pero puede ser inefectivo para las demás variables, a menos que sean altamente correlacionadas con la variable elegida.

Una forma es elegir una variable estimada para "representar" a las demás (la variable más importante, o entre todas las importantes la que tenga el mayor error muestral, etc.). Otra forma es construir una sola variable sintética basada en un número de variables estimadas "representativas" y luego construir un solo árbol. Se pueden construir varios tipos de variables sintéticas, por ejemplo, la técnica del Componente Principal.

Una solución alternativa es utilizar uno o varios criterios "combinados" para elegir la mejor variable clasificadora para cada grupo. Sin embargo, esto sólo se puede aplicar bajo el procedimiento multirama de pasos. Estos criterios se pueden basar en funciones de  $R^2$  (ó en  $(1-R^2)(1+D)(1+D)$ ) que se obtienen para las diferentes variables objetivos de cada paso de la construcción del árbol. Vamos a examinar algunos criterios sencillos.

Deja que  $R^2_{ij}$  sea el valor de  $R^2$  para la variable estimada j-th ( $j=1,2,\dots,J$ ) y para la variable clasificadora i-th ( $i=1,2,\dots,I$ ). La variable clasificadora i con el valor más alto de una de las siguientes es la mejor.

- (1) El número de variables estimadas para el que  $\text{Max}_i(R^2_{ij})$  se obtiene - con esto todas las variables estimadas tienen la misma influencia, incluso aquellas para que  $R^2$  es pequeño para todas las variables clasificadoras.
- (2) La media  $\bar{R^2_i} = \frac{\sum R^2_{ij}}{J}$  - con esto, las variables estimadas para que los  $R^2$  son altos contribuirán más a la determinación de las variables clasificadoras. Esto es equivalente a minimizar el total de las "pérdidas" de todas las variables estimadas si se hace uso de una variable clasificadora diferente a la mejor para cada una de ellas.
- (3) La media  $\bar{R^2_i} = \frac{\sum R^2_{ij}/\text{Max}_i(R^2_{ij})}{J}$  - con esto, incluso las variables estimadas para que  $R^2$  es bajo para todas las variables clasificadoras, pero con un alto ratio, pueden ejercer mucha influencia sobre la elección de la variable clasificadora. Esto es equivalente a minimizar el total de las "pérdidas relativas" de todas las variables estimadas al elegir una variable clasificadora diferente a la mejor para cada una de ellas.

Se pueden usar variaciones de estos criterios al ponderar las variables estimadas, por ejemplo, por calificar su importancia, o por su varianza, o por la varianza que queda en cada paso de la clasificación. Se pueden aplicar criterios combinados de este tipo sólo a un número limitado de variables de estimación, porque puede disminuir la eficiencia para cada una de las variables de estimación por separado y también a causa de dificultades técnicas.

#### 7.4 Ilustración de los Arboles de Clasificación

Se ilustra aquí la construcción de un árbol, primero para una variable estimada como en la Encuesta sobre Ingresos de Israel. Esta encuesta se parece en muchos aspectos a la EGF y, por lo tanto, se usa aquí para propósitos ilustrativos. Luego, se ilustra la construcción de un árbol para un conjunto de variables estimadas como en la EGF de Israel. En ambas encuestas se aplicaba el procedimiento de pasos.

a) Construir un Árbol para el Ingreso Neto por Familia en la Encuesta sobre Ingresos

Hasta 1984, la post-estratificación en las Encuestas sobre Ingresos se basaba en una clasificación cruzada de tres variables: sexo y edad del cabeza económico de la familia, y tamaño de la localidad. Desde 1985, se ha implantado la post-estratificación usando un árbol de clasificación y las variables clasificadoras elegidas eran: tamaño de familia, número de personas empleadas en la familia, tamaño de la localidad, y las características del cabeza económico de la familia - nivel de empleo, edad, continente de nacimiento, años de estudios, ocupación, rama económica, y horas semanales de trabajo.

En la primera etapa, se clasificaba según los tres grupos primarios de cabezas de familia: empleados, autónomos y sin trabajo, dado que todas las estimaciones de la encuesta se dan por separado para estos grupos. En los pasos siguientes, resultó que se usaba todas las variables clasificadoras, excepto el tamaño de la localidad y las

horas semanales de trabajo porque eran las que menos contribuyeron a la reducción de la varianza.

Los valores acumulativos de  $R^2$  se dan en la Tabla 7.1, correspondiente a los varios pasos de la clasificación, para todas las familias y para cada uno de los tres subgrupos primarios.

**Tabla 7.1**  
 **$R^2$  Acumulativo por Etapa de Clasificación y por Tres Subgrupos**  
**Ingreso Neto por Familia en la Encuesta sobre Ingresos de Israel**

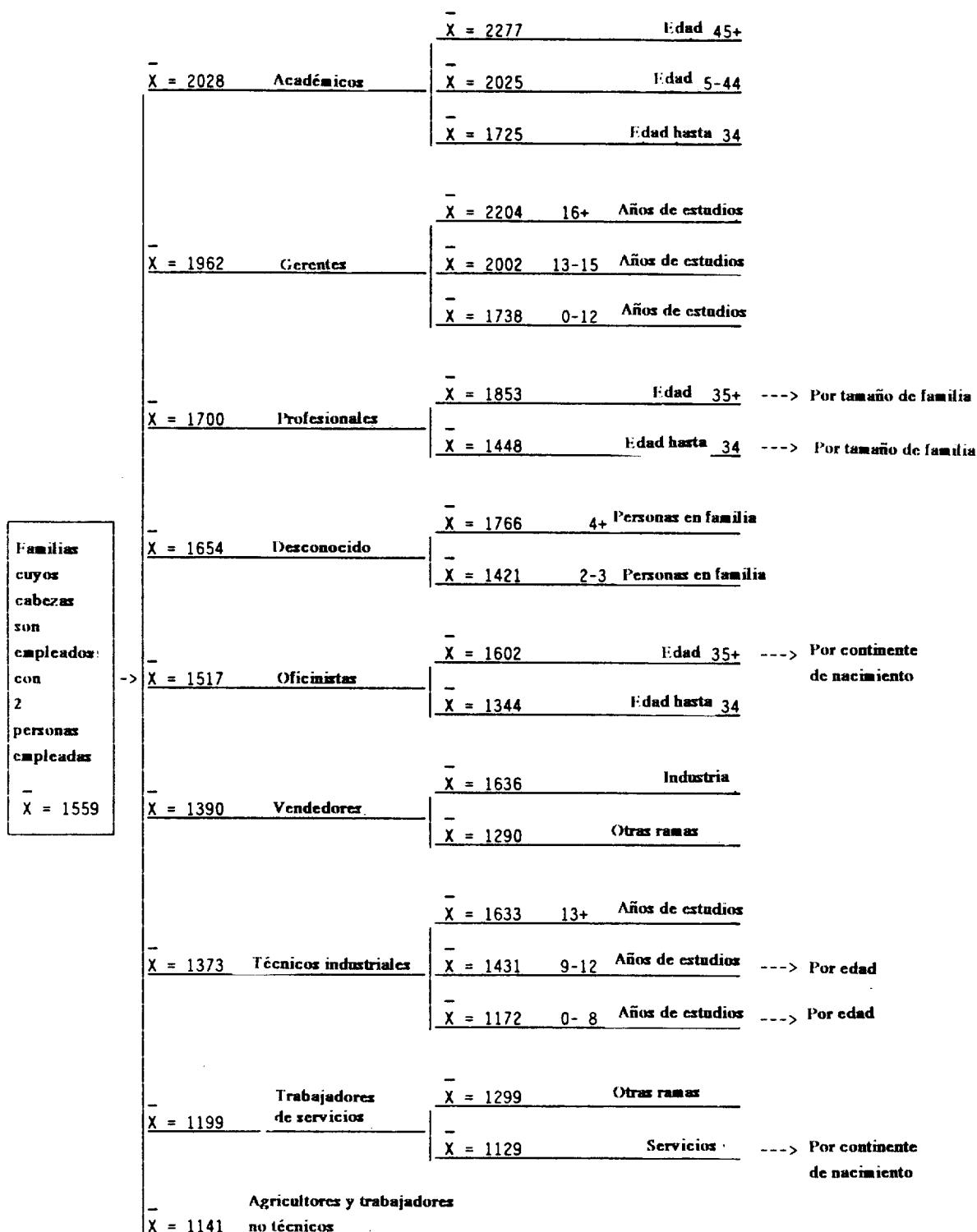
| Etapa | Total        |                           | Cabezas de familia     |                           |                       |                           |                          |                           |
|-------|--------------|---------------------------|------------------------|---------------------------|-----------------------|---------------------------|--------------------------|---------------------------|
|       |              |                           | Empleados<br>w = 63.4% |                           | Autónomos<br>w = 7.7% |                           | Sin trabajo<br>w = 28.9% |                           |
|       | Nº de Grupos | Cumulativa R <sup>2</sup> | Nº de Grupos           | Cumulativa R <sup>2</sup> | Nº de Grupos          | Cumulativa R <sup>2</sup> | Nº de Grupos             | Cumulativa R <sup>2</sup> |
| 1     | 3            | 0.277                     | 1                      | -                         | 1                     | -                         | 1                        | -                         |
| 2     | 9            | 0.452                     | 3                      | 0.279                     | 3                     | 0.132                     | 3                        | 0.095                     |
| 3     | 40           | 0.527                     | 23                     | 0.395                     | 9                     | 0.180                     | 8                        | 0.165                     |
| 4     | 84           | 0.560                     | 57                     | 0.447                     | 13                    | 0.199                     | 14                       | 0.192                     |
| 5     | 100          | 0.569                     | 70                     | 0.463                     | 14                    | 0.201                     | 16                       | 0.198                     |

Los valores de  $R^2$  obtenidos por el árbol eran: para total población 0.57; para familias cuyo cabeza es empleado 0.46; y para los otros dos subgrupos, sólo 0.20. Las familias cuyo cabeza es autónomo constan de menos de 8% de la población total y teniendo en cuenta las limitaciones impuestas por el tamaño de la muestra, no se podía usar una clasificación más detallada.

El grupo de familias cuyo cabeza no trabaja, aunque mayor, se clasificaba en sólo 16 categorías a pesar del hecho que este grupo consta de aprox. 30% de la población total. Se podría haber conseguido una mejor clasificación de ese grupo si se hubiera elegido variables clasificadoras más adecuadas. Un árbol revisado (con más variables clasificadoras) que se ha construido recientemente usando el software CART, logró un valor más alto de  $R^2$  (sobre 0.35) para ese mismo grupo.

Es difícil presentar aquí el árbol de clasificación completa con sus 100 subgrupos. Como un ejemplo, la Figura 7.1 presenta una de las ramas de familias cuyos cabezas son empleados y con 2 personas empleadas por familia. Este subgrupo primero se clasificaba por 9 grupos de ocupación. Luego se clasificaban por edad a los académicos, profesionales, y oficinistas, por número de años de estudios a los gerentes y técnicos industriales y por rama económica a los trabajadores de servicios y ventas, mientras que no se clasificaba más al grupo de agricultores y trabajadores industriales.

**Figura 7.1**  
**Arbol de Clasificación para Ingreso Neto por Familia ( $\bar{X}$ )**  
**en la Encuesta sobre Ingresos de Israel**  
**Ejemplo de una Sub-rama**



b) Construir un Arbol para un Conjunto de Variables de Estimación en las Encuestas de Ingreso Familiar

Desde la encuesta de 1979-80, se han usado los árboles de clasificación para la post-estratificación con respecto a un conjunto de 11 variables de estimación: Gasto total de la familia y los 10 principales grupos de gasto. Las variables potenciales clasificadoras eran: tamaño de familia, tamaño de la localidad, y las características del cabeza de familia - edad, continente de nacimiento, último colegio, ocupación y nivel de empleo (se usaba el período de investigación en una etapa separada de la ponderación). En encuestas anteriores, se determinaban las ponderaciones por una post-estratificación basada en una clasificación cruzada, por tamaño de familia, tamaño de localidad y período de investigación.

En cada etapa, los valores de  $R^2$  se calculaban para cada una de las 11 variables de estimación según cada una de las variables clasificadoras todavía no empleadas en esa etapa. Se examinaban 5 criterios combinados diferentes en cada etapa: los tres mencionados en la sección anterior y también dos criterios ponderados. Las ponderaciones eran relativas a las varianzas muestrales del gasto medio estimado por familia en la población total, aunque hubiera sido mejor determinar estas ponderaciones relativas a la varianza muestral residual en el subgrupo. En algunas etapas había una variable clasificadora que era la mejor según todos los criterios, mientras que en otras etapas, había que juzgar qué criterios a usar.

Para ilustrar el proceso, se demuestra la determinación de las variables clasificadoras para sólo algunos grupos en las primeras etapas. En la primera etapa todas las variables clasificadoras participaron y la Tabla 7.2 presenta la contribución de cada variable clasificadora.

