

1998

the replication method for  
estimating sampling errors

el método de replicación  
para la estimación de  
errores de muestreo

DAVID MORGANSTEIN

37

Seminario internacional de estadística en euskadi

1998

the replication method for  
estimating sampling errors

el método de replicación  
para la estimación de  
errores de muestreo

DAVID MORGANSTEIN

37



*Lanketa / Elaboración:*

**Euskal Estatistika-Erakundea /**  
Instituto Vasco de Estadística - EUSTAT

*Argitalpena / Edición:*

**Euskal Estatistika-Erakundea /**  
Instituto Vasco de Estadística - EUSTAT  
Duque de Wellington, 2 - 01010 Vitoria-Gasteiz

© Euskal A.E.ko Administrazioa  
Administración de la C.A. de Euskadi

*Botaldia / Tirada:*

500 ale / ejemplares  
XI-1998

*Inprimaketa eta koadernaketa: /*

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

*ISBN: 84-7749-322-0*

*Lege-gordailua / Depósito legal: S.S. 1.395/98*

## AURKEZPENA

Nazioarteko Estatistika Mintegia antolatzean, hainbat helburu bete nahi ditu EUSTAT-Euskal Estatistika Erakundeak:

- Unibertsitatearekiko eta, batez ere, estatistika-sailekiko lankidetza bultzatzea.
- Funtzionarioen, irakasleen, ikasleen eta estatistikaren alorrean interesatuta egon daitezkeen guztienei lanbide-hobekuntza erraztea.
- Estatistika alorrean mundu mailan abangoardian dauden irakasle eta ikertzaile ospetsuak Euskadira ekartzea, horrek eragin ona izango baitu, zuzeneko harremanei eta esperientziak ezagutzeari dagokicenez.

Jarduera osagarri gisa, eta interesatuta egon litekeen ahalik eta pertsona eta erakunde gehienetara iristearren, ikastaro horietako txostenak argitaratzea erabaki dugu, betiere txostengilearen jatorrizko hizkuntza errespetatuz; horrela, gai horri buruzko ezagutza gure herrian zabaltzen laguntzcko.

Vitoria-Gasteiz, 1998ko Azaroa

GORKA KNÖRR BORRÀS  
Zuzendari Nagusia

## PRESENTATION

In promoting the International Statistical Seminars, EUSTAT-The Basque Statistics Institute wishes to oachieve several aims:

- Encourage the collaboration with the universities, especially with their statistical 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 worlwide 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 people and institutions as possible, it has been decided to publish the papers of these courses, always respecting the original language of the author, to contribute in this way towards the growth of knowledge concerning this subject in our country.

Vitoria-Gasteiz, November 1998

GORKA KNÖRR BORRÀS  
General Director

## **PRESENTACION**

Al promover los Seminarios Internacionales de Estadística, el EUSTAT-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, Noviembre 1998

**GORKA KNÖRR BORRÀS**  
Director General

## BIOGRAFI OHARRAK

David Morganstein Westat-eko lehendakariordea da, eta bertako estatistika-langileen zuzendaria. Aditua da laginketa-diseinu konplexuetarako bariantzen zenbatespenean eta kalitatearen hobekuntzan, estatistika-metodoak erabiliz. Morganstein jaunak hainbat erakunderekin lan egin du (45tik gora), besteak beste, mundu zabaleko zenbait estatistika-bulegorekin, kalitatea hobetzeko metodoei buruz. Kalitatearen hobekuntzari buruzko ikastaro asko eman ditu, besteak beste, Bristol-eko Biltzarrean eta ISIko hitzaldietan eman zituenak; era berean, hitzaldi asko eman ditu laginketa gaiei buruz, hala nola laginketa-diseinuei eta bariantzen zenbatespenei buruz.

David Morganstein Purdue-ko Unibertsitateko ingenaria da, eta Michigan-go Unibertsitateko Estatistikako lizenziatura. Irakaslea da inkesta-metodoetarako programa bateratuan, zeina Maryland-eko Unibertsitateak, Michigan-go Unibertsitateak eta Westat-ek babestu baitute. Amerikako Estatistika Elkarteko zuzendaritzako kidea izan da sei urtez; horietariko sei, diruzain lanetan. Amerikako Estatistika Elkarteko kidea da, eta ASAko (Amerikako Estatistika Elkarteko) sortzaileen saria eman diote. Era berean, ISIko (International Statistical Institute) eta IASSeko (International Association of Survey Statisticians) kidea da.

## BIOGRAPHICAL SKETCH

David Morganstein is a Vice President of Westat, and Director of its statistical staff. He is a senior statistician specializing in variance estimation for complex sample designs and quality improvement using statistical methods. Mr Morganstein has worked with more than 45 organizations, including statistical agencies worldwide on methods for improving quality. He has taught many courses in quality improvement, including courses at the ASA sponsored Bristol Conference and the ISI meetings, and in sample survey topics such as sample design and variance estimation.

Mr Morganstein is a graduate of Purdue University in Engineering and the University of Michigan in Statistics. He is an instructor in the Joint Program for Survey Methods, co-sponsored by the University of Maryland, the University of Michigan and Westat. He has served on the American Estatistical Association Board of Directors for seven years and as the ASA's Treasurer for four of those years. He is a Fellow of the ASA and a recipient of the ASA Founder's Award. He is also a member of the ISI and the International Association of Survey Statisticians.

## NOTAS BIOGRAFICAS

David Morganstein es Vicepresidente de Westat, y Director de su personal estadístico. Es experto en la estimación de varianza para diseños de muestreo complejos y mejora de la calidad utilizando métodos estadísticos. El Sr. Morganstein ha trabajado con más de 45 organizaciones, incluyendo agencias estadísticas de todo el mundo, sobre métodos para mejorar la calidad. Ha impartido muchos cursos sobre mejora de la calidad, incluidos los impartidos en la Conferencia de Bristol y en las conferencias del ISI, y también sobre temas de muestreo tales como diseños muestrales y estimación de varianza.

David Morganstein es ingeniero por la universidad de Purdue y licenciado en Estadística por la Universidad de Michigan. Es profesor en el Programa Conjunto para Métodos de Encuestas, patrocinado por la Universidad de Maryland, la Universidad de Michigan y Westat. Ha colaborado en el Consejo de Dirección de la Asociación Americana de Estadística durante siete años, cuatro de ellos como Tesorero. Es miembro de la Asociación Americana de Estadística y se le ha otorgado el Premio de los Fundadores de la ASA (Asociación Americana de Estadística). Es también miembro del ISI (International Statistical Institute) y de la IASS (International Association of Survey Statisticians).

## **LAGINKETAKO ERROREAK ZENBATESTeko IHARDESPEN-METODOA: IHARDESTEAk NOLA SORTU**

David Morganstein, J. Michael Brick, Pam Broene eta Mary G. Nixon  
Westat, Inc

Ihardespina metodo erabilgarria da estatistika-datu ez-linealak laginketa-inkesta konplexuetan izaten dituzten bariantzeta hurbiltzeko. Emaitza asintotikoki baliokideak eskaintzen ditu, metodo alternatiboaren (linealizazio-metodoaren) aldean, eta abantaila aipagarriak erakusten ditu praktikan sortzen diren egoera arruntenetako batzuetan. Azterlan honetan, azaltzen dugu nola aplika daitekeen ihardesporen teoria inkestadiseinu bakoitzean. Ihardespina erabiltzeak aukera ematen digu helburu garrantzitsu batzuk lortzeko, zeinak izatez oso zailak baitira linealizazio-metodoak erabiliz. Hona hemen zertan diren ihardesporen abantailak: oso erraza da ihardespina datuen erabiltzaileei azaltzea, zeren ez baita beharrezko laginketa-inkestetako bariantzei buruzko prestakuntza berezi bat izatea; ia estatistika-datu guztiei aplika daki, jardunbide berea erabiliz eta softwarea aldatu gabe; zenbatespen-estrategia asko erabil ditzake, ponderazioak doitzeko bide bat edo bestea erabiliz; jardunbide finkoa da, datuen azpimultzoekin eta datu absenteekin lan egiteko; metodo erraza eta dotorea da, luzera-lagineko diseinuak kudeatzeko, eta diagnostikorako ezaugarriak eskaintzen ditu, zeinak baliagarriak izaten baitira ponderazio-arazoak aurkitzeko eta inuesta berrien planak egiteko.

1. atalean, inuesta-zenbatespeneko laginketa-erroreak zenbatesteko metodo batzuei buruzko sarrera labur bat egiten dugu: metodo zehatzak, linealizazioa eta zenbait ihardespenei metodo. Sartutako erreferentziek adierazten dutenez, hurbilketa-metodo horiek asintotikoki baliokideak dira. Beraz, hurbilketa-metodoen artean aukeratzea beste ezaugarri batzuen mende dago, hala nola erabilerraztasuna eta ulerterraztasuna. 1. atal horretan, ihardesporen teoria berrikusten da, eta linealizazio-metodoari aplikatzen ez zaizkion abantaila batzuen zerrenda eskaintzen du.