**Tabla 7.2**  
**Determinación de la Primera Variable Clasificadora**  
**para la Población Total**  
**EGF de Israel**

| Gasto medio<br>por familia<br>(Variable estimada) | Pondera-<br>ción por<br>criterio | Variable clasificadora  |                           |                   |                |                       |      |                                    |
|---|----------------------------------|-------------------------|---------------------------|-------------------|----------------|-----------------------|------|------------------------------------|
|   |                                  | Tamaño<br>de<br>Familia | Tamaño<br>de<br>Localidad | Último<br>colegio | Ocupa-<br>ción | Nivel<br>de<br>Empleo | Edad | Conti-<br>nente<br>naci-<br>miento |
| Gasto total                                       | 1                                | 14.3                    | 2.1                       | 12.1              | <u>18.4</u>    | 9.5                   | 13.4 | 4.2                                |
| Comida(excl. Verd. y fruta)                       | 1                                | <u>41.1</u>             | 1.1                       | 1.3               | 9.7            | 7.7                   | 16.7 | 6.0                                |
| Veduras y fruta                                   | 1                                | <u>32.5</u>             | 0.6                       | 0.5               | 4.7            | 4.1                   | 11.8 | 4.7                                |
| Mantenimiento vivienda                            | 3                                | 10.8                    | 1.7                       | <u>11.6</u>       | 10.9           | 4.1                   | 7.5  | 3.8                                |
| Transporte y comunicación                         | 8                                | 4.7                     | 2.0                       | <u>16.2</u>       | 15.5           | 5.6                   | 7.0  | 3.8                                |
| Educación, cultura y ocio                         | 5                                | 9.0                     | 0.6                       | 9.1               | <u>13.8</u>    | 8.7                   | 11.4 | 2.8                                |
| Vivienda  | 2                                | 1.3                     | 4.0                       | 5.0               | <u>6.2</u>     | 1.2                   | 2.3  | 2.8                                |
| Ropa y calzado                                    | 6                                | <u>10.4</u>             | 0.4                       | 1.7               | 4.7            | 3.8                   | 6.3  | 1.4                                |
| Muebles y equipo hogar                            | 10                               | 3.6                     | <u>8.6</u>                | 3.2               | 5.5            | 4.5                   | 7.7  | 1.8                                |
| Varios bienes y servicios                         | 4                                | 4.3                     | 0.5                       | 4.5               | <u>7.1</u>     | 4.3                   | 6.0  | 1.8                                |
| Salud   | 5                                | 3.1                     | 1.7                       | <u>4.5</u>        | 4.1            | 2.0                   | 1.6  | 0.7                                |

**Tabla 7.3**  
**Determinación de la Segunda Variable Clasificadora**  
**En Familias de Tamaño 1**  
**EGF de Israel**

| Gasto medio<br>por familia<br>(Variable estimada) | Pondera-<br>ción por<br>criterio | Variable clasificadora  |                           |                   |                |                       |             |                                    |
|---|----------------------------------|-------------------------|---------------------------|-------------------|----------------|-----------------------|-------------|------------------------------------|
|   |                                  | Tamaño<br>de<br>Familia | Tamaño<br>de<br>Localidad | Último<br>colegio | Ocupa-<br>ción | Nivel<br>de<br>Empleo | Edad        | Conti-<br>nente<br>naci-<br>miento |
| Gasto total                                       | 1                                | -                       | 2.7                       | <u>16.8</u>       | 13.3           | 8.8                   | 8.3         | 0.4                                |
| Comida(excl. Verd.y Fruta)                        | 1                                | -                       | <u>6.8</u>                | 2.2               | 4.8            | 5.3                   | 4.7         | 5.7                                |
| Veduras y fruta                                   | 1                                | -                       | 4.7                       | 1.1               | 2.0            | 2.9                   | 7.2         | <u>14.0</u>                        |
| Mantenimiento vivienda                            | 3                                | -                       | 1.1                       | <u>4.7</u>        | 3.9            | 0.2                   | 3.5         | 2.3                                |
| Transporte y comunicación                         | 8                                | -                       | 0.5                       | <u>15.2</u>       | 11.4           | 8.4                   | 13.1        | 0.5                                |
| Educación, cultura y ocio                         | 5                                | -                       | 5.0                       | <u>10.2</u>       | <u>10.3</u>    | 9.5                   | 10.0        | 0.2                                |
| Vivienda  | 2                                | -                       | 8.7                       | <u>9.7</u>        | 4.7            | 1.3                   | 1.0         | 4.2                                |
| Ropa y calzado                                    | 6                                | -                       | 0.8                       | <u>6.7</u>        | 3.7            | 1.7                   | <u>7.8</u>  | 0.5                                |
| Muebles y equipo hogar                            | 10                               | -                       | 5.7                       | 9.5               | 11.3           | 8.3                   | <u>11.5</u> | 0.5                                |
| Varios bienes y servicios                         | 4                                | -                       | 0.5                       | 2.2               | <u>2.3</u>     | 1.5                   | 1.5         | 0.1                                |
| Salud   | 5                                | -                       | <u>3.9</u>                | 2.1               | 3.0            | 0.9                   | 1.2         | 1.4                                |
| a) Número de Max( $R^2$ )                         |                                  | -                       | 2                         | <u>4</u>          | 2              | 0                     | 2           | 1                                  |
| b) Media simple de $R^2$                          |                                  | -                       | 3.7                       | <u>7.3</u>        | 6.4            | 4.4                   | 6.3         | 2.7                                |
| c) Media ponderada de $R^2$                       |                                  | -                       | 3.2                       | <u>8.3</u>        | 7.5            | 5.2                   | 7.8         | 1.2                                |
| d) Media simple de $R^2/\text{Max}(R^2)$          |                                  | -                       | 0.45                      | <u>0.78</u>       | 0.72           | 0.45                  | 0.67        | 0.30                               |
| e) Media ponderada de $R^2/\text{Max}(R^2)$       |                                  | -                       | 0.39                      | <u>0.85</u>       | 0.80           | 0.50                  | 0.78        | 0.16                               |

En las etapas siguientes, se realizó más clasificación, usando el mismo procedimiento, aunque con menos variables clasificadoras. Por ejemplo, en la tercera etapa, para las familias de tamaño 1 cuyos cabezas tenían el nivel de educación más bajo, la mejor variable clasificadora para todos los criterios era continente de nacimiento, mientras que, para aquellos cabezas de familia con el nivel de educación más alto, la mejor variable clasificadora era edad.

El resultante  $R^2$  acumulado en cada etapa para cada una de las 11 variables de estimación se presentan en la Tabla 7.4.

**Tabla 7.4**  
**R<sup>2</sup> Acumulado por Etapas de Clasificación**  
**EGF de Israel**

| Gasto medio<br>por familia<br>(Variable estimada) | R <sup>2</sup> Acumulado |         |         |         |
|---|--------------------------|---------|---------|---------|
|   | Etapa 1                  | Etapa 2 | Etapa 3 | Etapa 4 |
| Gasto total                                       | 0.143                    | 0.282   | 0.391   | 0.409   |
| Comida(excl. Verd.y fruta)                        | 0.411                    | 0.429   | 0.479   | 0.491   |
| Verduras y fruta                                  | 0.325                    | 0.345   | 0.394   | 0.408   |
| Mantenimiento vivienda                            | 0.108                    | 0.225   | 0.288   | 0.308   |
| Transporte y comunicación                         | 0.047                    | 0.199   | 0.263   | 0.289   |
| Educación, cultura y ocio                         | 0.090                    | 0.186   | 0.257   | 0.270   |
| Vivienda  | 0.013                    | 0.072   | 0.190   | 0.199   |
| Ropa y calzado                                    | 0.104                    | 0.128   | 0.178   | 0.193   |
| Muebles y equipo hogar                            | 0.036                    | 0.080   | 0.167   | 0.190   |
| Varios bienes y servicios                         | 0.043                    | 0.089   | 0.138   | 0.155   |
| Salud   | 0.031                    | 0.061   | 0.117   | 0.148   |
| Número de sub-grupos                              | 6                        | 26      | 80      | 117     |

El R<sup>2</sup> obtenido por esta clasificación es relativamente alto para las estimaciones de la media del Gasto total, para Comida y para Verduras y fruta (entre 40% y 50%), pero es más bajo para todas las demás estimaciones (entre 15% a 30%).

Las primeras tres etapas contribuyen la mayor parte de la homogeneidad y la contribución de la última etapa es mínimo a pesar de que el número de grupos incrementó notablemente. La contribución marginal de cada etapa no es similar para los varios grupos de gasto. Para Comida, Verduras y fruta la primera etapa contribuye la mayoría de los R<sup>2</sup>, mientras que en otros grupos, la contribución marginal de las últimas etapas es relativamente alta. Para la EGF de 1992-93, se planifica construir un árbol revisado con un número mayor de variables clasificadoras potenciales, para poder obtener una mejor clasificación. También, se estudiará la eficiencia de usar software CART con una variable sintética.

### 7.5 Eficacia de los Arboles de Clasificación

Normalmente, se construye un árbol usando los datos de cierta encuesta y se aplica para la post-estratificación en una encuesta consiguiente. Si la muestra de la encuesta posterior hubiera sido estratificada inicialmente por las mismas post-estratas que el árbol, la reducción esperada en la varianza muestral hubiera sido por el factor  $1-R^2$ , con tal que no hubiera ocurrido cambios en la población relativos a la variable de estimación. Esto también sería el caso con una muestra simple de post-estratificación por las mismas post-estratas si los tamaños de las muestras son suficientemente grandes, si los ratios de respuesta son similares en las post-estratas y si se saben las estimaciones externas. Sin embargo, en la práctica no se cumplen todas estas condiciones, así que, hay que examinar la eficiencia empírica real de la post-estratificación por arboles de clasificación.

a) Para la Encuesta sobre Ingresos de 1988 se estudiaban los resultados empíricos comparando tres métodos:

- A: Auto-ponderación (un grupo).
- B: Post-estratificación por clasificación cruzada (100 grupos) - método antiguo.
- C: Post-estratificación por árbol de clasificación (100 grupos) - método nuevo.

La Tabla 7.5 presenta una comparación de los tres métodos para la varianza muestral relativa de las estimaciones del ingreso neto por familia.

**Tabla 7.5**  
**Ratios de las Varianzas Muestrales Relativas**  
**entre Métodos de Ponderación**  
**Estimaciones de Ingreso Neto por Familia**  
**Encuesta sobre Ingresos, 1988 de Israel**

| Sub-grupo                                    | Ratio entre Varianzas Muestrales Relativas |      |      |                          |      |       |                   |      |
|--|--|------|------|--------------------------|------|-------|-------------------|------|
|  | Con ponderación conocido                   |      |      | Con ponderación estimada |      |       | $\epsilon$ efecto |      |
|  | B/A  | C/A  | C/B  | B'/A                     | C'/A | C'/B' | B'/B              | C'/C |
| Total  | 0.85                                       | 0.55 | 0.65 | 0.92                     | 0.69 | 0.75  | 1.08              | 1.25 |
| <b>Nivel de empleo de cabeza de familia</b>  |  |      |      |                          |      |       |                   |      |
| Empleado                                     | 0.92                                       | 0.62 | 0.67 | 0.96                     | 0.75 | 0.79  | 1.04              | 1.22 |
| Autónomo                                     | 1.05                                       | 0.92 | 0.87 | 1.05                     | 0.93 | 0.89  | 1.00              | 1.02 |
| Sin trabajo                                  | 1.34                                       | 0.80 | 0.60 | 1.35                     | 0.82 | 0.61  | 1.01              | 1.03 |
| <b>Años de estudios de cabeza de familia</b> |  |      |      |                          |      |       |                   |      |
| 0- 8   | 0.95                                       | 0.79 | 0.83 | 0.98                     | 0.94 | 0.95  | 1.04              | 1.19 |
| 9-12   | 1.04                                       | 0.68 | 0.65 | 1.06                     | 0.74 | 0.70  | 1.02              | 1.10 |
| 13+  | 0.90                                       | 0.62 | 0.69 | 0.93                     | 0.74 | 0.80  | 1.04              | 1.20 |
| <b>Tamaño de familia</b>                     |  |      |      |                          |      |       |                   |      |
| 1  | 1.09                                       | 1.02 | 0.94 | 1.10                     | 1.08 | 0.98  | 1.01              | 1.06 |
| 2  | 0.99                                       | 0.94 | 0.95 | 1.01                     | 1.00 | 0.99  | 1.02              | 1.06 |
| 3  | 1.02                                       | 0.96 | 0.93 | 1.03                     | 0.99 | 0.96  | 1.01              | 1.04 |
| 4  | 1.09                                       | 0.71 | 0.65 | 1.09                     | 0.74 | 0.67  | 1.00              | 1.03 |
| 5  | 1.02                                       | 0.89 | 0.88 | 1.02                     | 0.95 | 0.93  | 1.01              | 1.07 |
| 6+   | 1.03                                       | 0.97 | 0.95 | 1.03                     | 1.02 | 0.99  | 1.01              | 1.05 |

Se puede ver que si se saben los  $W_h$  (métodos B y C), para la población entera, la varianza muestral relativa en el método antiguo se reduce a 0.85 comparado con 0.55 en el método nuevo. Así que, por el método nuevo hay un beneficio de 0.65 comparado con el método antiguo. La reducción de la varianza muestral con el método nuevo (C) es menor que el valor inicial 0.43 de  $1-R^2$ , principalmente a causa de los efectos de  $\sigma$  y D. Sin embargo, dado que  $W_h$  son estimaciones (métodos B' y C') de la Encuesta de la Fuerza Laboral, un factor adicional,  $\epsilon$ , contribuye más al incremento de la varianza en el método nuevo que en el antiguo, porque  $S^2_b$  es mayor. Para la población total,  $\epsilon$  en el método nuevo, incrementa la varianza muestral relativa por 1.25 comparado con 1.08 en el método antiguo. Por lo tanto, en la práctica, el beneficio real de eficiencia obtenido con el método nuevo es sólo 0.75.

Para las estimaciones dentro de los subgrupos, el método antiguo sólo reduce ligeramente la varianza muestral y a veces, incluso puede incrementarla como resulta de los ratios diferenciales de la no respuesta. Con el método nuevo, se reducen las varianzas muestrales pero, en general, por una menor cantidad que en la población total. Como era de esperar, con los dos métodos  $\epsilon$  es menor, cuanto más pequeños sean los grupos, y a la inversa, mayor cuanto mayor sea el  $S^2_b$  dentro del grupo.

Las mismas estimaciones obtenidas de los tres métodos también se comparaban tanto para la variable de estimación utilizada para determinar la clasificación (ingreso neto por familia) como las otras dos variables (ingreso bruto por familia e ingreso neto por persona), y se presentan en la Tabla 7.6.

**Tabla 7.6**  
**Comparación entre Estimaciones y Ratio de MSE**  
**por Métodos de Ponderación (Porcentajes)**  
**Encuesta sobre Ingreso 1988 de Israel**

| Sub-grupo                                    | Comparación entre Estimaciones |          |                              |          |                             |          | Ratio de MSE de<br>Ingreso neto<br>por familia |      |      |
|--|--------------------------------|----------|------------------------------|----------|-----------------------------|----------|--|------|------|
|  | Ingreso neto<br>por persona    |          | Ingreso bruto<br>por familia |          | Ingreso neto<br>por familia |          |  |      |      |
|  | (C'/A)-1                       | (B'/A)-1 | (C'/A)-1                     | (B'/A)-1 | (C'/A)-1                    | (B'/A)-1 | B'/A   | C'/A | C'/B |
| Total  | 2.0                            | -0.2     | 6.4                          | 3.6      | 5.0                         | 3.0      | 0.18   | 0.03 | 0.14 |
| <b>Nivel de empleo de cabeza de familia</b>  |                                |          |                              |          |                             |          |  |      |      |
| Empleado                                     | 1.9                            | -0.1     | 2.2                          | 1.0      | 1.9                         | 0.8      | 0.46   | 0.14 | 0.31 |
| Autónomo                                     | 0.2                            | -1.0     | 1.6                          | 0.7      | 1.7                         | 0.6      | 0.88   | 0.69 | 0.78 |
| Sin trabajo                                  | -2.9                           | -3.5     | 0.1                          | 1.0      | -0.2                        | 0.9      | 1.50   | 0.82 | 0.55 |
| <b>Años de estudios de cabeza de familia</b> |                                |          |                              |          |                             |          |  |      |      |
| 0- 8   | 1.6                            | -1.1     | 7.1                          | 4.4      | 5.7                         | 3.7      | 0.20   | 0.09 | 0.47 |
| 9-12   | 0.4                            | -0.3     | 4.6                          | 2.5      | 3.4                         | 2.0      | 0.31   | 0.12 | 0.38 |
| 13+  | 2.6                            | -0.9     | 5.0                          | 1.8      | 3.9                         | 1.4      | 0.46   | 0.10 | 0.21 |
| <b>Tamaño de familia</b>                     |                                |          |                              |          |                             |          |  |      |      |
| 1  | 3.5                            | 2.7      | 5.4                          | 3.7      | 3.5                         | 2.7      | 0.43   | 0.39 | 0.90 |
| 2  | 4.2                            | 2.2      | 5.9                          | 3.1      | 4.2                         | 2.2      | 0.34   | 0.15 | 0.44 |
| 3  | 3.9                            | 0.9      | 5.1                          | 1.2      | 3.9                         | 0.9      | 0.68   | 0.23 | 0.34 |
| 4  | 2.2                            | 0.7      | 3.1                          | 0.9      | 2.2                         | 0.7      | 0.78   | 0.37 | 0.48 |
| 5  | 1.3                            | 0.2      | 2.0                          | 0.4      | 1.3                         | 0.2      | 0.93   | 0.65 | 0.70 |
| 6+   | 2.9                            | 0.0      | 3.6                          | 0.4      | 2.8                         | 0.2      | 0.91   | 0.38 | 0.42 |

Se puede ver que, en general, las estimaciones del nuevo método ( $C'$ ) son más altas que del método antiguo ( $B'$ ) y normalmente con ambos métodos son más altas que las estimaciones obtenidas sin usar la ponderación (A). La excepción a esto es en las familias cuyos cabezas no trabajan, donde todo lo contrario es verdad.

Si se puede asumir que las estimaciones en el método nuevo tienen menos sesgo, entonces la reducción en sesgo por el nuevo método es considerable comparado con el método antiguo. Esto es el caso tanto para la población total como para la mayoría de los subgrupos. Sin embargo, no está claro si el sesgo en cada uno de los grupos se corrige hasta el mismo punto, dado que la eficiencia de la clasificación varía de grupo en grupo, un tema que necesita más examen.

El MSE (Mean Square Error) relativo de las estimaciones del ingreso neto por familia por los tres métodos, bajo la asunción que por el método C las estimaciones no están sesgadas, también se presentan en la Tabla 7.6. dado que los sesgos relativos son mayores que los errores muestrales relativos, el MSE en el método nuevo es mucho menor que en el método antiguo, y, por supuesto, que el de la estimación simple. Por lo tanto, el nuevo método es muy eficiente para la población total pero menos eficiente para los subgrupos pequeños (como las personas autónomas).

b) Para la Encuesta de Gasto Familiar 1979-80 se calculaba la diferencia entre las dos estimaciones obtenidas por el procedimiento antiguo de ponderación y el nuevo, relativo a la estimación por el nuevo método. La Tabla 7.8 presenta estos ratios para estimaciones del gasto medio por familia para la población total y para el tamaño de familia. Estos ratios pueden indicar una reducción de sesgos (debido a la no respuesta) por el nuevo método porque se espera que sea menos sesgado.

**Tabla 7.7**  
**Reducción de Sesgo por la Nueva Ponderación**  
**Relativa a la Antigua Ponderación por Tamaño de Familia**  
**EGF de 1979-80 de Israel**

| Grupo de gasto                                 | Tamaño de familia |       |       |       |       |       |       |
|--|-------------------|-------|-------|-------|-------|-------|-------|
|  | Total             | 1     | 2     | 3     | 4     | 5     | 6+    |
| <b>Estimaciones de Gasto medio por Familia</b> |                   |       |       |       |       |       |       |
| Gasto total                                    | + 4.7             | + 9.2 | + 7.9 | + 2.2 | + 4.1 | + 3.7 | + 0.7 |
| Veduras y fruta                                | + 1.0             | - 0.2 | + 2.4 | + 0.4 | + 2.1 | - 0.4 | - 0.6 |
| Comida(excl. Verd.y fruta)                     | + 1.6             | - 0.6 | + 1.0 | + 0.6 | + 0.5 | - 0.1 | - 1.3 |
| Muebles y equipo hogar                         | + 2.2             | - 3.2 | + 8.8 | - 1.3 | + 2.9 | + 2.5 | - 4.0 |
| Ropa y calzado                                 | + 3.6             | + 1.8 | +10.8 | + 0.1 | + 4.0 | + 2.0 | + 1.0 |
| Vivienda                                       | + 4.9             | + 8.1 | + 8.1 | - 1.0 | + 4.9 | + 3.5 | + 5.2 |
| Mantenimiento vivienda                         | + 5.0             | + 2.5 | + 7.4 | + 3.9 | + 5.4 | + 6.2 | + 1.7 |
| Salud  | + 5.4             | +12.4 | + 6.1 | + 7.5 | + 3.2 | + 2.5 | + 1.7 |
| Varios bienes y servicios                      | + 6.6             | +20.4 | +11.8 | + 1.0 | + 5.1 | + 1.5 | + 3.7 |
| Educación, cultura y ocio                      | + 8.3             | +18.1 | +14.7 | + 8.3 | + 4.3 | + 6.8 | + 1.2 |
| Transporte y comunicación                      | + 9.3             | +28.9 | +11.3 | + 5.8 | + 7.5 | + 9.7 | - 1.1 |
| <b>Estimaciones de Ponderaciones de Gasto</b>  |                   |       |       |       |       |       |       |
| Gasto total                                    | 0.0               | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   | 0.0   |
| Veduras y fruta                                | - 3.5             | - 8.6 | - 5.1 | - 1.8 | - 1.9 | - 4.0 | - 1.3 |
| Comida(excl. Verd.y fruta)                     | - 3.0             | - 9.0 | - 6.4 | - 1.6 | - 3.5 | - 3.7 | - 2.0 |
| Muebles y equipo hogar                         | - 2.4             | -11.4 | + 0.8 | - 3.4 | - 1.2 | - 1.2 | - 4.7 |
| Ropa y calzado                                 | - 1.1             | - 6.8 | + 2.7 | - 2.1 | - 0.1 | - 1.6 | + 0.3 |
| Vivienda                                       | + 0.2             | - 1.0 | + 0.2 | - 3.1 | + 0.8 | - 0.2 | + 4.5 |
| Mantenimiento vivienda                         | + 0.3             | - 6.1 | - 0.5 | + 1.7 | + 1.2 | + 2.4 | + 1.0 |
| Salud  | + 0.7             | + 2.9 | - 1.7 | + 5.2 | - 0.9 | - 1.2 | + 1.0 |
| Varios bienes y servicios                      | + 1.8             | +10.3 | + 3.6 | - 1.2 | + 1.0 | - 2.1 | + 3.0 |
| Educación, cultura y ocio                      | + 3.4             | + 8.2 | + 6.3 | + 6.0 | + 0.2 | + 3.0 | + 0.5 |
| Transporte y comunicación                      | + 4.4             | +18.0 | + 3.2 | + 3.5 | + 3.3 | + 5.8 | - 1.8 |

Se puede ver que con el método antiguo la estimación del gasto total era 4.7% más alto que con el método nuevo, pero los sesgos difieren para los diferentes tamaños de familia. Por ejemplo, para tamaño 1, el sesgo relativo es de +9.2% y para tamaño 3 es sólo +2.2%. Para los diferentes grupos de gasto, se obtienen sesgos relativos diferentes, siendo éstos particularmente altos para Transporte y Comunicación (+9.3%) y para Educación y Cultura (+8.3%), y relativamente bajo para Comida (+1.6%) y para Verduras y fruta (+1.0%).

Así que, las ponderaciones estimadas para el Indice de Precios al Consumo también están sesgadas, como se demuestra en la Tabla 7.7. Para la población total, se puede ver claramente que la ponderación del gasto en Comida está sesgada a menos (3.0%) como la ponderación de gasto en Educación y Cultura; Transporte y Comunicación están sesgadas a más (3.4% y 4.4% respectivamente).

La reducción de los sesgos por el uso del árbol de clasificación para la población total está influenciada por el árbol entero, incluida, por supuesto, la primera clasificación por tamaño de familia. Un examen de los sesgos de las ponderaciones en cada tamaño de familia revela los efectos de usar el árbol de clasificación en cada uno de los grupos por tamaño de familia. Se encontraban sesgos relativos especialmente altos de estas ponderaciones para familias del tamaño 1, donde para Comida la estimación tiene un sesgo hacia abajo por aprox. 9.0%; Muebles y Equipo de casa hacia abajo por 11%, y Ropa y Calzado hacia abajo por 7%. La ponderación de gasto para Transporte y Comunicación está sesgada hacia arriba por aprox. 18% y para Educación y Cultura por 8%.

En resumen, estos resultados empíricos (y resultados adicionales no presentados aquí), muestran sin duda que un árbol da una mejor clasificación en grupos homogéneos que una clasificación cruzada y así se incrementa la eficiencia de la post-estratificación. También es razonable decir que al usar el árbol de clasificación, se logra una reducción del sesgo debido a la no respuesta. Sin embargo, no está claro si la corrección de los sesgos es la misma en cada uno de los subgrupos y aún así no está claro si las estimaciones de las diferencias entre los grupos tenga menos sesgo. Este es un problema común a la mayoría de los métodos de ponderación.

## 7.6 Ponderaciones Externas de las Familias

Para las personas, generalmente se disponen de estimaciones externas fiables y actuales para la clasificación cruzada por características básicas, como por ejemplo, sexo, edad, religión o grupo étnico, tipo o tamaño de localidad, etc. Sin embargo, normalmente no se disponen de datos actuales de las características de la familia. La Encuesta de la Fuerza Laboral actual (EFL), o una encuesta similar de familias, a menudo es la única fuente de estas estimaciones. Las estimaciones de la EFL se pueden usar para ponderar otras encuestas de familia que se basan en muestras más pequeñas que las de la EFL y que a menudo tienen un ratio de la no respuesta considerablemente más alto.