2., 3., 4. eta 5. atalek laginketa-diseinu arruntak azaltzen dituzte, eta ihardespenei nola sortu azaltzen du. 2. atalak etapa bakarreko laginketa-discinuei buruzkoa da, geruzapena dutenei eta ez dutenei buruzkoa. 3. atalak bariantzak zenbatesteko hurbilketa-metodo bat azaltzen du, baldin eta laginketa-unitate bakarra hautatzen bada. Era berean, populazio bukakorraren laginketarekin lotutako gai garrantzitsu batzuk erabiltzen ditu, hala nola biztanleria bukakorraren zuzenketaren erabilpena eta bariantzak zenbatesteko askatasun-maila. Honako kontzeptu hau erakusten da: bi oinarritzko laginketa-unitate (OLU) eta geruzak nola konbinatu, diseinuei dagokien kalkulu-ahalegina murrizteko; zeren, bestela, ihardeste asko izango bailitzkete. 4. atalak 2. ataleko gaiak sakontzen ditu, eta etapa askotariko ordainen gaia erabiltzen du. Unitate buru-ordezkarriak (ziurtasuna) daudenean ihardespenei nola sortzen diren, berriz, 5. atalean azaltzen da.



## C O N T E N T

INTRODUCTION .....	15
1.- BACKGROUND .....	16
1.1 Alternative Methods .....	16
1.2 Review of Replication Theory .....	18
1.3 Some Strengths of Replication .....	19
Ease of Use .....	19
Handling Different Estimation Methods .....	20
Handling Subsets and Missing Data .....	21
Handling Longitudinal or Panel Survey Designs .....	23
Diagnostic Features .....	23
2.- SINGLE-STAGE SAMPLE DESIGNS .....	24
2.1 Sampling Without Stratification .....	24
2.2 Stratified Sampling .....	25
3.- ISSUES IN FINITE SAMPLING APPLICATIONS .....	26
3.1 Finite Population Correction .....	27
3.2 Degrees of Freedom of Variance Estimate .....	28
3.3 Combining PSUs for Variance Estimation .....	31
3.4 Combining Strata for Variance Estimation .....	33
3.5 Collapsing Strata for Variance Estimation .....	34
4.- MULTI-STAGE SAMPLE DESIGNS .....	35
4.1 Issues in Multi-stage Samples .....	35
4.2 Procedures for Some Sample Designs .....	36
5.- SELF-REPRESENTING UNITS .....	39
5.1 Single-Stage Designs .....	39
5.2 Multi-Stage Designs .....	39

6.- CREATING REPLICATES FOR THE NHIS .....	41
6.1 Introduction to the NHIS .....	41
6.2 Overview of the NHIS Sample Design and Estimation .....	41
6.3 Variance Estimation for the NHIS .....	42
6.4 Properties of the Variance Estimators .....	45
6.5 Comparison of the Methods .....	45
7.- SUMMARY .....	47
REFERENCES .....	48
APPENDIX .....	51
BOOKS PUBLISHED UP TO DATE .....	123

# ÍNDICE

INTRODUCCION .....	69
1.- Introducción .....	70
1.1 Métodos alternativos .....	70
1.2 Revisión de la Teoría de Replicación.....	72
1.3 Algunas ventajas de la Replicación .....	73
Facilidad de uso.....	73
El manejo de diferentes métodos de estimación.....	74
Gestión de subconjuntos de datos y datos ausentes.....	76
Manejo de diseños longitudinales o encuestas de panel.....	77
Diagnóstico .....	78
2.- DISEÑOS MUESTRALES MONO-ETAPICOS .....	78
2.1 Muestreo sin estratificación .....	79
2.2 Muestreo estratificado.....	80
3.- TEMAS RELACIONADOS CON APLICACIONES DE MUESTREO FINITO .....	81
3.1 Corrección de población finita.....	81
3.2 Grados de libertad de la estimación de varianza.....	83
3.3 La combinación de UMP para la estimación de varianza.....	84
3.4 La combinación de estratos para la estimación de varianza.....	86
3.5 Agrupación de estratos para la estimación de varianza.....	88
4.- DISEÑOS MUESTRALES MULTI-ETAPICOS .....	88
4.1 Aspectos de los muestreos multi-etápicos.....	89
4.2 Procedimientos para algunos diseños muestrales .....	90
5.- UNIDADES AUTO-REPRESENTANTES .....	92
5.1 Diseños mono-etápicos .....	92
5.2 Diseños multi-etápicos .....	93

6.- CREACION DE REPLICADAS PARA LA NHIS .....	95
6.1 Introducción a la NHIS .. ..	95
6.2 Visión global del diseño muestral y la estimación de la NHIS.....	95
6.3 Estimación de varianza dc la NHIS.....	96
6.4 Propiedades de los estimadores de varianza .....	99
6.5 Comparación de los métodos.....	100
7.- CONCLUSION.....	101
BIBLIOGRAFIA.....	103
APENDICE.....	105
CURSOS PUBLICADOS HASTA LA FECHA .....	123

## **THE REPLICATION METHOD FOR ESTIMATING SAMPLING ERRORS: CREATING REPLICATES**

David Morganstein, J. Michael Brick, Pam Broene, and Mary G. Nixon  
Westat, Inc.

Replication is a very useful method of approximating the variance of nonlinear statistics in complex sample surveys. It provides asymptotically equivalent results to the alternative linearization method, and it has distinct strengths under some very common situations that arise in practice. In this paper, we describe how replication theory can be implemented in a variety of survey designs. Using replication provides the opportunity to achieve some important objectives that are inherently difficult with linearization methods. The strengths of replication include: it is easy to explain replication to data users without special training in variance estimation for sample surveys; it can be applied to nearly all statistics using the same procedure without modifying the software; it can handle a variety of estimation strategies including adjustments of various kinds to the weights; it is a consistent procedure for handling subsets of the data and missing data items; it is a simple and elegant method for handling longitudinal sample designs; and it offers diagnostic features useful for uncovering weighting problems and planning new surveys.

In Section 1, we provide some statistical background including a brief review of various methods for estimating the sample errors of survey estimates: exact methods; linearization, and several replication approaches. The included references indicate that these approximate methods are asymptotically equivalent. Thus, a choice between approximate methods depends on other features, such as ease of use and understandability. Section 1 includes a review of replication theory, and then lists a number of advantages not found with the linearization method.

Sections 2 through 5 discuss common sample designs and indicate how replicates can be created for them. Section 2 covers single-stage sample designs both with and without stratification. Section 3 describes an approximate method for estimating variances when only a single sampling unit is selected. It also discusses a number of important issues in finite population sampling, such as the use of the finite population correction and the degrees of freedom for estimating variances. The concept of combining both primary sampling units (PSUs) and strata to reduce computational effort for designs that would otherwise have many replicates is introduced. Section 4 expands the topics in Section 2 to cover multi-stage counterparts. Replicate creation when there are self-representing (certainty) units is discussed in Section 5. Section 6 describes a practical application of replication to the U.S. National health Interview Survey in which a procedure requiring 70 replicates provides reasonably precise estimates of sampling errors for subnational as well as national statistics.

## 1. BACKGROUND

Several approximate methods have been developed for estimating the sampling errors of complex statistics from surveys where the sample designs include multiple stages, stratification, clustering, and unequal selection probabilities. Given that survey weights are often subjected to adjustments for survey non-response and undercoverage, it is desirable for the sampling error approach to reflect the effect of these weight adjustments.

Section 1.1 discusses commonly used approaches to estimating sampling errors. Section 1.2 provides the statistical background behind the replication method. A number of advantages of the replication, particularly its ability to reflect almost all aspects of weight adjustments, are listed in Section 1.3.

### 1.1 Alternative Methods

When a survey is conducted using a complex sample design, the design must be taken into account to produce unbiased estimates and standard errors of the estimates. Most standard statistical software such as SPSS and the SAS® System ignore the sample design and assume the observations are independent and identically distributed. Using weighting features, the standard software can produce unbiased point estimates, but it is not possible to correctly estimate standard errors. In fact, if either the sample design or the estimator is complex, the exact expression for the variance of the estimator is often intractable and approximate methods are needed.

One approach to approximating the variance of an estimate is to use replication methods which include the jackknife (with the variants we call JK1, JK2 and JK $n$ ), balanced repeated replication (BRR), and bootstrap methods. The choice between methods depends in large part on the nature of the sample design, as discussed in detail later. Regardless of the choice, each approach involves dividing the full sample into  $G$ , not necessarily mutually exclusive, subsamples known as replicates so that each replicate sample, when properly weighted, provides an estimate of the population characteristic of interest. For the most part, in this paper we only discuss jackknife and BRR options to illustrate the main points.

In JK1, which is appropriate for unstratified designs, replicates are formed by excluding a PSU (or groups of combined PSUs) from a replicate subsample and adjusting the weights of the remaining PSUs so that they weight up to the population. In general, replicates are formed in ways that mirror the original sampling of PSUs. (On a historical note, one of the original replication methods, the random group approach, is essentially the JK1 jackknife method described here.)