Las EFL en muchos países se llevan a cabo como encuestas de panel, para que cada familia participe en la encuesta durante varias rondas. Este también es el caso de Israel, donde la EFL se hace continuamente, basándose en un panel de trimestres, y cada vivienda se investiga durante dos trimestres consecutivos y, después de un descanso de dos trimestres, dos trimestres adicionales. El procedimiento de ponderación de la EFL para individuos se basa principalmente en las estimaciones de ratio para asegurar la consistencia con estimaciones actuales independientes de la distribución de la población por grupos de sexo-edad-religión-región. Por lo tanto, en general, las diferentes ponderaciones se asignan a diferentes miembros de la misma familia.

Las estimaciones del número de familias y sus características se obtienen en Israel actualmente al asignar a las familias la ponderación del cabeza de familia como se determinaba al ponderar los individuos. Hay una clara evidencia que las estimaciones de familias obtenidas por este tipo de ponderación están sesgadas, por ejemplo, el porcentaje estimado de las familias de 1 persona es severamente sesgado hacia abajo. Ni el uso de ponderaciones de otras personas "representativas", ni el uso de varias combinaciones de las ponderaciones individuales de miembros de las familias era satisfactorio.

Un abordamiento distinto a este problema, no basado en las ponderaciones individuales, se ha empleado en Israel para las EGF y para algunas otras encuestas desde 1980, y ahora se adopta para usar como base actual. Las etapas del procedimiento se describen a continuación:

- (a) Se realizan ajustes preliminares de las ponderaciones iniciales para corregir ciertos tipos de falta de cobertura en la muestra de la encuesta (como nuevas construcciones y unidades de viviendas temporales de nuevos inmigrantes).
- (b) Se hacen imputaciones para las personas de no respuesta "temporal", es decir, la no respuesta en una o más rondas, que contestan al menos una vez en las cuatro rondas. Los datos conocidos, para una selección de variables adecuadas para una imputación de este tipo, del trimestre más cercano se imputan al trimestre que falta, después de unos ajustes, si hacen falta, para la diferencia en el tiempo entre los trimestres.
- (c) Se hacen ajustes para compensar por la no respuesta "permanente", es decir, las familias que no contestaron en ninguna ronda. Hay varias alternativas: asignar un factor de ajuste uniforme a todos los encuestados o sólo a los de la no respuesta "temporal"; o asignar diferentes factores a los diferentes grupos de la no respuesta "temporal" (según las razones de la no respuesta) para compensar por los correspondientes grupos de no respuesta "permanente".
- (d) Se aplica un tratamiento especial para los paneles todavía no investigados en las cuatro rondas, especialmente para el panel investigado sólo una vez.
- (e) Controlar la distribución externa de las personas por grupos de sexo-edad-religión-región. (ver la sección 7.7).

Incluso con este procedimiento, las estimaciones para las familias todavía pueden estar algo sesgadas. Un método alternativo es el de usar los datos del último censo de la población en conjunto con las estimaciones de los cambios con el tiempo de las EFL. Dado que es razonable creer que las estimaciones mejoradas de las EFL estén menos sesgadas que otras estimaciones, al estimar los cambios con el tiempo, se pueden obtener mejores estimaciones externas para la post-estratificación.

### 7.7 Control de la Distribución de Personas con una Ponderación de Familia

Los procedimientos basados en la post-estratificación (o sus variantes), que controlan la distribución de las familias, no necesariamente controlan la distribución auxiliar de los individuos por sus características claves, por ejemplo, grupos de sexo-edad. Las estimaciones de la encuesta relacionadas con personas pueden no coincidir con los controles incluso cuando las estimaciones externas de las familias, que se usan para ponderar las familias, no tienen sesgo. Además de un incremento en la varianza, también puede resultar un sesgo tanto en las estimaciones de la encuesta relacionadas con las familias (dado que se espera una correlación entre el gasto en varios grupos de artículos y la composición de la familia según las características de sus miembros) como en las estimaciones de la encuesta relacionadas con grupos específicos de personas.

Por lo tanto, hay que considerar un ajuste para asegurar compatibilidad con la distribución auxiliar de las personas como una etapa adicional en el procedimiento de la ponderación de las familias, para evitar posibles sesgos en las estimaciones de la encuesta de esta fuente.

Se han empleado ajustes de este tipo en las EGF de 1979-80 y 1986-87 de Israel (además, en otras encuestas familiares). Se hacían después de determinar las ponderaciones de las familias por un procedimiento multietapa, es decir, post-estratificación (descrita en las secciones anteriores), en conjunto con un procedimiento de "rastrillado" para otros ajustes (por ejemplo, para corregir la desigual distribución de la muestra en el tiempo).

Normalmente, no se dispone de una distribución auxiliar fiable de las familias por los grupos de sexo-edad de sus miembros. Por lo tanto, el ajuste para coincidir con los controles de las personas se hace directamente por medio de un proceso iterativo. En principio, los perfiles de las familias se definen por la composición edad-sexo de sus miembros y se asigna una serie de factores de corrección a cada familia según su perfil, en varios ciclos hasta conseguir la convergencia, como se explica a continuación.

Deja que  $N_k$  sea el número de personas de la población en el grupo  $k$ ,  $k=1,2,\dots,G$ . Estos son los datos auxiliares y generalmente son fiables.

Deja que  $n_{jk}$  sea el número de personas en la familia  $i$  que pertenece al grupo  $k$ , y  $n_i$  el tamaño total de la familia  $i$ . Entonces, para

cada familia  $i$ , se define un perfil por el vector  $\{n_1, n_2, \dots, n_G\}_i$ , y son posibles  $F$  perfiles diferentes ( $f=1, 2, \dots, F$ ).

El número de personas en la muestra de todas las familias de perfil  $f$  que pertenecen al grupo  $k$  es  $n_{kf}$  y el número total de personas en todas las familias de perfil  $f$  es  $n_f$ .

Deja que  $N_k^*$  sea el tamaño estimado del grupo  $k$  obtenido por la ponderación de las familias antes de hacer el ajuste para personas.

Si  $N_k^* = N_k$  para todos los  $k$ , no requiere ningún ajuste, pero normalmente esto no es el caso. Entre varias alternativas, aquí se describen dos procedimientos de ajuste para la distribución de personas: el primer procedimiento no coincide completamente con los datos auxiliares de la familia, y el segundo, que es de tipo rastrillado, sí coincide. La elección entre estos dos procedimientos depende de la calidad de la distribución externa usada para la ponderación de las familias. Cuando se cree que estos datos sean fiables y sin sesgo, el segundo procedimiento es preferible.

En ambos procedimientos, a veces, hay que asignar un factor de ajuste preliminar a todos los perfiles para corregir una posible falta de cobertura o una sobrecobertura de las estimaciones externas de las familias.

#### Con el primer procedimiento:

En la primera ronda, se calculan  $A_k(1) = N_k/N_k^*$  para cada  $k$ . Entonces, para cada perfil  $f$ , se calcula el primer factor de ajuste.

$$\alpha_f(1) = \sum A_k(1)n_{kf}/n_f$$

y la ponderación de cada familia se multiplica por este factor y se obtienen nuevas estimaciones del número de personas en cada grupo  $k$ ,  $N_k^*(1)$ .

En cada ronda consiguiente,  $r$ , se obtiene un conjunto de factores de ajuste de la misma manera:

$$\alpha_f(r) = \sum A_k(r)n_{kf}/n_f,$$

donde  $A_k(r) = N_k/N_k^*(r-1)$  y  $N_k^*(r-1)$  es la estimación obtenida en la ronda anterior.

#### Con el segundo procedimiento:

En cada ronda, hay dos pasos: en el primer paso se calculan  $\alpha_f(r)$ , como en el primer procedimiento, y en el segundo paso, los factores de ajuste,  $\beta_h(r)$ , se dan a cada grupo de familias  $h$ ,

$$\beta_h(r) = L_h / L_h^*(r),$$

donde,  $L_h$  es el número externo de familias en el grupo  $h$ , y  $L_h^*(r)$  es la nueva estimación de este número obtenida por las ponderaciones ajustadas después del primer paso de la ronda  $r$ .

Ambos procedimientos se paran al lograr una coincidencia suficiente, o no se obtienen mejores coincidencias. Pero cuando la relvarianza de las ponderaciones llegue a ser demasiado grande, se puede parar el procedimiento en etapas anteriores, al coste de menos control de la distribución de personas, es decir, hay que hacer concesiones mutuas entre estos dos factores. También se puede parar el primer procedimiento para evitar más deterioro de la coincidencia en la distribución externa de las familias.

Se debería señalar que a veces no se puede obtener un ajuste satisfactorio, por ejemplo, con una muestra demasiado pequeña, o cuando los datos externos de las familias están significativamente sesgados, y entonces no siempre se puede lograr la convergencia para un incremento razonable de la varianza muestral.

Estos tipos de ajuste, se pueden aplicar para la muestra entera como un grupo, o para varios subgrupos de la población por separado, que puede ser preferible, con tal que se disponga de muestras bastante grandes para cada subgrupo.

Se debería mencionar que ambos procedimientos también se pueden aplicar usando "perfíles binarios" de familias en vez de "perfíles completos", como antes. Con esto, para cada familia, variables indicadoras,  $\delta_{ik}$ , se definen para cada grupo k, donde  $\delta_{ik}=1$  si  $n_{ik} \geq 1$ , y si no,  $\delta_{ik}=0$ . El vector  $\{\delta_1, \delta_2, \dots, G\}_i$  define el "perfil binario" de la familia i y son posibles B diferentes "perfíles binarios" ( $b=1, 2, \dots, B$ ) y  $B \leq F$ , dado que cada uno de los "perfíles binarios" puede constar de varios "perfíles completos". El procedimiento de ajuste funciona de una forma similar que con "perfíles completos", donde  $n_{kb}$  y  $n_b$  sustituyen  $n_{kf}$  y  $n_f$ . Aunque el uso de perfíles así agrupados no es tan bueno como el uso de perfíles completos, y se usa cuando la muestra es demasiado pequeña.

## CAPITULO 8. CODIFICACION

Para cada variable no cuantitativa que se investiga en una encuesta (o en un censo), tiene que definirse una variable categórica con un conjunto preespecificado de categorías exhaustivas y mutuamente exclusivas, además, se determina un código (normalmente numérico) para cada categoría. Sólo se pueden procesar los datos después de que toda la información verbal de cada variable no cuantitativa haya sido convertida a los códigos apropiados.

### 8.1 Técnicas de Codificación en General

Se presentan a continuación varias técnicas de codificación, que dependen del tipo de variable investigada y de la estructura de los formularios usados para la recogida de datos.

- (1) Técnicas de precodificación - Cuando se emplean preguntas de tipo cerrado para la recogida de datos, se imprimen en el cuestionario todas las categorías con todas las posibles respuestas junto con sus códigos. El código que corresponde a la respuesta elegida está marcado durante la recogida de los cuestionarios. Las preguntas precodificadas se emplean ampliamente, pero sólo son factibles para las variables que tienen un número limitado de categorías y para respuestas cortas sin perder claridad.

Otra técnica de precodificación es cuando, para cada categoría, se asigna cierto espacio en el cuestionario para registrar información adicional. El código de cada categoría es preimpreso e imputado en el ordenador junto con la otra información numérica, o, la localización de la información de la categoría en el cuestionario determina el código, y entonces no hace falta teclear el código.

- (2) Codificación manual - Los códigos de la información verbal recogida con preguntas de tipo abierto se determinan por los codificadores con la utilización de libros de códigos, es decir, una lista de posibles respuestas ordenada alfabéticamente con sus categorías y códigos, y una lista de códigos en orden numérico, con las posibles respuestas para cada código. Los códigos apropiados se registran en los formularios y más tarde se imputan en el ordenador junto con toda la demás información numérica.

Otra técnica manual es cuando los códigos se teclean directamente en el ordenador (en vez de escribirlos primero en los formularios y luego pasarlo al ordenador), entonces se ahorra una etapa en el proceso de la encuesta. Sin embargo, esta técnica es adecuada sólo para las variables con un número limitado de categorías. Otra versión de esta técnica es cuando aparece en la pantalla del ordenador un diccionario corto y el codificador elige el código apropiado con el cursor. Esto se hace cuando no es aconsejable utilizar preguntas precodificadas porque el número de códigos diferentes es demasiado grande, o porque la estructura del cuestionario no lo permite.

(3) Codificación automática - En la codificación automática, primero se almacena en el ordenador un "diccionario" que sustituye al libro de códigos de la codificación manual. Tiene que ser tan comprensivo como sea posible, incluyendo toda la información verbal posible que se espera que aparezca en los cuestionarios (a no ser que sea muy infrecuente) y se debe evitar las descripciones ambiguas. En segundo lugar, la información verbal recogida en la encuesta tiene que ser imputada en el ordenador y entonces, se busca cada descripción de artículo en el diccionario. Si encaja la descripción con la respuesta, se asigna el correspondiente código, si no, se refiere a un proceso suplementario, es decir, la codificación manual. El algoritmo de la búsqueda y las normas correspondientes se deben adaptar al tipo de variable codificada y a la estructura del lenguaje.

En general, hay dos tipos de emparejamiento: exacto e inexacto. Con el emparejamiento exacto, la descripción estandarizada del artículo tiene que ser idéntica a la descripción encontrada en el diccionario y sólo entonces se asigna el código asociado.

El emparejamiento inexacto permite que se asigne el código asociado si sólo hay una similitud parcial entre la descripción del artículo y la definición en el diccionario. Se puede especificar el grado de similitud utilizando varios conjuntos de normas de emparejamiento basados en probabilidades empíricas y en la estructura del lenguaje, por ejemplo, ignorando el orden de las palabras, asignando ponderaciones heurísticas a las palabras según el contenido de cada palabra, etc. (Appel y Hellerman, 1983). En general, los emparejamientos exactos se espera que produzcan menos errores de codificación que los emparejamientos inexactos, pero lleva un ratio más alto de descripciones sin código. Sin embargo, no siempre es el caso con las variables de una estructura simple, como bienes y servicios en las EGF, donde se puede lograr un alto ratio de emparejamientos exactos.

La codificación automática se puede realizar de dos formas principales:

- (a) Codificación asistida por ordenador - La información verbal de un artículo a la vez se imputa en el ordenador. Luego, todas las descripciones del diccionario parecidas a ese artículo aparecen en la pantalla junto con sus códigos. El codificador elige de las alternativas el código que le parece más apropiado y lo imputa en el ordenador, o tecleando el código o usando el cursor. A veces, el codificador tiene acceso a información adicional relacionada con el artículo codificado que se puede usar para obtener ayuda del ordenador. Una versión de esta técnica es cuando el codificador sólo imputa los primeros caracteres de una palabra (por ejemplo, las primeras tres letras) y entonces aparece en la pantalla un subconjunto de definiciones del diccionario para que eliga una. (Embrey, 1988).
- (b) Codificación totalmente automática - La información verbal de todos los formularios de cierto lote de códigos se

teclea consecutivamente en el ordenador. Luego, el fichero que incluye toda la información verbal de todos los artículos se empareja con el diccionario por medio de un algoritmo de búsqueda especial adaptado específicamente a esta operación. Entonces, se asignan códigos automáticamente a las descripciones emparejadas. Un fichero de entradas que no se empareja automáticamente pasa a la codificación manual, o primero, a la codificación asistida por ordenador y luego, si hace falta, a la codificación manual.

Un buen procedimiento de codificación automática es el que minimiza el ratio de codificación errónea, el número de entradas que requieren codificación manual y el tiempo de búsqueda del ordenador para cada entrada. Claramente, existe un conflicto entre estos objetivos, es decir, minimizar el ratio de la codificación errónea puede resultar un incremento en la cantidad de codificación manual y viceversa, y la reducción de tiempo de búsqueda, por ejemplo, reduciendo el número de artículos de referencia en el diccionario, puede resultar un incremento en el ratio de errores. Por lo tanto, hay que hacer concesiones mutuas entre estos diferentes objetivos para poder elegir el método más eficiente.

Un buen proceso de codificación automática también asegura la flexibilidad para incorporar información acumulada durante el período de la codificación. Así que, se requiere una continua evaluación del proceso para poner al día el diccionario y para mejorar el proceso.

Tanto con la codificación manual como con la automática, se necesitan comprobaciones de la codificación y de la operación del teclado de los códigos basándose en los procedimientos estándares de muestras de control de calidad. Sin embargo, al contrario que los sistemas de control que se usan para la fabricación industrial, está el problema adicional de determinar el código verdadero. Por lo tanto, se han desarrollado métodos especiales para controlar la operación de codificación.

Hay dos procedimientos principales de validación: una validación dependiente y una independiente. En la validación dependiente, un técnico inspecciona el código asignado por el codificador y decide si es correcto o no. En la validación independiente, el técnico asigna códigos sin haber visto los códigos iniciales. Luego se emparejan los códigos iniciales con los códigos controlados (en el ordenador) y un experto examina las discrepancias y decide cual es el código correcto, o incluso puede elegir otro código. El técnico dependiente tiende a estar influenciado por el código inicial y así el ratio de códigos incorrectos que no se detectan puede ser sustancial.

La validación independiente tiene un ratio mucho más alto de detección de códigos erróneos, pero es más cara y más difícil de implantar que la validación dependiente. Con cualquier tipo de validación se descubrirán más errores si se repite muchas veces, pero, porque los recursos son limitados generalmente sólo se hace una validación.

Cada una de las técnicas de codificación tiene sus ventajas e inconvenientes.

Las técnicas de precodificación evitan el proceso de codificación y por lo tanto, no tienen errores de codificación, pero tienen un uso limitado a ciertas variables y sólo se pueden utilizar con cierto tipo de estructura del formulario de la encuesta.

Normalmente, la codificación manual está sujeta a un alto ratio de codificación y tecleo erróneos, por lo tanto, hay que comprobarla a fondo con procedimientos de control de calidad. Además, la codificación manual es costosa y consume mucho tiempo, y el entrenamiento de un gran número de codificadores impone una carga a la organización.

La ventaja principal de la codificación automática está en su potencial a proporcionar una mejor calidad de codificación que con la codificación manual. Además, esto se puede obtener a un coste reducido y en menos tiempo. La codificación automática ahorra la mayoría del trabajo de oficina de la codificación manual, pero *no todo*, dado que la codificación automática normalmente necesita alguna codificación manual también, excepto para ciertas variables para las cuales se puede obtener una codificación automática total. Por otra parte, teclear las respuestas textuales que se requiere con la codificación automática es más caro que teclear sólo los códigos en el proceso manual. La codificación automática también reduce la carga a la organización dado que siempre se disponen de mecanógrafos que necesitan poco entrenamiento y su trabajo es más fácil de comprobar. La codificación automática no está libre de errores y no se puede evitar la validación, pero no es tan costosa y complicada que la validación requerida para la codificación manual.

La codificación asistida por ordenador, a veces es preferible a la codificación automática, porque combina la inteligencia humana con la capacidad del ordenador. Pero, mientras que la codificación totalmente automática permite codificadores con poco entrenamiento, la codificación asistida por ordenador necesita codificadores entrenados.

## **8.2 Codificación de Bienes y Servicios en las Encuestas de Gastos Familiares**

En las EGF, se usan varias técnicas de codificación para las diferentes variables no cuantitativas. Se usan preguntas precodificadas, por ejemplo, para las características de los miembros de la familia, con tal que consten de un número limitado de categorías (sexo, último colegio, etc.) y a menudo se usa la técnica de categorías precodificadas para los artículos de gasto en el cuestionario. La otra información verbal se codifica manualmente o automáticamente.

En Israel, algunas variables se codifican manualmente y sus códigos se escriben primero en los cuestionarios y luego se imputan en el ordenador (por ejemplo, Ocupación e Industria), mientras que otras se codifican manualmente pero los códigos se teclean directamente en el ordenador (por ejemplo, Tipo de comercio, en los diarios). En Israel y en algunos otros países, la codificación de bienes y

servicios se hace automáticamente, pero en la mayoría de los países todavía se emplea la codificación manual. A continuación vamos a centrarnos en la codificación de la información verbal, dado que es específica en las EGF.

(a) Ambito y dificultades

Aunque las muestras de las EGF normalmente constan de sólo varios miles de familias (por ejemplo, 5.000 en la encuesta israelí de 1986-87, 6.000 en la encuesta sueca de 1985, aprox. 7.000 al año en la encuesta continua de los EE.UU.), la codificación de bienes y servicios tiene una escala mucho mayor que la codificación de información en otras encuestas de muestreo. Es porque el número medio de artículos registrados en los diarios es bastante grande. Por ejemplo, en Israel había que codificar aprox. 750.000 artículos de diario en la encuesta EGF de 1986-87, dado que se registraron una media de 150 artículos en cada uno de los diarios de dos semanas.

La codificación de estos artículos es una tarea enorme, y además se encuentran otras dificultades. Dado que la mayoría de los registros en los diarios los hacen los mismos encuestados (excepto en los casos raros cuando lo hace el entrevistador), los técnicos o los mecanógrafos tienen que interpretar miles de diferentes caligrafías. Además, la codificación es incluso más difícil en los países donde la población consta de grupos de diferentes orígenes, nuevos inmigrantes, o en los países multilingües, especialmente cuando se usan lenguas con diferentes caracteres, como en Israel donde el hebreo, árabe y, últimamente, ruso son permitidas para registrar en los diarios. A veces los encuestados describen los bienes sólo por su nombre comercial, sin los detalles adicionales necesarios para su codificación, y esto puede ser más frecuente en los países desarrollados donde hay una amplia variedad de bienes del mismo tipo. Aunque los entrevistadores están instruidos a comprobar los registros en los diarios, quedan muchas descripciones deficientes.

(b) El Sistema de Clasificación

Todos los artículos en las EGF relacionados con transacciones financieras se clasifican por un sistema jerárquico comprensivo para simplificar la agregación de grupos más detallados. En general, el sistema de clasificación jerárquica consta de cuatro clases principales: (1) Ingreso; (2) Gastos no de consumo; (3) Gastos de consumo; y (4) Ahorros. Dentro de estas cuatro clases se definen grupos principales y dentro de ellos, subgrupos que se divide en los grupos finales (que contiene un artículo específico o un grupo de artículos).

Como ejemplo, el sistema de clasificación que se usa en las EGF de Israel se basa en un código jerárquico de 5 dígitos (más un dígito de control para reducir los errores mecanográficos en la codificación manual). El número de clases se presenta en la Tabla 8.1. En algunos países el número de códigos es mucho mayor, como en Holanda (3.000) y en Dinamarca (1.500), mientras que otros países usan menos como en el Reino Unido (400) y en Francia (350) (Oficina Estadística de los Paises Europeos, 1981).

**Tabla 8.1**  
**Número de Clases en el Sistema de Clasificación**  
**EGF de 1986-87 de Israel**

| Clase primaria<br>(1º dígito) | Grupos princip.<br>(2º dígito) | Subgrupos<br>(3º dígito) | Grupos finales<br>(4º,5º dígitos) |
|-------------------------------|--------------------------------|--------------------------|-----------------------------------|
| Ingreso                       | 4                              | 15                       | 41                                |
| Gasto no consumo              | 2                              | 4                        | 6                                 |
| Gasto consumo                 | 10                             | 70                       | 686                               |
| Ahorros                       | 3                              | 11                       | 14                                |
| <b>Todas clases primarias</b> | <b>19</b>                      | <b>100</b>               | <b>747</b>                        |

Aunque se intenta decidir unas normas de clasificación consistentes, no es fácil lograrlo. Los diferentes países podrían usar diferentes sistemas de clasificación con diferentes números de clases finales, adaptados a sus propias necesidades y a sus hábitos de consumo específicos. Incluso en una encuesta en particular, las normas de clasificación no siempre son consistentes. Es porque las EGF son de multipropósito y no es siempre posible construir un solo sistema de clasificación jerárquica adecuado para todos los usos de la encuesta sin tener muchas excepciones.

Una norma general que se podría aplicar, como un ejemplo, sería clasificar una comodidad según su uso. Según esta norma, todos los productos de comida se clasifican bajo un grupo principal de gasto, aunque a veces, verduras y fruta se consideran como un grupo principal por separado.

Dentro de estos grupos principales, se clasifican los productos de comida según su naturaleza general, su valor nutricional, su método de preparación y según otras consideraciones específicas. Esto no siempre es práctico, por ejemplo, el gasto en "comida fuera de casa" no se puede dividir en cada artículo consumido y entonces se clasifica según el lugar de la comida (restaurante, quiosco, cafetería del trabajo, etc.) y por el tipo de comida (bocadillo, comida completa, etc.).

Si se hubiera aplicado la norma de clasificar una comodidad según su uso, un televisor sería clasificado bajo "entretenimiento" o "educación", un frigorífico y platos bajo "comida", una lavadora bajo "ropa", etc. Pero, considerando los usos de la encuesta, no se puede aplicar totalmente este principio, así que, la mayoría de los aparatos eléctricos (aunque no todos) normalmente se clasifican bajo "electrodomésticos", y platos bajo "equipamiento de hogar".

Otro ejemplo de la dificultad en clasificar es cuando se adquiere materia prima de una fuente y el trabajo de otra. El gasto general del producto final se puede clasificar según su uso, o dividirlo en componentes, es decir, el gasto se divide entre la materia prima y la labor. Por ejemplo, cuando se suministra tela a un sastre, el gasto de la tela se incluye bajo "hilos" y el pago al sastre bajo "costura y sastrería". Sin embargo, si es un traje a medida se incluye todo el gasto bajo "trajes".

Un ejemplo de inconsistencia es cuando "reparaciones, "seguros", "licencias", etc. para cualquier tipo de equipamiento se clasifican bajo la misma clase que el equipamiento, mientras que, "seguro de vivienda" se clasifica en una clase separada bajo "vivienda", y no se divide entre los diferentes artículos que este seguro incluye.

También puede haber diferencias de opinión sobre cómo clasificar a los varios artículos. Un ejemplo es el gasto en los diferentes tipos de actividades deportivas: la pregunta es si se deberían clasificar bajo "salud" o como parte de "educación, cultura y ocio". También surgen problemas al clasificar el gasto en "viajes al extranjero", si se clasifica bajo "transporte" o "entretenimiento", o, cada parte del gasto por separado bajo "comida", "transporte", "alojamiento" etc.

En cuanto a "regalos" (excepto transferencias de dinero), una norma podría clasificar los regalos dados a otras familias como "regalos" (bajo gasto no de consumo) y clasificar los regalos recibidos, según su tipo, como consumo. Otra norma es la de clasificar los regalos dados a otros como gasto en consumo según el tipo de comodidad.