Several methods including JK2, JK $n$  and BRR are used with stratified designs. Two of the methods, JK2 and BRR, are appropriate for designs involving exactly 2 PSUs in each stratum. In

the traditional BRR, replicates are formed by excluding one of the two sampled PSUs in each strata. Thus, only half of the PSUs are used in the estimate for each replicate. If there are  $H$  strata, then there are  $2^H$  possible BRR replicate samples. McCarthy (1966) shows that a much smaller number of properly selected “half-samples” contain the same information as the entire set of  $2^H$  possible samples. This minimal set is called a balanced half-sample and can be constructed using orthogonal patterns. (WesVar uses a set of orthogonal Hadamard matrices to create balanced half-samples when BRR replicates are requested). In a variation on the BRR due to Fay (Judkins, 1990), the two PSUs in the strata are both included but given different weights. One is given a weight between 0 and 1, say  $k$ , while the other PSU is given a weight of  $2-k$ . In all of the examples that follow where BRR is appropriate, the Fay method can be used.

In JK2, only a single PSU is excluded from a single strata to form each replicate. A complete set of JK2 replicates for a 2 PSU per stratum design requires  $H$  replicates, each corresponding to a replicate subsample that includes one PSU from a stratum. In a sense, a single JK2 replicate estimates the variance contributed by a single stratum.

JKn is used with designs involving more than 2 PSUs in some strata. It is similar to the JK2 approach, but requires more replicates. If a sampling stratum contains  $n_h$  sampled PSUs, then  $n_h$  replicates are formed by leaving out each of the sampled PSUs in turn. Thus, in a stratum with 2 PSUs, the JK2 method would result in two replicates where the JK2 method requires only one replicate.

An alternative approach is to linearize the estimator using a Taylor series expansion and then use standard sample survey variance estimation methods to estimate the precision of the linearized statistic. Both the linearization and replication methods are described in more detail in Wolter (1985).

Many theoretical and empirical studies compare replication and linearization methods of variance estimation. On the theoretical side, early work by Krewski and Rao (1981) show the asymptotic equivalence of these methods. Subsequent theoretical work, including that of Rao, Wu and Yue (1992), show that the linearization and replication methods are asymptotically equivalent to a reasonably high order. On the empirical side, simulation studies have been conducted to compare the precision of the estimates using both methods. One of the earliest of these studies is Kish and Frankel (1974). More recent work by Kovar, Rao and Wu (1988), Rao, Wu, and Yue (1992), and Valliant (1990) show that the methods generally result in consistent estimates. The jackknife's properties are very similar to those of linearization, while the properties of the BRR and bootstrap methods are very similar to each other. Most of the comparisons consider both the accuracy of the variance estimates and the derived confidence intervals. Rust and Rao (1996) review some of this literature and examine other aspects of the two approaches to approximating variances of estimates from complex sample surveys.

In this paper, we review circumstances in which the replication method has some distinct advantages over the linearization method for estimating variances in sample surveys. Since both methods result in estimates that are asymptotically equivalent, the criteria for assessing the strengths of the replication method are factors other than the precision of the estimates. Some of the features examined are the survey design and estimation strategies that can be more easily reflected appropriately with replication than linearization. Of course, in some situations linearization is simpler or easier to implement but most of these situations are covered in other literature (Shah et al., 1992).

The strengths of a method depend heavily on the software that is available for use in computing. WesVar Complex Samples, a Windows-based software package for computing variances using replication methods, is used to illustrate the strengths of the replication method in practical applications. WesVar Complex Samples is designed to compute functions of weighted totals and fit models to data from complex samples using a variety of replication techniques (Westat, 1998). WesVar Complex Samples (version 3) can be purchased from SPSS, Inc. (<http://www.spss.com>). (An earlier and less powerful version, WesVarPC, is available free-of-charge from the World Wide Web URL of Westat <http://www.westat.com>. The earlier, free version and its documentation can be downloaded from this site (Westat. 1997).)

In the next section, we provide a brief introduction to replication theory and methods. Particular features of sample surveys that may be more easily and accurately addressed using replication methods are discussed subsequently, with examples of how this can be accomplished using WesVar.

## 1.2 Review of Replication Theory

The basic idea underlying replication is to select subsamples repeatedly from the whole sample, to calculate the statistic of interest for each of these subsamples, and then to use the variability among these subsample or replicate statistics to estimate the variance of the full sample statistics. There are different ways of creating subsamples from the full sample. The subsamples are called replicates and the statistics calculated from these replicates are called replicate estimates.

Replicate variance estimators,  $v(\hat{\theta})$ , take the form

$$v(\hat{\theta}) = c \sum_{k=1}^G h_k (\hat{\theta}_{(k)} - \hat{\theta})^2 \quad (1)$$

where

- $\theta$  is an arbitrary parameter of interest,
- $\hat{\theta}$  is the estimate of  $\theta$  based on the full sample,

- $\hat{\theta}_{(k)}$  is the  $k$ -th estimate of  $\theta$  based on the observations included in the  $k$ -th replicate,
- $G$  is the number of replicates,
- $c$  is a constant that depends on the replication method, and
- $h_k$  is a stratum specific constant that is only required for certain designs.

The value of  $c$  depends on the replication method; for BRR and bootstrap  $c = 1/G$ ; for the unstratified jackknife (JK1),  $c = (G-1)/G$ , for the paired jackknife (JK2) and JK $n$ ,  $c = 1$ . The factor  $h_k$  is needed for jackknife designs with an unequal number of PSUs per stratum (JK $n$ ) and can also be used to adjust for finite population corrections in some designs. This issue is not discussed further in this paper. Rust and Rao (1996) describe replication methods more completely.

Designing the replicates is a critical feature that has important implications for the validity of the estimates of the variances. WesVar is software specifically designed to accommodate replicate methods; however, for some designs, replicate preparation requires considerable ingenuity so that precise variance estimates can be produced while not requiring extensive computations. For many designs, it is very simple. Sections 2 through 5 describe how various replication methods can be used to develop replicates for specific sample designs.

### 1.3 Some Strengths of Replication

#### Ease of Use

The choice of which method of variance estimation to use can be examined in a number of ways. One important dimension is the ability to describe how the variance estimates are produced to users without advanced training in variance estimation for sample surveys. Replication is a very well-known method in statistics and many users readily accept the basic premise of replication. Dividing the full sample into subsamples and estimating the variability of replicate subsamples around the full sample estimate is often the only explanation required.

This aspect of ease of use is reinforced by the fact that the same procedure is followed for nearly all statistics: means, totals, ratios, and more complex functions of the sample data. For example, the log-odds ratio for a two-way table from a sample can be written as:

$$\hat{\theta} = \log\left(\frac{\hat{N}_{11}\hat{N}_{22}}{\hat{N}_{12}\hat{N}_{21}}\right) \quad (2)$$

where the ‘^’ above the quantities indicate these are consistent, approximately unbiased sample estimates of the population characteristics. The variance of the sample log-odds ratio can be estimated exactly the same as the sample mean or total using (1).

This simplicity contrasts with the linearization method that requires each nonlinear statistic be linearized before the variance can be estimated. The linearized form of the statistic must be developed and incorporated into the software. While the derivation of the linearized form is often not difficult (Binder, 1996), the linearization software must either have the specific statistic already explicitly included in its list of statistics or it must numerically approximate the statistic. Either of these options places some limits on the data user.

A very important strength of replication is that it allows the user to define nearly any statistic and compute the variance of it using (1) without having to modify the software. WesVar does this in two ways. First, the user can specify a "Compute" statistic as a function of variables on the data set, as described above. For example, some "Compute" statistics that are easily defined in WesVar are ratios of ratios, the log of ratios, and the square root of a difference of ratios. An easy to use graphical interface is used to enter the statistic and the program computes (1) for that statistic. Second, the user can specify a "Function" statistic as a function of estimated quantities in the cells of a table. For example, the log-odds ratio example mentioned above can be estimated in WesVar by naming the four cells of the two-by-two table and specifying the log of the ratio of the estimated totals in the cells. User specified statistics, such as differences, contrasts, and logs of differences or ratios can be calculated this way. Once again, the simplicity is possible because the variance is estimated using (1) for almost all types of statistics.

### **Handling Different Estimation Methods**

In most sample surveys, it is not sufficient to simply weight the observed data by the reciprocal of the probability of selection. Different estimation strategies are used to reduce the bias and/or the variance of the estimates. For example, almost all surveys are subject to nonresponse and the weights are adjusted to compensate for this nonresponse (Brick and Kalton, 1996). Similarly, estimation strategies such as poststratification, raking, or generalized regression estimation are often used to take advantage of auxiliary data and reduce both the bias and the variances of the estimates. Replication offers a simple and elegant method of incorporating these adjustments in the estimates of variance (Yung, 1996).

Rust and Rao (1996) show that a simple way of creating the replicate estimates required in (1) is to create replicate weights and attach these to each observation just as the full sample weight is normally attached to an observation. The replicate estimates can then be computed by using the replicate weights the same way the full sample weights are used to create the full sample estimate. They also point out that, if the agency responsible for data collection includes the replicate weights with the data, this results in three important advantages: first, the user does not have to know any of the features of the sample design to properly estimate sample variances (the information is all contained in the replicate weights); second, the agency can reflect all stages of estimation appropriately in the replicate weights so as to reflect the full estimation scheme used in

the full sample weights; and, third, the software for replicate variance estimation does not have to be modified to compute the estimates properly (equation (1) is still used).