No es sorprendente que sea imposible obtener un sistema de clasificación único cumpliendo estrictamente unas normas sencillas. Por eso, no es fácil determinar los códigos correctos, y por lo tanto la tarea de codificación se complica aún más.

### 8.3 Evaluación de la Codificación Manual

Aunque cualquier estudio de evaluación de errores de codificación, al igual que con la mayoría de los estudios sobre la evaluación de los errores de respuesta, está sujeto a la dificultad de decidir los valores verdaderos (incluso si los determinan los expertos que hacen esfuerzos especiales para encontrar el verdadero código), los estudios de evaluación pueden proporcionar unos resultados útiles. Se evaluaba la codificación manual (después de validación) de los artículos en los diarios mensuales en la EGF de Israel de 1979-80 (Kantorowitz, 1981). Esto se hizo, primero, para estimar los efectos de la codificación errónea sobre la exactitud de las estimaciones de la encuesta, y en segundo lugar, para estudiar los ratios de la codificación errónea por tipo y fuente de error para poder mejorar la codificación de las futuras encuestas.

Se seleccionaba una submuestra controlada de 120 familias, equilibrada por tamaño de familia, educación del cabeza de familia, y mes de investigación. Para cada uno de los 27.200 artículos incluidos en la submuestra, se definía un código "verdadero". Dos expertos comprobaron independientemente todos los códigos originales, grabando cada caso con error. Cualquier desacuerdo entre los dos expertos se comprobaba de nuevo. Se hizo un esfuerzo especial para determinar los códigos "verdaderos", por ejemplo, se usaban los precios individuales de ciertos artículos cuando la descripción en el diario era insuficiente; se buscaba información adicional sobre los artículos descritos sólo por su nombre comercial; se utilizaba información de los cuestionarios, etc. Para cada error encontrado, se codificaba también el tipo y fuente del error. A continuación se presentan algunos resultados de este estudio.

(a) Codificación de Errores por Tipo y Fuente de Error

De todos los códigos, 8.2% eran erróneos y su distribución por tipo y fuente se presenta en la Tabla 8.2. Como se esperaba, los errores de codificación son más frecuentes si la clasificación requerida es muy detallada. De hecho, de todos los errores, 70.7% eran entre los grupos finales dentro de los subgrupos (tipo A); 14.8% eran entre los subgrupos dentro de los grupos principales (tipo B); 11.4% eran entre los grupos principales de gasto (tipo C); 2.4% eran entre los grupos de consumo y los grupos de no consumo (tipo D); y 0.7% eran dentro de los grupos de no consumo (tipo E).

**Tabla 8.2**  
**Codificación de Errores, por Tipo y Fuente de Error (porcentajes)**  
**EGF de 1979-80 de Israel**

| Fuente de error          | Tipo de error |       |       |       |       |
|--------------------------|---------------|-------|-------|-------|-------|
|                          | Total         | A     | B     | C     | D & E |
| <b>TOTAL</b>             | 100.0         | 70.7  | 14.8  | 11.4  | 3.1   |
| (1) Libro de códigos     | 100.0         | 75.8  | 12.6  | 11.2  | 0.4   |
| (2) Codificador          | 100.0         | 62.8  | 18.2  | 14.8  | 4.2   |
| (3) Descripción artículo | 100.0         | 76.7  | 12.2  | 7.3   | 3.8   |
| - Evitable(*)            | 100.0         | 70.0  | 17.1  | 5.9   | 7.0   |
| - Inevitable             | 100.0         | 78.9  | 10.6  | 7.8   | 2.7   |
| <b>TOTAL</b>             | 100.0         | 100.0 | 100.0 | 100.0 | 100.0 |
| (1) Libro de códigos     | 24.9          | 26.8  | 21.3  | 24.4  | 2.9   |
| (2) Codificador          | 41.9          | 37.2  | 51.4  | 54.3  | 56.5  |
| (3) Descripción artículo | 33.2          | 36.0  | 27.3  | 21.3  | 40.6  |
| - Evitable(*)            | 8.4           | 8.3   | 9.7   | 4.4   | 18.9  |
| - Inevitable             | 24.8          | 27.7  | 17.6  | 16.9  | 21.7  |

(\*) Al emplear precios, o datos de los cuestionarios.

Fuente: Kantorowitz (1981)

Las deficiencias en el libro de códigos causaron 25% de todos los errores porque las categorías no eran mutuamente excluyentes o no eran del todo incluyentes, es decir, cuando dos artículos diferentes con nombres parecidos no se diferenciaban suficientemente, o cuando el diccionario no tenía un código asignado para un artículo. La mayoría de estos errores implicaban sólo un número limitado de artículos y entonces hubiera sido relativamente fácil evitarlos. Los errores debidos a equivocaciones de los codificadores eran 42% de todos los errores y éstos se relacionaban con la experiencia y entrenamiento del codificador, y también con la eficiencia del control de calidad.

El resto, 33%, eran debidos a una descripción deficiente del artículo en el diario, y se podría haber evitado un 8%, codificando correctamente si se hubiera hecho un esfuerzo adicional para utilizar los precios disponibles o la información del cuestionario de la familia. Aproximadamente 40% de los errores de tipos D y E son de esta fuente y más o menos la mitad se podría haber evitado al usar la información adicional disponible. Una cuarta parte de los

otros errores de esta fuente se podría haber evitado si tanto los entrevistadores como los encuestados hubieran sido más conscientes de la necesidad de una información más específica.

Los errores de codificación no se distribuyeron proporcionalmente sobre todos los grupos de artículos. Los ratios de error por tipo se estimaban para los diferentes grupos de artículos para guiar las mejoras en futuras operaciones de codificación, concentrándose en los grupos más vulnerables. Los resultados, por los grupos principales se presentan en la Tabla 8.3. Por ejemplo, el grupo de Transporte y comunicación tenía un ratio significativamente alto de codificación errónea (24%), mayoritariamente de tipo A (22%). Ropa y calzado, y Muebles y equipo de hogar también tenían ratios de error relativamente altos (un 18% cada uno), pero menos que dos tercios eran del tipo A en los dos grupos y tipo B (5% y 2% respectivamente). Verduras y fruta tenía ratios de error relativamente bajos para todos los tipos (en total 4%).

**Tabla 8.3**  
**Codificación de Errores por Principales Grupos de Gasto**  
**(porcentajes)**  
**EGF de 1979-80 de Israel**

| Grupos principales de gasto (*) | Porcentajes de todos los códigos | Porcentaje de códigos erróneos | Ratio de codificación errónea(**) |
|---------------------------------|----------------------------------|--------------------------------|-----------------------------------|
| TOTAL                           | 100.0                            | 100.0                          | 8.1                               |
| Transporte y comunicación       | 5.7                              | 17.0                           | 24.1                              |
| Muebles y equipo hogar          | 1.2                              | 2.7                            | 18.5                              |
| Ropa y calzado                  | 2.2                              | 4.8                            | 17.7                              |
| Salud                           | 1.0                              | 1.9                            | 13.9                              |
| Educación, cultura y ocio       | 6.0                              | 6.5                            | 8.8                               |
| Mantenimiento vivienda          | 3.8                              | 4.1                            | 8.7                               |
| Varios bienes y servicios       | 4.8                              | 4.3                            | 7.3                               |
| Comida(excl.Verde y fruta)      | 54.3                             | 47.3                           | 7.0                               |
| Verduras y fruta                | 21.0                             | 11.4                           | 4.4                               |

\*) No incluye Vivienda porque rara vez necesitaba codificación

\*\*) Excluyendo tipo E.

**(b) Efectos de los Errores de Codificación sobre las Estimaciones de la Encuesta**

Los errores de codificación pueden afectar a las estimaciones de la encuesta al incrementar la varianza o, si los errores no se anulan mutuamente, al causar sesgo. El sesgo relativo,  $\beta(k)$ , de la estimación del diario debido a los errores de codificación para cada grupo de gasto  $k$ , se estimaba por la diferencia porcentual entre la estimación basada en los códigos iniciales ( $I$ ) y la que se basaba en los códigos "verdaderos" ( $T$ ), relativa a la estimación "verdadera" por  $\beta(k) = [X(k)_I/X(k)_T] - 1$ .

La estimación de la encuesta para cada grupo se deriva del diario (como Comida), o del cuestionario (como Electrodomésticos), o de

ambas fuentes. Así que la estimación del sesgo relativo de la estimación de la encuesta debido a errores de codificación es  $\beta^*(k) = \beta(k)P^*(k)$ , donde  $P^*(k)$  es la proporción de la contribución del diario a la estimación total de cada grupo  $k$ , y  $1-P^*(k)$  de la estimación está libre de errores de codificación.

La Tabla 8.4 presenta los sesgos relativos estimados,  $\beta(k)$  y  $\beta^*(k)$ , para los grupos principales de gasto y para algunos subgrupos seleccionados de artículos. Además, los errores relativos de muestreo,  $C(k)$ , de los mismos grupos de artículos de la misma encuesta, también se presentan.

Por lo tanto,  $E(k) = C^2(k) / [C^2(k) + \beta^2(k)]$  expresa la disminución del tamaño efectivo de la muestra debido a errores de codificación (ignorando otras fuentes de error), o, hasta qué punto se podría reducir los costes de campo para un nivel dado de exactitud de las estimaciones de la encuesta, si no existieran errores de codificación.

**Tabla 8.4**  
**Sesgo Relativo, Error Muestral Relativo y Disminución**  
**de Eficiencia Debido a Errores de Codificación**  
**Grupos Principales de Gasto y Subgrupos Seleccionados**  
**EGF de 1979-80 de Israel**

| Grupos de gasto                 | Porcentaje de gasto de diario<br>P(k) | Sesgo relativo(%)       |                         | Error muestral relativo<br>C(k) | Eficiencia relativa<br>E(k) |
|---------------------------------|---------------------------------------|-------------------------|-------------------------|---------------------------------|-----------------------------|
|                                 |                                       | de diario<br>$\beta(k)$ | general<br>$\beta^*(k)$ |                                 |                             |
| TOTAL                           | 44.1                                  | -0.13                   | -0.06                   | 1.1                             | 1.00                        |
| <b>Grupos principales(*):</b>   |                                       |                         |                         |                                 |                             |
| Comida(excl.Verde. y fruta)     | 100.0                                 | +0.30                   | +0.30                   | 1.0                             | .92                         |
| Verduras y fruta                | 100.0                                 | +0.80                   | +0.80                   | 1.1                             | .65                         |
| Ropa y calzado                  | 100.0                                 | -0.53                   | -0.53                   | 2.6                             | .96                         |
| Varios bienes y servicios       | 42.8                                  | +3.63                   | +1.55                   | 2.5                             | .72                         |
| Transporte y comunicación       | 33.1                                  | +1.34                   | +0.44                   | 2.5                             | .97                         |
| Mantenimiento vivienda          | 31.9                                  | -0.17                   | -0.05                   | 1.4                             | 1.00                        |
| Educación, cultura y ocio       | 27.8                                  | +0.14                   | +0.04                   | 2.3                             | 1.00                        |
| Muebles y equipo hogar          | 17.8                                  | -1.80                   | -0.32                   | 3.3                             | .99                         |
| Salud                           | 8.9                                   | -0.73                   | -0.06                   | 2.5                             | 1.00                        |
| <b>Subgrupos seleccionados:</b> |                                       |                         |                         |                                 |                             |
| Pan, cereales y productos pasta | 100.0                                 | -1.01                   | -1.01                   | 1.3                             | .62                         |
| Aceites vegetales y productos   | 100.0                                 | -0.07                   | -0.07                   | 2.9                             | 1.00                        |
| Carne y aves                    | 100.0                                 | +1.10                   | +1.10                   | 1.5                             | .65                         |
| Productos lácteos               | 100.0                                 | +0.34                   | +0.34                   | 1.1                             | .91                         |
| Azúcar y productos              | 100.0                                 | -2.47                   | -2.47                   | 2.0                             | .40                         |
| Comidas fuera de casa           | 100.0                                 | -6.48                   | -6.48                   | 4.9                             | .36                         |
| Verduras frescas                | 100.0                                 | +0.72                   | +0.72                   | 1.5                             | .81                         |
| Fruta                           | 100.0                                 | -0.13                   | -0.13                   | 1.5                             | .99                         |
| Servicio doméstico              | 100.0                                 | -2.39                   | -2.39                   | 6.9                             | .89                         |
| Varios artículos de hogar       | 100.0                                 | -0.01                   | -0.01                   | 2.4                             | 1.00                        |
| Decoración de casa              | 40.9                                  | -9.96                   | -4.07                   | 6.5                             | .72                         |
| Ropa caballero                  | 100.0                                 | +14.97                  | +14.97                  | 7.6                             | .20                         |
| Ropa mujer                      | 100.0                                 | +0.34                   | +0.34                   | 4.9                             | 1.00                        |
| Ropa niño                       | 100.0                                 | +26.21                  | +26.21                  | 4.5                             | .03                         |
| Calzado                         | 100.0                                 | -1.16                   | -1.16                   | 3.8                             | .91                         |

\*) No incluye Vivienda porque rara vez necesita codificación 227

Como se puede ver en la Tabla 8.4, incluso cuando  $\beta^*(k)$  es bastante pequeño,  $E(k)$  es significante. Por ejemplo, aunque en Verduras y fruta  $\beta^*(k)=+0.8\%$  es pequeño,  $E(k)=0.65$ ; para Comida (excluyendo Verduras y fruta)  $\beta^*(k)=0.3\%$  y  $E(k)=0.92$ . Para muchos subgrupos de artículos  $E(k)$  es aún más bajo, por ejemplo, para Azúcar y sus productos  $E(k)=0.40$ .