WesVar allows users to input replicate weights that have been adjusted for nonresponse, ratio estimation, poststratification, raking, calibration estimation, or any other commonly used method of estimation. Thus, the ordinary user can take advantage of the advanced estimation procedures used by the agency collecting the data to reduce the bias and the variance of the estimates, provided the agency places these replicate weights on the data file. For example, if nonresponse adjustments are used in creating the final weights, these same nonresponse adjustments can be applied separately to each replicate (each replicate thus has a different nonresponse adjustment) to reflect this procedure in the computed estimates. In addition, WesVar provides an option to allow the user to create replicate weights with poststratification or raking adjustments if the control totals required for these procedures are available and the replicate weights do not already include these adjustments.

This simplicity contrasts with the linearization approach, where each different estimation strategy requires a different development. For example, poststratification can be handled in linearization theory as a type of ratio estimation, but this implies that the estimator is different for the poststratified and the simple Horvitz-Thompson type of estimator. Thus, linearization software requires that the specific linearized poststratified estimator be included in the software to compute the variance appropriately. Nonresponse adjustments could also be handled in the linearization approach, but again it places a severe burden on the software to accommodate each type of estimation scheme. We are not aware of any linearization software that handles nonresponse adjustments in a simple manner. In replication, this is very simple if the agency providing the data includes this component in the replicate weights. The process to do this is straightforward and needs to be done only once when the weights are created.

### **Handling Subsets and Missing Data**

Since the replicate weights carry all the information needed for estimating variances, replication is also extremely well suited for handling data that are subset or missing. For example, a sample of adults might have been selected, but some analysts may wish to only estimate characteristics for a subset of the adults, say females between the ages of 30 and 44. This type of analysis is simple to do with replication methods without concerns about the implications for variance estimation.

For example, in WesVar the subset of the data can be identified using a simple subsetting statement and a subset file is created that contains only the records for the included adults, along with their full sample weight and replicate weights. The detailed analysis of just these records can then be conducted with the subset file since the replicate weights contain all the information needed to properly estimate the precision of the estimates.

This process is not as simple in most linearization software because the design information may no longer be accurate if the subset file is used alone. The basic problem is that the subset may not include observations from all of the sampled primary sampling units and most of the linearization software requires that all these units be identified in the data file used for analysis. Graubard and Korn (1996) show what might happen in this situation and illustrate the consequences with two examples from actual surveys. Here again, it is feasible to handle this situation under the linearization approach, but the steps required are more complex and more susceptible to error than the simple replication approach.

Another circumstance that is common in sample surveys but difficult to handle is missing data. Some of the issues associated with handling nonresponse adjustments of the weights that are normally used to adjust for unit nonresponse were mentioned earlier. Now, we turn our attention to item nonresponse and the missing data from this source.

Handling missing items in analysis is a long-standing problem without a fully accepted, general-purpose solution. It is too ambitious to expect to handle missing data in variance estimation to everyone's satisfaction when a method has not been accepted for handling the estimates themselves. However, a more modest goal is to handle the missing items in a consistent manner for creating the estimate and for estimating the variance of the estimates. In this regard, replication has a decided advantage over linearization methods.

Once a procedure has been determined for handling missing data for creating an estimate, that same procedure can be applied to create both the full sample and the replicate estimates. WesVar does this; the full sample and replicate weights are used in exactly the same way with respect to missing data for creating the components needed for computing (1). For example, in estimating a total with missing item responses, the full sample estimate and the replicate estimates may both be biased downward because of item nonresponse (if no imputation is done for the item), but the variance of the estimate is computed appropriately for the procedure.

In the linearization approach, this type of consistency is more difficult to achieve. In fact, many of the linearization software packages require all the items have responses in order to be able to compute variances. Other linearization software packages allow missing data, but do it in ways that are not consistent because the variance depends on the sample design and the specific form of the estimator. For example, the problem mentioned above with subsetting may occur due to missing item responses, but it may be very hard to detect that this has occurred. The methods of handling missing data also may result in estimates for strata being computed differently than estimates from the full sample.

The procedures for handling missing data can have important consequences. For example, the software may compute estimates of totals by the equivalent of imputing a value of zero for the

missing item, while means may be computed by ignoring the observations with missing data. The result is that different statistics, means and totals, are not treated the same. An example shows how important this can be. Suppose there are 4 respondents, 3 reporting values of 99, 100, and 101, while 1 has a missing value. In WesVar and replication methods, the estimated total for the full sample would be 300 and each replicate estimate would be within 100 to 300, so that the estimated variance would not be very large. On the other hand, if the linearization program imputed a value of zero for the missing value, then the estimated variance would be very large (the 4 observations for computing the variance would be 99, 100, 101, and 0).

### **Handling Longitudinal or Panel Survey Designs**

Panel or longitudinal sample surveys have become very popular in the last several decades, especially because they offer the possibility of measuring change over time more precisely than independent cross-sectional surveys. The increased precision is the result of the correlation between the units sampled over time. However, some of the benefits of having a more precise estimate cannot be realized unless the estimates of precision reflect the correlated responses.

Replication methods provide a rather simple and elegant method for incorporating the correlation. The replicates defined for units sampled for the baseline sample can be used also for the follow-up data collections so that the replicate estimates defined in (1) naturally contain the appropriate overlap and thus the estimates of precision reflect the design. Hinkins et al (1996) discuss this issue in more detail. Once again, WesVar can be used without modification to produce appropriate estimates of precision for estimates, provided the replicate weights for the follow-up design are produced correctly.

For linearization software, the problem is more complex. For some estimates it is possible to construct new variables that are the difference between the data collected in the different waves of the survey, but this may involve additional burdens on the data users. In some cases, elaborate procedures are required to address this issue.

### **Diagnostic Features**

In addition to being used in the computation of sampling error estimates for survey statistics, the set of replicated estimates can play a diagnostic role as well. Weighting problems may be detected by examining the replicated estimates and looking for an unusual, 'outlying' value. The JK2 and JK<sub>n</sub> methods provide a situation where individual strata variances can be examined. Examination of the replicate variance estimates can be helpful in assessing how the variance components are distributed across strata.

## 2. SINGLE-STAGE SAMPLE DESIGNS

In this section, we discuss single-stage designs and some approaches to constructing replicates for these designs. It should be noted that the approaches given are useful but there are alternatives that are also valid. Designs involving two or more stages are covered in Section 4. The section begins by discussing designs without stratification and then examines the topic of stratified designs.

Unstratified designs are not very common in the United States. Perhaps the most typical design falling into this category is a systematic sample from an ordered list. The JK1 approach is suggested for the unstratified, single-stage design.

The discussion of stratified designs is further subdivided by the number of PSUs selected in each stratum. When exactly 2 PSUs are sampled from every stratum, either the BRR or JK2 methods can be used. When more than 2 PSUs are selected, the JK $n$  method can be applied directly so long as the number of replicates are not too large as to result in extensive computing effort. If the computational effort that would result from a JK $n$  approach is too large, PSUs can be combined to reduce the number of replicates required. The combining of PSUs to reduce the number of replicates is described in Section 3.3.

Two terms used in the remaining sections are first discussed. In any sample design, there are primary sampling units, which could be the only sampling units, as well, for example, a single-stage list sample. In most samples, strata are also used. In many instances, the sampled PSUs and sampling strata will be used to create replicates. In other cases, the sampled PSUs and strata will be combined or collapsed (the distinction between the two will be clarified later) in order to reduce the number of replicates and the resulting computational effort. The actual clusters that are excluded from replicates, which may be PSUs or they may be combined or collapsed PSUs, will be denoted as VarUnits. The strata used in variance estimation will be denoted as VarStrat. (These terms are used in the WesVar software, as well).

### 2.1 Sampling Without Stratification

Two single-stage, unstratified designs commonly used are a simple random sample; and a systematic sample, possibly involving a sorted sampling frame that uses one or more variables in a sort order to achieve implicit stratification. In either case, the recommended solution is the same, the JK1 method.

#### **DESIGN 2.1: Single-stage designs where PSUs are sampled without stratification.**

In this design, use the JK1 replication method. Define each PSU as a VarUnit.

**Example 2.1:** A simple random sample of 50 patients is selected from a hospital for a study. The sampled patients are identified by a patient ID variable with values 1 to 50.

In this situation, use JK1 and use the patient ID variable as the VarUnit (see Appendix Example 2.1).

**Variant 2.1:** A systematic sample of PSUs is taken, possibly from an ordered list.

Although this variant is more common than DESIGN 2.1, there is no unbiased method for estimating the sampling errors of a single systematic sample. This design can be viewed as selecting a single cluster of units. With only a single selection, variances cannot be estimated in an unbiased way.

A common approach is to divide the systematic sample into a set of mutually exclusive and exhaustive, interleaved systematic subsamples. The sampling error is estimated as the variance among these smaller systematic samples. For example, if the systematic sample consists of 1,000 units, the replicates can be 50 subsamples drawn systematically, each consisting of 20 units. The first replicate contains units 1, 51, 101, ... 951 (every 50<sup>th</sup> unit starting with the first sampled unit). The second subsample contains units 2, 52, 102, ... 952 (every 50<sup>th</sup> unit starting with the second sampled unit). The JK1 method is used with the 50 replicates as the VarUnits.

## 2.2 Stratified Sampling

The stratified sampling case is discussed in two parts. We begin with the simplest situation, that of exactly two PSUs sampled from each stratum (DESIGN 2.2). This case is well described in the literature. The BRR and JK2 methods were developed for variance estimation with such designs.

We next examine the case of more than two, but not a large number of PSUs sampled from one or more, but not a large number of strata (DESIGN 2.3). By ‘not a larger number,’ we mean a situation where the number of replicates required for a direct application of the JK<sub>n</sub> approach would result in acceptable computation time.

Other variations of stratified designs are covered later. When either a large number of PSUs are sampled or a large number of strata are used, it may be desirable to combine PSUs or strata to reduce the number of replicates and, therefore, the computational effort. Combining of PSUs and strata are covered in Section 3.3 and Section 3.4. The case of a single random selection from a stratum, another design in which unbiased estimates of sampling errors are not possible, is provided in Section 3.5. A discussion of estimating the sampling error when one or more certainty PSUs has been selected is provided in Section 5.

**DESIGN 2.2: Single-stage stratified sampling: 2 PSU are sampled from each stratum.**

In this design if the number of strata is not too large, then use either BRR (or Fay's) or JK2 method. Define the sampling strata as the VarStrat and let the sampled PSUs define the VarUnits. Randomly assign them values of 1 or 2 within each stratum. For brevity, we no longer mention the Fay method; however, it is a variant of BRR and appropriate whenever BRR is suggested.

**Example 2.2: The public elementary schools in a state are stratified by size, the number of students in the school. Two schools are randomly sampled in each of 50 size class stratum.**

In this example, use BRR or JK2. The variable identifying the size class of the schools is the VarStrat. The VarUnit is the variable used to identify the sampled school within VarStrat. The variable used should have two unique values, such as 1 and 2, in each VarStrat. (see Appendix Example 2.2). If BRR is used, 52 replicates can be defined that contain all the information for variance estimation. If JK2 is used, 50 replicates will be needed.

**DESIGN 2.3: More than 2 PSUs are sampled from a small number of strata.**

For this design, use the JK<sub>n</sub> replication method. Define the sampling strata as the VarStrat and PSUs as the VarUnit. The number of replicates will equal the number of PSUs.

**Example 2.3: A stratified simple random sample of 100 business establishments is selected from no more than 50 strata, each stratum having at least 2 PSUs sampled. The establishments could be sampled with equal probability or with unequal, size measures.**

In this example, use the JK<sub>n</sub> method. Label the sampling strata as the VarStrat. Establishments are the VarUnits. The number of replicates created is 100 (see Appendix Example 2.3).

### **3. ISSUES IN FINITE SAMPLING APPLICATIONS**

This section includes a number of important topics. We first discuss methods for reflecting the finite population correction (fpc) factor's effect in reducing the sampling error in without-replacement sample designs. Second, since the variability of variance estimates is captured in the degrees of freedom (df), we discuss how to assess the df in Section 3.2. Section 2 mentioned stratified designs in which a straightforward approach to constructing replicates would result in

extensive computational effort. Sections 3.3 and 3.4 discuss combining PSUs and combining strata as methods for reducing the number of replicates required to estimate variances. The last section discusses a method for estimating variances when only a single PSU is selected from a stratum.

### 3.1 Finite Population Correction

The fpc is a factor that reduces variance because a large fraction of the population is sampled. (If the entire population were to be included, there would be zero sampling error. We do not discuss the common, related but ambiguous case of estimating the variability of statistics from a census that has nonresponse.). Generally, the fpc is applicable only to single-stage samples with equal probability sampling. When incorporating the effect of the fpc, it usually is assumed that all sampled units respond.

- For SRS  $fpc = (1 - n/N);$  (3)

where  $n$  is the number of sampled and responding units, selected from a population of  $N$  units.

- For STSRS  $fpc_h = (1 - n_h/N_h);$  (4)

where  $n_h$  is the number of sampled and responding units, selected from a stratum population of  $N_h$  units.

The stratified version is difficult to apply with BRR unless there is a common fpc for all strata. See Wolter 1985 for one approach to this by revising respondent weights. On the other hand, the fpc is easily applied using JK2 or JKn since replicates are defined by strata. We recommend this approach in the samples below.

#### **DESIGN 3.1: Many PSUs are sampled from a small number of strata (and the fpc is significant).**

As long as the number of replicates is not excessive, use the JK $n$  replication method. Define the sampling strata as the VarStrat and PSUs as the VarUnit. If the fpc is significant for any VarStrat, it should be accounted for by multiplying that replicate's contribution by the appropriate variance reduction factor, (4). (WesVar can account for a significant fpc via the *Attach Factors* screen where the fpc factors are entered as if they were the  $h_k$  terms in (1).)

**Example 3.1: A stratified simple random sample of 50 business establishments is selected from four strata.**

In this example, use the JK<sub>n</sub> method. Label the sampling strata as the VarStrat and the establishments as the VarUnits. The number of replicates created is 50. For each of the 50 VarStrat, provide the fpc. (When using WesVar, use the *Attach Factors* screen where  $fpc_h = 1 - n_h / N_h$ . See Appendix Example 3.1).

**Variant 3.1.1: The sample has only 2 PSUs per stratum but the sampling fraction in some of the strata is large (i.e., greater than .20).**

In this variant, use the JK<sub>2</sub> method;. Unique fpc factors are used for each replicate or for sets of replicates with similar fpc factors. (If using the WesVar Program, use the *Attach Factors* screen to specify the desired fpc correction for each VarStrat). For strata with fpc's near 1, it may be unnecessary to include the fpc (see Appendix Variant 3.1.1).

**Example 3.1.1: A single-stage sample of high schools in a state is selected. The high schools are stratified into 50 size categories and two schools selected per stratum.**

**Suppose the percent of schools sampled in the largest five size category strata are: 50 percent in the largest, 30 percent in the next largest, 20 percent in the third largest, and 10 percent in the fourth and fifth largest. In all the remaining strata, less than 5 percent of the schools are sampled.**

Use the size category variable as the VarStrat. The school identification variables (labeled 1 and 2 within the size class) are the VarUnit. If the largest 5 values of VarStrat correspond to the largest size categories, attach factors of 0.5, 0.7, 0.8, 0.9, and 0.9 to these replicates. The sampling fractions in the remaining VarStrat are so small that a factor of 1.0 is reasonable for each of them (see Appendix Example 3.1.1).

### 3.2 Degrees of Freedom of Variance Estimate

The degrees of freedom of the variance estimate are used primarily to compute the width of the confidence interval, C.I., for the estimated statistic, often via a constant obtained from the t-distribution (or the equivalent for hypothesis testing). The C.I. for the parameter  $\theta$  is given by:

$$\hat{\theta} \pm t_{df, \alpha/2} \sqrt{v(\hat{\theta})} \quad (5)$$

The width of the C.I. is larger because the variance is estimated. The more precisely the variance is estimated, the smaller the C.I. The amount of the increase is the ratio of the t-statistic to a similar Z- statistic that would be used if the variance were known:

$$\frac{t_{df, \alpha/2}}{Z_{\alpha/2}} \quad (6)$$

Table 1 below shows this ratio for various df. Clearly, the ratio decreases as the df increases. Thus, it is desirable for the variance estimate to have an adequate number of degrees of freedom. A rough indication of the df for a replicated estimate is  $\sum n_h - L$ , the number of PSUs minus the number of strata. Thus, for a 2 PSU per stratum design, the df will be approximately the number of replicates. However, in some situations this can be misleading and the number of replicates will be an overstatement of the true df (see Section 6 for an example). The table indicates that there is very little penalty in increased C.I. width with df of 30 or more, with virtually no penalty for 60 or more df.

If the necessary information is available or can be estimated, the df can be approximated using the following formula:

$$df = \frac{2Var(\hat{\theta})^2}{Var(v(\hat{\theta}))} \quad (7)$$

where  $Var$  is variance and  $v(\hat{\theta})$  is the estimated variance.

When dealing with a stratified design and estimating population totals, the denominator of this expression is approximately:

$$Var(v(\hat{\theta})) = \sum_h \frac{N_h^4 S_h^4}{n_h^3} \left( \beta_h - \frac{n_h - 3}{n_h - 1} \right) \quad (8)$$

where  $\beta_h$  is the kurtosis and  $S_h$  is the standard deviation of the PSUs in stratum h.

Table 1. The increase in width of a 95 percent C.I. due to df of variance estimate

df	t	t/z
1	12.71	6.48
3	3.18	1.62
5	2.57	1.31
10	2.23	1.14
15	2.13	1.09
20	2.09	1.06
25	2.06	1.05
30	2.04	1.04
35	2.03	1.04
40	2.02	1.03
50	2.01	1.02
60	2.00	1.02
70	1.99	1.02
80	1.99	1.01
90	1.99	1.01
100	1.98	1.01
Infinity	1.96	1.00

When  $\beta_h = 3$  and all the  $n_h$  are large, then the  $df \approx \sum n_h - L$ ; when the sample allocation is different, the possible number of df may be much less than  $\sum n_h - L$ .

### 3.3 Combining PSUs for Variance Estimation

If considerably more than 200 PSUs have been sampled, then to use either the jackknife or BRR methods as described above, more than 100 replicates will be needed.

If BRR is used directly with no combining the number of replicates needed is the next multiple of four greater than the number of strata (e.g., with 50 strata, 52 replicates are needed). If JK2 is used, the number of replicates equal the number of strata. If JK $n$  is used, the number of replicates equals the number of PSUs.