El sesgo relativo estimado para casi todos los grupos principales era menos de 1%, excepto para Varios bienes y servicios, donde  $\beta(k)=+3.6\%$  y  $\beta^*(k)=+1.6\%$ . En cuanto a los subgrupos, los sesgos relativos eran generalmente más altos: algunos de ellos padecían un grave sesgo, como Ropa de niño ( $\beta^*(k)=+26\%$ ), y Ropa de caballero ( $\beta^*(k)=+15\%$ ). Para algunos de los artículos más detallados (no presentados en la Tabla 8.4), el sesgo era incluso mayor, por ejemplo, era de 4.3% para las estimaciones del gasto en tipos específicos de pan, mientras que para el grupo general de pan este sesgo era pequeño (-0.5%), es decir, era difícil distinguir entre artículos muy parecidos entre sí.

Estos resultados, a pasar de sus limitaciones, indican la importancia de invertir más esfuerzo en todos los medios de mejorar la exactitud de la codificación: un libro de códigos mejor, mejor entrenamiento de los codificadores, un control de calidad comprensivo, y el desarrollo de los sistemas automáticos de codificación.

#### 8.4 Codificación Automática

Los primeros intentos a desarrollar una codificación automática, hace unos 20 años, fueron para la Oficina de Censo de los Estados Unidos (Corbert, 1972 y O'Reagan 1972). Hoy en día se está experimentando con la codificación automática en muchos países y ya la están empleando, o planifican usarla en un futuro próximo, en algunas oficinas nacionales. En la mayoría de las oficinas estadísticas, el desarrollo de la codificación automática se originó con la codificación complicada de Ocupación e Industria, aunque en algunos países, la codificación automática se desarrolló para la codificación sencilla de las variables geográficas (lugar de residencia, país de nacimiento, lugar de trabajo, etc.), de la calificación educacional, etc. La codificación automática de bienes y servicios no es tan complicada como la codificación de Ocupación e Industria, pero no se puede considerar tan sencilla como de País de nacimiento o de Tipo de comercio.

Sin embargo, en las EGF, la codificación automática para bienes y servicios se usa en pocos países. La primera aplicación de la codificación automática de bienes era en la EGF sueca de 1978, y desde entonces se ha usado en las EGF suecas de 1985 y 1988, en Dinamarca, en Holanda y en Israel (Lyberg y Dean, 1992). En estos países se usan diferentes versiones de codificación automática de bienes y servicios y la versión que se implantaba primero en Israel en la EGF de 1986-87 (Berg y Har, 1990) que se describe a continuación.

Se debería mencionar que la construcción del diccionario para la codificación automática y algunas de las normas en el algoritmo de codificación se adecuaban a las características especiales de la

lengua hebrea que difiere del inglés en sintaxis y en estructura. Por ejemplo, no hay una palabra separada para "the" en hebreo, en lugar de esto se añade un prefijo al nombre, así reduciendo el número de palabras que describen un artículo; apenas existen vocales en hebreo, por lo tanto, las palabras son más cortas; los nombres vienen delante de los adjetivos, así que la búsqueda en el diccionario por la primera palabra en algunos casos es suficiente para la codificación; y hay una distinción de géneros y también de las formas singular/plural, lo que incrementa la posibilidad de identificar a un artículo.

(a) Construcción del Diccionario Almacenado en el Ordenador

Se preparaba un fichero inicial de las descripciones verbales de bienes y servicios con sus correspondientes códigos, basándose en varias fuentes: (1) el libro de códigos usado para la codificación manual en la encuesta anterior después de ponerlo al día; (2) una muestra de diarios con unos 22.000 registros de la EGF anterior; (3) una muestra de diarios de un pretest de la EGF de 1986-87 con unos 10.000 registros; y (4) listas de comodidades, con sus nombres comerciales, de las cadenas de supermercados y grandes almacenes.

Cada descripción del diccionario se transformó a una forma estándar de un máximo de tres palabras, cada palabra con un máximo de 8 caracteres. Se añadió a cada descripción el código apropiado, después de una validación completa. Para algunas descripciones se podría determinar el código con sólo la primera palabra, mientras que, para otras la primera palabra no era suficiente y se necesitaba la segunda palabra, o incluso la tercera.

Así que, para hacer más eficiente el proceso de codificación, se definían "indicadores" numéricos para cada palabra de la descripción del diccionario, para que fuera posible saber si la primera o segunda palabra sería suficiente para la codificación. Por ejemplo, todos los registros de "Vino" tenían el mismo código. Por lo tanto, los siguientes registros, escritos en hebreo (donde los nombres están delante de los adjetivos), tenían el mismo código: "VINO TINTO", "VINO DULCE", VINO SECO", etc. Así que para estos registros la primera palabra "VINO" determinó el código, y esto se indicaba en el diccionario. Otro ejemplo es queso. Una palabra no era suficiente para codificar los varios tipos de queso. Se necesitaban dos palabras para todos los tipos de "QUESO BLANCO", entonces, "QUESO BLANCO GRASO" y "QUESO BLANCO NOGRASO" tenían el mismo código. Pero los diferentes tipos de "QUESO AMARILLO" tenían diferentes códigos, así que el registro de "QUESO AMARILLO DURO" no se codificaba igual que "QUESO AMARILLO BLANDO", así que se necesitaban tres palabras para la codificación.

El diccionario incluía información adicional sobre la mayoría de los artículos, tal como las posibles ponderaciones de la unidad y posible tipo de comercio. Esta información se utilizaba para comprobar la validez de los códigos asignados.

Se borraban del diccionario los artículos que podrían causar ambigüedad, a menos que se pudieran comprobar estos artículos en etapas posteriores del proceso, por el precio, o ponderación, o tipo de comercio. Por ejemplo, si "Batería de coche" se registró en el

diario sólo como "Batería" (en inglés pila) esta ambigüedad se podría detectar luego por el precio, y entonces no se borraba del diccionario. La detección de las ambigüedades se realizó listando todas las referencias del diccionario en orden alfabético.

El diccionario se puso al día durante el proceso de la codificación, especialmente durante las primeras etapas. Los artículos sin código que ocurrieron frecuentemente se añadieron al diccionario, se corregieron los errores detectados y se borraban las ambigüedades. Al principio el diccionario incluía unos 2.500 artículos y, al final del proceso, este número aumentó hasta 3.000. Se debería mencionar que los artículos que aparecían rara vez en los diarios se excluían del diccionario, para prevenir una sobrecarga del mismo. Además, se preparaba un fichero inicial de "sinónimos" de unas 1.000 palabras que se suelen deletrear erróneamente y que también se ponía a día continuamente.

#### (b) El Proceso de Codificación

Se procesaba la codificación continuamente durante el período de investigación de la encuesta en 24 lotes, cada lote constaba de 200 diarios de dos semanas, con un total de unos 30.000 registros. Toda la información (numérica y descripción verbal) de todos los registros de los diarios de un cierto lote, se imputaba en el ordenador, tecleando la información. Los mecanógrafos tenían instrucciones de teclear la información verbal siguiendo unas normas sencillas, por ejemplo, suprimiendo ciertas preposiciones y sufijos.

La descripción de cada artículo se transformó por medio del proceso computerizado en la forma estándar de tres palabras, donde el criterio para acabar una palabra era al menos un espacio en blanco, y cada palabra se truncó después de los primeros 8 caracteres. Entonces, cada palabra de la descripción de los artículos se emparejaba con el fichero de sinónimos para "traducirla" si fuera necesario en la palabra correcta, igual que en el diccionario.

El siguiente paso era buscar la descripción en el diccionario y así el código correspondiente. El emparejamiento se hizo por medio de un algoritmo con el objetivo de minimizar el tiempo de búsqueda en el ordenador y que se conformaba a las características específicas del idioma hebreo. En principio, tanto los artículos del diccionario como la descripción estándar de los registros en los diarios se ordenaban en orden alfabético. Luego se buscaba la primera palabra de cada registro de diario entre las primeras palabras de los artículos del diccionario. Si se obtenía un emparejamiento exacto y el indicador asignado al diccionario indicaba que la primera palabra era suficiente para codificar, se asignaba el código; si no, se buscaba la segunda palabra, aplicando las mismas normas, y luego la tercera palabra si era necesario. Si no se obtenía un emparejamiento exacto para la primera palabra, no se buscaba más, e igualmente con la segunda y tercera palabras cuando se requerían dos o tres palabras.

La mayoría de los códigos iniciales pasaban por un número de validaciones para poder distinguir entre descripciones similares de artículos con códigos diferentes. Por ejemplo, "café" puede ser "paquete de café" comprado en el supermercado, o "taza de café" en

una cafetería. Tanto el tipo de comercio como el peso y especificación de la unidad (taza, paquete, etc.) se usaba en esta validación. Estos dos detalles del artículo en el diario se emparejaban con aquellos del diccionario para determinar la legitimidad del código asignado. Si se emparejaban se aceptaba el código inicial, si no, se borraba y no se obtenía ninguna codificación automática. No se validaban todos los artículos del diccionario porque no era necesario o porque no se disponía de datos de validación. Se debería señalar que si había más validación de los códigos asignados después del proceso de codificación, por ejemplo, se comprobaban el gasto y la cantidad comprada con el rango de precios posibles.

De la codificación automática para cada lote de códigos resultaban dos ficheros: un fichero que incluía los registros codificados y el segundo fichero que incluía los registros que fracasaban en la codificación automática. Los registros sin código aparecen en orden alfabético para su codificación manual, con la cual los códigos se teclean directamente en el ordenador. A menudo el codificador podría teclear un sólo código que se copiaba automáticamente para todos los artículos de la misma descripción.

En general, en la encuesta de 1986-87, 81% de todos los registros de los diarios se codificaban automáticamente y este ratio se incrementó del 75% en el primer lote hasta el 83% en el último (en la encuesta sueca de 1978, el ratio total de codificación era de 65%, pero en la encuesta de 1985 subió hasta 82%, y en Dinamarca el ratio era de 75% (Lyberg y Dean, 1992). Se debería señalar que 1% de los artículos del diccionario codificaban al 50% de todos los registros de diarios, es decir, unos pocos artículos del diario eran muy frecuentes, como tipos de pan, leche, periódicos y tabaco.

#### (c) Control de Calidad y Evaluación del Proceso de Codificación

Se llevó a cabo un control de calidad para el proceso de codificación por varias razones:

- (1) Para controlar el nivel de calidad de la operación de mecanografía
- (2) Para entrenar más a los mecanógrafos, y que identificasen sus errores más frecuentes
- (3) Para detectar deficiencias en el diccionario o en el fichero de sinónimos y para mejorarles para su uso en los lotes de códigos siguientes
- (4) Para controlar la codificación manual de aquellos registros que fallaron en la codificación automática .

Para el control de calidad se usaba una muestra de 10% de todos los registros de los diarios (de codificación automática o de codificación manual) de cada lote de códigos. Todos los registros de la muestra se codificaban de nuevo manualmente e independientemente de la codificación inicial y el codificador no sabía si el código había sido asignado automáticamente o no. Luego, se emparejaban los códigos iniciales con los códigos controlados. Cuando se emparejaban bien se suponía que no había ningún error en el código asignado. Si los dos códigos no eran iguales, un experto decidía independientemente cuál era el código correcto (no necesariamente

uno de los dos) y entonces se determinaba si el código inicial era erróneo o no. Cuando se asignaba un código erróneo inicialmente se corregía y se comprobaba la fuente del error: el diccionario, el fichero de sinónimos, la mecanografía, o la codificación manual. Se corregía el diccionario y el fichero de sinónimos, si hacia falta, y se avisaba al codificador y al mecanógrafo de los errores frecuentes.

En la encuesta anterior, donde sólo se usaba una codificación manual, los ratios estimados de códigos erróneos eran de 8.2% en la encuesta de 1986-87. Donde se usaba codificación automática con alguna codificación manual suplementaria, este ratio era de 3.6%. Después de la corrección de los códigos erróneos en la muestra de control de calidad, el ratio total estimado de códigos erróneos se redujo a 3.2% (bajó incluso más después de comprobar los precios en una etapa separada del proceso).

Como media de todos los lotes, 3.0% de todos los códigos automáticos tenían errores: 1.8% debido al diccionario o el fichero de sinónimos; 1.0% debido a errores de mecanografía en la descripción del artículo; y 2.0% debido a otras razones. Estos ratios eran más altos para las primeras etapas del proceso de codificación y eran más bajos en las últimas etapas, dado que se corregía continuamente el diccionario y los mecanógrafos eran más eficientes.

Como se ha mencionado antes, en la encuesta anterior el ratio total de errores de codificación manual se estimó en 8.2%, y el ratio de códigos erróneos de codificación manual en la EGF 1986-87 de artículos no codificados automáticamente era de 6.2%. Se podría esperar que el ratio de los errores en la codificación no automática en la encuesta de 1986-87 sería más alto que el ratio total en 1979-80, a causa de una codificación más complicada para artículos que no se codificaban automáticamente, por ejemplo, artículos menos frecuentes, descripciones inusuales en el diario, etc.