The computational effort of repeating the estimation process a large number of times, especially for iterative analyses such as logistic regression, may be unacceptable. What is more, it may be unnecessary, as discussed below. Of course, with the cost of computing significantly less than in the past, this is much less of a concern than it was when computing equipment and computing time were more costly. To reduce the computational effort, however, it is possible to combine PSUs without significant reduction in the df for variance estimates while reducing the number of replicates to a more computationally acceptable number. The goal is to reduce the computational requirements needed for a particular sample design by combining PSUs while retaining an unbiased and consistent estimate of variance.

Excessive and careless combining of PSUs can lead to inconsistent estimates. Valliant (1990) and Rao and Shao (1996) provide guidance on this issue. The discussion below includes practical examples and advice on combining PSUs.

The approaches to combining PSUs are outlined here. For the JK1 method, create  $G$  groups of PSUs, so that each group is likely to provide a good estimate of the population characteristic. For the BRR, JK2, and JK $n$  methods, within each stratum, create  $G_h$  groups (where  $G_h = 2$  for BRR and JK2), so each group is likely to provide a good estimate of the stratum population characteristic.

#### **DESIGN 3.3.1: A large number of PSUs is sampled as an SRS. Defining each PSU as a VarUnit in JK1 would result in more replicates than desired.**

In this situation, use the JK1 replication method. Determine the number of replicates that are desired, say  $G$ ; and combine PSUs into  $G$  VarUnits so that each VarUnit is approximately a random subset of the full sample.

**Example 3.3.1: A simple random sample of 500 patients is selected from a hospital, but only 50 replicates are desired.**

In this example, use the JK1 method and create a PSU variable by randomly assigning the values from 1 to 50 so that 10 patients are assigned to each PSU. Use this PSU variable as the VarUnit (see Appendix Example 3.3.1).

**DESIGN 3.3.2: A large number of PSUs is sampled systematically without explicit stratification.**

This situation is similar to DESIGN 2.1, except that a large number of PSUs has been sampled. In this case, the approach recommended earlier will work. In addition, however, we provide an alternative.

A solution using the JK2 method will also work. Denote the number of replicates desired as  $G$ . Combine the PSUs into  $G$  pseudo-strata. In combining the PSUs, take into account any systematic sampling. Although there were no actual strata, define pseudo-strata as units that are adjacent to each other in the systematic list, and assign them to the same VarStrat.

**Example 3.3.2: A sample of 400 businesses is sampled systematically with probability proportional to the square root of the number of employees. Each of the 400 businesses is a PSU. Form about 100 replicates.**

In this example, use the JK2 or BRR replication methods and list the 400 businesses in the order of sampling. Assign businesses 1-4 to VarStrat 1, identify two as VarUnit 1 and two as VarUnit 2. Assign businesses 5-8 to VarStrat 2, identify two as VarUnit 1 and two as VarUnit 2. Assign businesses in this way until businesses 397-400 are assigned to VarStrat 100 (see Appendix Example 3.3.2).

**DESIGN 3.3.3: In a stratified sample, the number of PSUs sampled in a few of the strata is so large that defining each PSU as a VarUnit would result in more replicates than desired.**

In this design, use the JK $n$  replication method; however, combine the PSUs into VarUnits within the strata using the approaches outlined above.

**Example 3.3.3:** There are 800 establishments sampled from 10 strata with probability proportional to size (PPS) where the size measure is total annual sales. VarStrat identifies the 10 sampling strata. In the first 8 VarStrat, the number of PSUs (establishments) sampled is relatively small. In the last 2 VarStrat many PSUs are selected (in stratum 9, 200 are sampled; in stratum 10, 500 are sampled).

For computational speed, no more than 200 replicates are desired. The PSUs in VarStrat 9 and 10 must be combined. In VarStrat = 9, list the 200 sampled establishments in the order of sampling and assign 1, 51, 101, and 151 to have VarUnit = 1. Assign 2, 52, 102, and 152 to have VarUnit = 2, etc.

Do the same thing for the 500 establishments sampled from VarStrat 10. This gives about 50 degrees of freedom for VarStrat 9 and 10 and the maximum number for the others (see Appendix Example 3.3.3).

### 3.4 Combining Strata for Variance Estimation

As with large numbers of PSUs, if considerably more than 100 strata have been used in the sample design, then to use either the jackknife or BRR methods, more than 100 replicates will be needed. Just as PSUs can be combined to reduce the computational effort, so too can strata be combined. Combining strata is distinct from collapsing, a method that is required to handle designs in which a single PSU is sampled from a stratum. Collapsing is discussed in Section 3.5.

As with combining PSUs, the goal is to reduce the computational effort, without introducing bias while retaining a large number of degrees of freedom. Also, as before, do not use data from the sample to make decisions about combining strata as this can lead to biases, generally under-estimates of sampling errors. Combining strata is referred to as “partial balancing” in the BRR design. The choice of strata to combine should follow the principle that the combined strata should provide a good estimate of the population of those strata.

An overly simplified example may help clarify the meaning of combining strata. Say there is a sample design involving four regions of a country (northeast, northwest, southeast and southwest). These regions are to be treated as strata with two PSUs selected from each. A direct JK2 solution would involve four replicates, each of which excluded one of the two PSUs in a stratum. By combining the four strata into two, an unbiased (though more variable) estimate of sampling error can be obtained using a total of only two replicates.

A combined estimate with a single degree of freedom can be formed as follows. A VarStrat variable will be created with the value of 1 for any PSU sampled from either the northeast or the southwest regions (strata). The VarUnit variable is still the PSU number (1 or 2) within the sampled stratum. A single replicate will be formed in which a PSU from both the northeast and

the southwest is excluded. The same can be done for the other two strata (northwest and southeast) to yield the second of the two combined strata replicates. This procedure estimates the variance contributed by sampling from both strata in an unbiased way; although, it will yield a more variable variance estimate than the direct method suggested in the previous paragraph.

If the two strata are not combined, then two replicates would be needed and an estimate with approximately two degrees of freedom would result. In one replicate one of the PSUs from the northeast stratum would be excluded and the other weighted up to represent the stratum total. In the second replicate the same procedure applies to the two PSUs from the southwest stratum.

By way of contrast, a collapsed stratum estimate that is biased can be produced by creating the same VarStrat variable with a value of 1 for PSUs in both the northeast or southwest but a VarUnit variable that is numbered from 1 to 4 for the four PSUs sampled in these two strata. In this situation, JK<sub>n</sub> would be used and four replicates would be produced in each collapsed stratum; however, the resulting estimate overstates the sampling error of the design because it creates a between-stratum component not inherent in the sample design.

**Example 3.4: There are 400 PSUs sampled from 200 strata, with 2 PSUs per stratum.**

If the 200 strata were assigned as VarStrat, there would be 200 replicates with JK2 or BRR. To improve the computational speed, reduce the number of replicates to  $G = 80$  by combining the 200 strata into 80 VarStrat. Use the BRR or JK2 replication method. List the 200 strata so that adjacent strata is as similar as possible. Assign strata 1, 81, and 161 to VarStrat = 1. Assign strata 2, 82, and 162 to VarStrat = 2. Continue until strata 80 and 160 are assigned to have VarStrat = 80. Let VarUnit be the PSU variable with values of 1 and 2 that were assigned before the strata were combined (see Appendix Example 3.4).

### 3.5 Collapsing Strata for Variance Estimation

When a single PSU is sampled from a stratum, there is no design unbiased way of estimating sampling variability within that stratum. The traditional method for handling this situation is to collapse any strata with single sampling units with another stratum. When this approach is taken, however, the ‘collapsed strata’ method is likely to overstate the variance of the actual, 1 PSU per stratum design. Using frame data, collapse so that PSUs in collapsed strata are as similar as possible.

**Example 3.5: A sample of 200 farms is sampled with one farm selected per stratum.**

Use the BRR or JK2 method. Arrange 200 strata so that they are as similar as possible with respect to the key estimates to be produced (region, type of farm, size, etc.). Collapse strata 1 and

2 into VarStrat = 1; 3 and 4 into VarStrat = 2; etc. Randomly assign each farm within a VarStrat so that VarUnit = 1 for one farm and VarUnit = 2 for the other.

## 4. MULTI-STAGE SAMPLE DESIGNS

Most procedures previously discussed for single-stage sample designs apply to multi-stage samples. However, care is needed in special circumstances, for example, when there are self-representing units, or when the finite population correction factor (fpc) is non-negligible. These special circumstances are discussed in Section 4.1 and Section 5. Section 4.1 discusses the finite population correction factor and other issues in multi-stage sampling. Section 4.2 describes how to create replicate weights for two types of designs: (1) exactly two PSUs sampled in each stratum, and (2) more than two PSUs per stratum.

### 4.1 Issues in Multi-stage Samples

Generally, use of the finite population correction factor is inappropriate and unnecessary with multi-stage sampling. When the sampling fractions at the first stage of sampling are small, e.g. less than 20 percent, it is generally not necessary to incorporate the fpc when calculating the variance. The fpc may not be appropriate as a means of adjusting the variance estimate if sampling at the first stage is done with unequal probabilities of selection.

The bias that results from ignoring the fpc in multi-stage sampling is (Wolter, 1985, eqn. 24.16):

$$Bias \doteq 2Var(\hat{\theta}_{WR}) - Var(\hat{\theta}_{WOR}) \quad (9)$$

where  $Var(\hat{\theta}_{WR})$  is the variance with replacement, and  $Var(\hat{\theta}_{WOR})$  is the variance without replacement. The bias occurs only in the between-PSU component estimate of the variance. If an overall fpc is applied to the with-replacement variance estimate, it will affect both the between and within components, so the fpc is generally not appropriate.

The procedures for combining PSUs and combining strata outlined previously for single-stage samples in Sections 3.3 and 3.4 also apply to multi-stage samples. PSUs should be combined in such a way that the new combined PSU provides an unbiased estimate of the population or stratum. Similarly, combined strata should provide an unbiased estimate of the population of PSUs in the set of strata chosen to be combined. Do not combine strata to such an extent that combined units contribute a large fraction to the variance.

#### **DESIGN 4.1: One PSU sampled from some strata.**

No unbiased estimate of the variance is possible when only 1 PSU per stratum is sampled. The collapsed stratum method is often appropriate in this situation. In this method, strata with only 1 sampled PSU are combined with other strata that are similar. The replication method most suitable to the collapsed strata is then applied.

**Example 4.1: A multi-stage sample of 1 PSU per stratum is selected. The strata are sorted into some useful order.**

This sample design can be handled by reducing the sample to a stratified sample with 2 PSUs selected per stratum, then using the BRR or JK2 methods. First, pairs of sampling strata that are adjacent in the sorted list are collapsed. The collapsed strata are labeled as the VarStrat. Each VarStrat should have two sampled PSUs that should be randomly classified as either VarUnit = 1 or 2 (see Appendix Example 4.1).

#### **4.2 Procedures for Some Sample Designs**

##### **DESIGN 4.2: Two PSUs sampled from each stratum.**

In this situation, either the BRR method (including Fay's variation) or the JK2 method may be used. Both of these methods require that there be exactly two VarUnits in each VarStrat. The first-stage sampling strata are defined as the VarStrat and the PSUs are defined as the VarUnit by randomly assigning them values of 1 or 2 within each stratum.

**Example 4.2: Schools are stratified by the number of students. Two schools are sampled PPS in each size class stratum. Classes within schools and students within classes are subsampled.**

The school size class variable is the VarStrat, because schools are stratified by size. VarUnit is the variable that identifies the sampled school within VarStrat (see Appendix Example 4.2).

##### **DESIGN 4.3: An even number of PSUs is sampled from some strata.**

In this design, either the BRR (including Fay's variation), JK2 or JKn method may be used. The first-stage strata are defined as the VarStrat. For the BRR and JK2, group the PSUs to form two VarUnits within each VarStrat. The VarUnit variable is assigned by first grouping all the even-numbered PSUs in the VarStrat into one group and all odd-numbered PSUs into a second group. Each group of PSUs is then randomly assigned a value of 1 or 2 within each VarStrat.

Note that this method is not appropriate if many PSUs are sampled in a small number of strata because it will result in too few degrees of freedom for the variance estimates. For JK<sub>n</sub>, simply set VarUnit equal to the PSU identifier. If this results in too many replicates, group the PSUs first before assigning VarUnit.

**Example 4.3: One stratum has 6 sampled PSUs.**

The BRR (including Fay's variation), JK2, or JK<sub>n</sub> method may be used. To create VarUnit for the BRR or JK2 method, randomly assign 3 of the PSUs to VarUnit = 1 and the other 3 to VarUnit = 2. If the PSUs are sampled in systematic order, (i.e., they are sorted within a stratum, then sampled using systematic sampling), then assign the first, third, and fifth PSUs to VarUnit=1 and the other three to VarUnit=2 (see Appendix Example 4.3). To use JK<sub>n</sub>, assign each PSU to its own VarUnit. This will result in the creation of 6 replicates for this stratum.

**Variant 4.3.1: Two PSUs are sampled in most strata, but 3 PSUs are sampled in some strata.**

Assuming the number of strata is large and no stratum contributes a large fraction to the variance of the estimates, a simple approach is to randomly combine two of the PSUs into 1 to fit the standard model for BRR or JK2 (see Appendix Variant 4.3.1).

**Example 4.3.1: A sample has 100 strata. Strata 1 to 99 have 2 PSUs sampled. Stratum 100 has 3 sampled PSUs, labeled *a*, *b*, and *c*.**

To use either BRR or JK2, randomly choose 2 PSUs in stratum 100, and label them as *a* and the other PSU as *b*. Use the stratum variable as the VarStrat, and the newly created PSU variable as the VarUnit.

If this procedure results in sizeable overestimates of the variance, then an alternative is to create two or more replicates for strata with three PSUs. A simple way to do this is to use JK<sub>n</sub>. In JK<sub>n</sub>, one replicate is created per VarUnit so that two replicates are created for each stratum with two PSUs, and three replicates are created for strata with three PSUs.

Another alternative is to create two replicates for strata with three PSUs. In the first replicate, two of the three PSUs (say *a* and *b*) are weighted by 1.5 and *c* is 0-weighted. In the second replicate, a different group of two PSUs (say *a* and *c*) are weighted by 1.5 and *b* is 0-weighted.

#### **DESIGN 4.4: An odd number of PSUs are sampled in some strata.**

When an odd number of PSUs (three or more) is sampled in some strata, the JK<sub>n</sub> method should be used. The number of replicates resulting will be equal to the number of PSUs sampled. If this number is too large, the number of replicates can be reduced by grouping the PSUs before assigning VarUnit.

**Example 4.4.1:** There are 60 schools sampled in 3 strata. The stratum sample sizes are 15, 20, and 25, and the sampling fractions are small. Teachers are sampled within schools.

In this situation, JK<sub>n</sub> may be used with VarStrat set equal to the school stratum and VarUnit set equal to the variable identifying the school. Sixty JK<sub>n</sub> replicate weights will be created for each school and each teacher by dropping out one school at a time.

**Example 4.4.2:** There are 1,000 schools sampled in 8 strata. Schools are sorted by enrollment within each stratum prior to sampling. The stratum PSU sample sizes vary from 100 to 255 and the school sampling fractions are small. Students are subsampled within schools.

If JK<sub>n</sub> were used here with VarStrat assigned to the school stratum and VarUnit assigned to the variable identifying the school, 1,000 replicates would result. To reduce the number of replicates, group the schools prior to assigning VarUnit. For example, if 100 replicates are desired, then group the schools into 100 VarUnits. To do this, sort the schools within each stratum in the order they were selected. Then, number the schools from 1 to 100 repeatedly, beginning over again with 1 until the end of the stratum is reached. In each stratum, set VarUnit equal to the sequence number assigned to the school. Since the sequence numbers range from 1 to 100, this will result in 100 VarUnits and 100 replicate weights for each school and student. In a stratum containing 255 sampled schools, this is equivalent to assigning schools 1, 101, 201 to VarUnit = 1; schools 2, 102, 202 to VarUnit = 3; schools 3, 103, 203 to VarUnit = 3, and so on.

## **5. SELF-REPRESENTING UNITS**

In a single-stage sample, certainty or self-representing (SR) units do not contribute to the sampling error. They can easily be handled by setting the replicate weight equal to the full sample weight. If the sample is multi-stage with subsampling, treat each SR unit like a stratum, and define replicates using the second stage units. If there are many second-stage units, use methods to combine them. If the SR unit is larger than the average stratum, assign more replicates to it by creating pseudo-strata within it.

### **5.1 Single-Stage Designs**

**DESIGN 5.2:** The design is a single-stage stratified sample with a take-all (SR) stratum.

Set the replicate weights equal to the full sample weight for units from the SR stratum. The SR stratum will then contribute nothing to the estimated variances. In the remaining strata, use BRR, JK2, or JKn to create replicates, depending on the number of PSUs in the strata.

**Example 5.2:** A single-stage sample of universities in a state is selected. State universities are selected with certainty and others are stratified by size with two sampled in each of 10 strata. The stratum variable takes on the value of 1 for state universities and 2 to 11 corresponding to the 10 size strata. Within the 10 size strata, universities are selected with equal probabilities.

The JK2 method can be used since there are two universities sampled in each of the 10 strata. Each of the sampled universities in VarStrat 2 to 11 has a variable with the VarUnit equal to 1 or 2 that is randomly assigned. In VarStrat 1, set all the replicate weights equal to the full sample weight. If the sampling fraction is large in any of the remaining VarStrat, include the fpc.

### **5.2 Multi-Stage Designs**

**DESIGN 5.3:** One or more PSUs is sampled with certainty (SR) in a multi-stage sample.

A multi-stage sample of PSUs has some SR PSUs. In all the SR PSUs, second-stage units (SSUs) are subsampled. Any SR PSU is treated as a separate variance stratum (VarStrat). SSUs are treated as sampled PSUs and are the VarUnits.

If there are no SSUs in some of the SR PSUs, these SR PSUs do not contribute to the sampling error of estimates. To insure that their responses are included in the estimate but do not

contribute to the sampling error estimate, their replicate weights should be equal to the full-sample weight.

**Example 5.3.1:** A multi-stage sample of 2 area PSUs per stratum is selected, except that in one stratum a single SR PSU is selected. This SR PSU is divided into blocks and 2 blocks are sampled.