Esto se puede explicar por varias razones: había un mejor libro de códigos para esta encuesta (basado en los resultados del estudio de evaluación de la codificación manual en la encuesta anterior); codificadores más calificados en la encuesta de 1986-87; en la encuesta anterior la codificación era peor debido a la fatiga y aburrimiento; la asignación del mismo código a varios artículos con la misma descripción que resultó con menos errores en 1986-87; y quizás, debido a menos control de calidad en 1986-87.

Para concluir, la codificación automática resultó ser preferible a la codificación manual. Primero, el ratio de codificación automática era satisfactorio y se obtenía un ratio mucho menor de codificación errónea. Segundo, comparada con la codificación manual usada en la encuesta anterior (en la misma escala de codificación), se ahorraba 60% del coste del trabajo. Sin embargo, se invertía un esfuerzo considerable en desarrollar el procedimiento de codificación automática para su primer empleo, pero a largo plazo, esto se puede considerar como una inversión.

## BIBLIOGRAFIA

- Appel M. and Hellerman E. (1983). Census Bureau Experience with Automated Industry and Occupation Coding. Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 32-40.
- Berg A. and Har S. (1990). Report on Automatic Coding in the 1986/87 Family Expenditure Survey in the Israel Central Bureau of Statistics, Jerusalem (Unpublished - Hebrew).
- Breiman L., Friedman J., Olshen R. and Stone C. (1984). Classification and Regression Trees. Monterey, California: Wadsworth, Inc.
- Corbert J.P. (1972). Encoding from Free Word Descriptions. Unpublished manuscript, U.S. Bureau of the Census.
- Cowan C. (1977). Incentive Effects on Amounts Reported in an Expenditure Diary Survey. Proceedings of the Section on Social Statistics, American Statistical Association, Part I, pp. 498-503.
- Elliot D. (1991). Weighting Expenditure and Income Estimates from the United Kingdom Family Expenditure Survey to Compensate for Nonresponse. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Elteto O. (1991). Nonresponse Rates and a Way of Adjusting for Nonresponse in Household Expenditure Surveys in Hungary. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Embrey B. (1988). The ASCO Computer Assisted Coding System. Paper presented at the Social Research Conference, University of Queensland, Brisbane.
- Glaude M. (1982). A Decade of Methodological Investigations on Family Expenditure Surveys in France. Paper presented at the International Meeting on Analysis of Sample Survey Data, Jerusalem, Israel.
- Harrison R. (1991). Respondent Burden and Respondent Fatigue in the 1988-89 Australian Household Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- de Heer W.F. (1991). A New Method for the Budget Survey of 1988. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Israel Central Bureau of Statistics (1963). Family Expenditure Surveys 1950/51-1956/57-1959/60. Special Series No. 148. Jerusalem, Israel.
- Jacobs E., Jacobs C. and Dippo C. (1989). The U.S. Consumer Expenditure Survey. ISI Proceedings of the 47<sup>th</sup> Session, Paris, Vol. 1, pp. 123-142.
- Kantorowitz M. (1981). Evaluation of Some Aspects of the Israeli Family Expenditure Survey. ISI Proceedings of the 43<sup>rd</sup> Session, Buenos Aires, Vol. 3, pp. 1563-1582.
- Kantorowitz M. (1990). Improving Post-stratification by Constructing Hierarchical Classification Trees of Homogeneous Groups. Paper presented at the IOS/IASS Summer Conference on Survey Design, Methodology and Analysis, Essex, England.
- Kemsley W.F.F. and Nicholson J.L. (1960). Some Experiments in Methods of Conducting Family Expenditure Surveys. Journal of the Royal Statistical Society, Series A, Vol. 123, pp. 307-328.

- Kemsley W.F.F. (1961). The Household Expenditure Enquiry of the Ministry of Labour: Variability in the 1953-54 Enquiry. *Applied Statistics, Journal of the Royal Statistical Society*, Vol. 10 No. 3, pp.117-135.
- Kemsley W.F.F. (1975). Family Expenditure Survey. A Study of Differential Response Based on a Comparison of the 1971 Sample with the Census. *Statistical News*, No. 31, pp. 16-21.
- Kordos J. and Kubiczek A. (1991). Methodological Problems in the Household Budget Surveys in Poland. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Lindstrom H.L. (1991). An Experiment with Incentives. *Research & Development Report, Statistics Sweden*, No. 10, pp. 3-13.
- Lyberg L. and Dean P. (1992). Automated Coding of Survey Responses: An International Review. Paper presented at the Work Session on Statistical Data Editing, Conference of European Statisticians, Working Paper No. 3. Washington D.C.
- Matheson J. (1991). Application of Computer Assisted Interviewing to the United Kingdom Family Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- McWhinney I. and Champion H. (1974). The Canadian Experience with Recall and Diary Methods in Consumer Expenditure Surveys. *Annals of Economic and Social Measurement*, 3/2, pp. 411-437.
- Moller B. (1991). The Relationship Between Data from Accounting Books, Interviews and Registers. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Nevraumont U. (1991). Evaluation of the Canadian Food Diary Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- O'Reagan R.T. (1972). Computer-assigned Codes from Verbal Responses. *Communications of the ACM*, Vol. 15, pp. 455-459.
- Pearl R. B. (1979). Reevaluation of the 1972-73 U.S. Consumer Expenditure Survey. Technical Paper No. 46. U.S. Bureau of the Census, Washington, D.C.
- Prais S.J. (1958). Some Problems in the Measurement of Price Changes with Special Reference to the Cost of Living. *Journal of the Royal Statistical Society, Series A*, Vol. 121 Part 3, pp. 312-332.
- Redpath B. (1986). Family Expenditure Survey: a Second Study of Differential Response, Comparing Census Characteristics of FES Respondents and Non-respondents. *Statistical News*, No. 72, pp. 13-16.
- Russo A. and Coccia G. (1991). On the Characteristics and Effects of Non-responses in the Italian Consumer Expenditure Survey. Paper presented at the Workshop on Diary Surveys, Stockholm, Sweden.
- Silberstein A. (1989). Recall Effects in the U.S. Consumer Expenditure Interview Survey. *Journal of Official Statistics*, Vol. 5 No. 2, pp. 125- 142.
- Statistical Office of the European Communities (1981). The 1979 Harmonized Family Budget Surveys. Review of the Methodological Aspects of the Questionnaires in the Ten Member States.
- Statistics Canada (1990). Survey of Family Expenditures in 1990, Interviewer's Manual. Ottawa, Canada.

- Sudman S. and Ferber R. (1971). Experiments in Obtaining Consumer Expenditures by Diary Methods. *Journal of the American Statistical Association*, Vol. 66 No. 336, pp. 725-735.
- Tucker C. (1992). The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey. *Journal of Official Statistics*, Vol. 8 No. 1, pp. 41-61.
- Turner R. (1961). Inter-week Variation in Expenditure Recorded During a Two-week Survey of Family Expenditures. *Applied Statistics*, *Journal of the Royal Statistical Society*, Vol. 10 No. 3, pp. 136-146.
- U.S. Bureau of Labor Statistics (1988). BLS Handbook of Methods. Bulletin 2285. U.S. Department of Labor, Washington, D.C.

## **SEMINARIO INTERNACIONAL DE ESTADISTICA NAZIOARTEKO ESTATISTIKA MINTEGIA**

### **TEXTOS PUBLICADOS HASTA LA FECHA ORAIN ARTE ARGITARATUTAKO GAIAK**

- N.<sup>º</sup> 1. Zbkia.: LINEAL STATISTICAL INFERENCE  
(Inglés, Euskera, Español)  
Autor/Egilea: C.R. RAO  
Año/Urtea: 1983
- N.<sup>º</sup> 2. Zbkia.: MUESTREO Y APLICACIONES  
(Español, Euskera)  
Autor/Egilea: E. CANSADO  
Año/Urtea: 1983
- N.<sup>º</sup> 3. Zbkia.: STATISTICAL EDUCATION  
(Inglés, Euskera, Español)  
Autor/Egilea: V. BARNETT  
Año/Urtea: 1983
- N.<sup>º</sup> 4. Zbkia.: ANALYSE DES DONNEES  
(Francés, Euskera, Español)  
Autor/Egilea: P. CLAPIER  
Año/Urtea: 1983
- N.<sup>º</sup> 5. Zbkia.: DESIGN OF EXPERIMENTS  
(Inglés, Euskera, Español)  
Autor/Egilea: D.J. FINNEY  
Año/Urtea: 1984
- N.<sup>º</sup> 6. Zbkia.: ASPECTOS DE TEORIA Y APLICACIONES EN EL MUESTREO  
(Español, Euskera)  
Autor/Egilea: F. AZORIN POCH  
Año/Urtea: 1984
- N.<sup>º</sup> 7. Zbkia.: CURSO BASICO INTENSIVO DE MUESTREO  
(Español, Euskera)  
Autor/Egilea: J.L. SANCHEZ-CRESPO  
Año/Urtea: 1985

- N.<sup>o</sup> 8. Zbkia.: ANALYSE DES SERIES CHRONOLOGIQUES: LES INDICES STATISTIQUES  
(Francés, Euskera, Español)  
Autor/Egilea: J. FOURASTIE  
Año/Urtea: 1985
- N.<sup>o</sup> 10. Zbkia.: METHODOLOGY AND TREATMENT FOR NON-RESPONSE  
(Inglés, Euskera, Español)  
Autor/Egilea: R. PLATEK  
Año/Urtea: 1986
- N.<sup>o</sup> 11. Zbkia.: STATISTICAL OPERATIONS BY SAMPLING  
(Inglés, Euskera, Español)  
Autor/Egilea: L. KISH  
Año/Urtea: 1986
- N.<sup>o</sup> 12. Zbkia.: ANALISIS DE SERIES TEMPORALES: ALGUNAS TECNICAS DE PREDICCION  
(Español, Euskera)  
Autor/Egilea: I. GALLASTEGI  
Año/Urtea: 1986
- N.<sup>o</sup> 13. Zbkia.: BASES DE DATOS  
(Español, Euskera)  
Autor/Egilea: F. SALTOR  
Año/Urtea: 1987
- N.<sup>o</sup> 14. Zbkia.: METODOS ESTADISTICOS PARA LA INVESTIGACION EPIDEMIOLOGICA  
(Español)  
Autor/Egilea: L.C. SILVA  
Año/Urtea: 1987
- N.<sup>o</sup> 15. Zbkia.: SAMPLING AND NON-SAMPLING ERRORS IN SURVEYS  
(Inglés)  
Autor/Egilea: A. MARTON  
Año/Urtea: 1988
- N.<sup>o</sup> 16. Zbkia.: LES ENQUÊTES TELEPHONIQUES<sup>A</sup>  
(Francés)  
Autor/Egilea: V. SALVY  
Año/Urtea: 1988
- N.<sup>o</sup> 17. Zbkia.: GENERALIZED LINEAR MODELS IN EPIDEMIOLOGY  
(Inglés)  
Autor/Egilea: J.C. DUFFY  
Año/Urtea: 1989
- N.<sup>o</sup> 18. Zbkia.: NEW TECHNOLOGIES IN COMPUTER ASSITED SURVEY PROCESSING  
(Inglés)  
Autor/Egilea: J.G. BETHLEHEM and W.J. KELLER  
Año/Urtea: 1989
- N.<sup>o</sup> 19. Zbkia.: EVALUATION OF QUESTIONNAIRE DESIGN EFFECTS  
(Inglés)  
Autor/Egilea: G. NATHAN  
Año/Urtea: 1990
- N.<sup>o</sup> 20. Zbkia.: PROCEDIMIENTO DE DEPURACION DE DATOS ESTADISTICOS  
(Inglés)  
Autor/Egilea: I. VILLAN CRIADO / M.S. BRAVO CABRIA  
Año/Urtea: 1990

N.<sup>o</sup> 21. Zbkia.: THE X11ARIMA / 88 SEASONAL ADJUSTMENT METHOD  
(Inglés)

Autor/Egilea: E. BEE DAGUM  
Año/Urtea: 1990

N.<sup>o</sup> 22. Zbkia.: RENTA Y DISTRIBUCION DE LA RIQUEZA, DESIGUALDAD Y POBREZA: TEO-  
RIA, MODELOS Y APLICACIONES  
(Español)

Autor/Egilea: C. DAGUM  
Año/Urtea: 1991

N.<sup>o</sup> 23. Zbkia.: LA CONTABILIDAD NACIONAL COMO MARCO DE LAS ESTIMACIONES DE  
VARIABLES ECONOMICAS  
(Español)

Autor/Egilea: V. ANTON VALERO  
Año/Urtea: 1991

N.<sup>o</sup> 24. Zbkia.: MACRO-EDITING. METHODS FOR RATIONALIZING THE EDITING OF QUANTI-  
TATIVE DATA.  
(Inglés)

Autor/Egilea: L. GRANQUIST  
Año/Urtea: 1991

N.<sup>o</sup> 25. Zbkia.: METHODOLOGICAL ISSUE IN FAMILY EXPENDITURE SURVEYS  
(Inglés, Euskera, Español)

Autor/Egilea: M. KANTOROWITZ  
Año/Urtea: 1992