Either the BRR or JK2 methods can be used, since 2 blocks are sampled in the SR PSU and 2 PSUs are sampled in the remaining strata. Label the SR PSU as a VarStrat and label each of the blocks in the SR PSU as the VarUnits (see Appendix Example 5.3.1). In the remaining strata, assign VarStrat to the stratum and VarUnit to the PSU identifier.

**Variant 5.3.2: The number of second-stage units sampled within the SR PSU is large.**

First, determine the number of VarStrat to be assigned to each SR PSU. If the SR PSU is approximately the same size as the other VarStrat, then assign it a single VarStrat. If it is much larger, assign a larger number of VarStrat to it. After creating the desired number of VarStrat, treat the SSU within the SR PSU as VarUnits. If this results in too many replicates, then combine the SSU. Note that if a stratified sample were drawn within the SR PSU with  $m$  strata, then  $m$  replicates would be created from that SR PSU unless strata are combined.

**Example 5.3.2: City blocks are sampled systematically, without stratification, in a large city that is a SR PSU. Assume that the city is about four times the population of the average VarStrat.**

In this example, it is assumed that JK $n$  will result in too many replicates. Therefore, either the BRR or JK2 method will be used by creating VarStrat and VarUnit as follows. Assign 4 VarStrat to the SR PSU by arranging blocks in the order sampled and labeling the first one-fourth of the blocks to VarStrat = 1, the second fourth to VarStrat = 2, etc. Within VarStrat, randomly assign half the blocks to VarUnit = 1 and half the blocks to VarUnit = 2 (see Appendix Example 5.3.2). Alternatively, VarUnit could be assigned by sorting the blocks within each VarStrat in a serpentine order such as 1,2,2,1,1,2,2,1. It would not be appropriate to assign the VarUnit as 1,2,1,2 because this could create a bias. For example, if the blocks were in order by size, VarUnit 1 would tend to contain smaller blocks than VarUnit 2.

## **6. CREATING REPLICATES FOR THE NHIS**

### **6.1 Introduction to the NHIS**

Using the methods discussed in Section 2 through 5, this section discusses the creation of replicates for variance estimation in the National Health Interview Survey (NHIS), a large-scale household survey of the civilian, noninstitutionalized population of the United States. The NHIS began in 1957 and has undergone modifications to its content and sampling design approximately every 10 to 15 years. It consists of a core questionnaire and several specialized supplemental questionnaires. Since the NHIS provides estimates of important characteristics about the health of the United States population, reliable measures of precision for those estimates are needed.

The NHIS employs a complex sample design that includes stratification and several stages of sampling. Sampling weights are adjusted for nonresponse and poststratified to known population totals. The complex sample design and estimation procedures complicate variance estimation. Additionally, confidentiality issues result in suppressing full information on stratum and PSU identifiers in public release files. Because of these concerns, the National Center for Health Statistics (NCHS) offers simplifying assumptions for NHIS variance estimation that have slightly masked stratum and PSU variables.

The documentation accompanying the 1995 NHIS public release file discusses two methods for estimating variances. The first method treats the NHIS sample as a two PSU per stratum design and is appropriate for replication software. The second method incorporates more information about the sample design; however, as presented leads to a very large number of replicates. Both methods allow the user to specify the masked stratum and PSU variables for each record on the public release file so that any software appropriate for variance estimation from a complex design can be used.

This section presents a third alternative to the two strategies documented in the 1995 NHIS documentation using methods discussed in the previous sections.

### **6.2 Overview of the NHIS Sample Design and Estimation**

The NHIS sample design is multi-stage, with stratification and clustering at several stages. The universe is partitioned into approximately 1,900 PSUs, with many large PSUs included in the sample with certainty (SR PSUs). The remaining PSUs are stratified by geography and PSU characteristics within state. Two PSUs in each of the nonself-representing (NSR) strata are selected with probability proportional to the population in the PSU. SSUs are formed within each selected PSU and stratified on black and Hispanic population concentration. SSUs are sampled at different rates, and within-sampled SSUs households may be subsampled. Persons within sampled households may be subsampled as well.

The NHIS weights account for the various stages of sampling and are adjusted for nonresponse and poststratified to age/sex/race ethnicity totals. Additional weight adjustments were made to the 1995 data to compensate for data lost as a result of the 3-week government shutdown at the end of that year.

### 6.3 Variance Estimation for the NHIS

Two methods of estimating variances are discussed in the documentation that accompanies the 1995 public use data file. These methods were developed originally to enable users to compute variances using replication or linearization approaches without having extensive knowledge of the sample design and estimation procedures of the NHIS. Although they were prepared for an earlier NHIS design, they were adapted to the new design. The rationale and procedures are given by Parsons and Casady (1986) and Parsons, Chan, and Curtin (1990). These two methods are described first, then an alternative method is presented.

Method 1 serves for replication software that requires exactly 2 PSUs in each stratum. The sample is partitioned into 187 variance strata and each variance stratum contains exactly 2 pseudo-PSUs. The documentation indicates that the strata and PSUs are based on the exact sample design as much as possible for the NSR strata where 2 PSUs are sampled per stratum (although collapsing is required when only 1 PSU was sampled in a stratum). SR PSUs are treated as strata and all the SSUs within a SR PSU are identified as being in 1 of 2 pseudo-PSUs.

This first procedure results in 187 variance strata, each with 2 PSUs. If BRR is used, then a fully balanced design requires 188 replicates and replicate weights. For ease of comparison with the other methods, we use the stratified paired-jackknife replication method (JK2), which requires 187 replicates and replicate weights. The actual methods used for this evaluation are slightly more complex than discussed here. Nixon et al. (1998) gives full details.

Method 2 is the alternative suggested for use with linearization software. This method handles NSR strata almost the same as Method 1, with 2 PSUs for each NSR strata and collapsing as necessary for strata in which only 1 PSU is sampled per stratum. The main difference is in handling the SR PSUs. In Method 2, all the SSUs in a SR unit are treated as PSUs rather than pairing them into pseudo-PSUs as is done in Method 1.

Method 2 is considered to be more statistically efficient because it incorporates more design information for the SR PSUs. However, the method results in 4,079 PSUs in the SR PSUs and 259 PSUs in the NSR strata. Although Method 2 should produce more stable variance estimates than Method 1, Method 2 is computationally intensive if replication methods are used because it would result in thousands of replicates and replicate weights. Although Method 2 is more stable than Method 1, it would be grossly inaccurate to suppose that the number of df for variance estimates computed using Method 2 could be estimated by the difference between the number of PSUs and the number of strata (Section 3.2). The df under this and other methods is discussed below.

This section describes a third alternative, denoted Method 3, that takes greater advantage of the design information in the SR PSUs than Method 1 without requiring the large number of replicates needed by Method 2. The main objective in designing the alternative is to produce stable and reliable variance estimates for national and domain estimates while reducing the number of replicates to a reasonable level. This is accomplished by using Method 1 design information for the NSR strata, Method 2 design information for the SR PSUs, and combining units to reduce the number of replicates as described in Section 3.3 and 5.2. The strategy for combining units takes into account the important analysis domains and the stability of the variance estimates.

Distributions from the NHIS can be used to assess how many replicates are needed. The first columns of Table 2 show the number of Method 2 PSUs and the estimated population totals for each region and type of PSU (SR or NSR). Because NSR strata are less metropolitan (an important analysis domain) and larger in total population than individual SR PSUs, the amount of collapsing of these strata is limited. The maximum number of variance stratum for the NSR strata in regions 1, 2, and 4 are 14, 38, and 18, respectively. By not combining these NSR strata, the number of df from the NSR strata is large, but there are only 70 ( $14 + 38 + 18$ ) replicates between the three strata. To keep the total number of replicates from all four NSR regions to 70, the NSR strata in region 3 must be combined with NSR strata from the other regions, but this still maximizes the df for region 3 estimates.

Table 2. Method 3 Assignment of 70 VarStrat for the NHIS

Region	Type	Number of PSUs	Population percent	Region percent	Number of VarStrat	Percent in VarStrat	Assigned VarStrat
1	NSR	29	3.5%	17.9%	14	20.0%	01 to 14
	SR	994	16.1%	82.1%	56	80.0%	15 to 70
	Total		19.6%				
2	NSR	76	10.2%	42.9%	38	54.3%	15 to 52
	SR	766	13.6%	57.1%	32	45.7%	01 to 14/53 to 70
	Total		23.8%				
3	NSR	117	16.7%	47.5%	58	82.9%	01 to 58
	SR	1,210	18.5%	52.5%	70	*	59 to 70,01 to 59
	Total		35.3%				
4	NSR	37	5.2%	24.3%	18	25.7%	53 to 70
	SR	1,109	16.1%	75.7%	52	74.3%	01 to 52
	Total		21.3%				
Total	NSR	259	35.6%				
	SR	4,079	64.4%				
	Total	4,338	100.0%				

\*In this category, the SSUs are first assigned to VarStrat 59 to 70 until they are roughly equal in size to the NSR VarStrat, then they are distributed across all the VarStrat.

The SR PSUs are combined by noting that in the NHIS SSUs are sampled systematically. A reasonable first approximation is to pair consecutive SSUs and treat these as paired selections from a stratum within the SR PSUs. Since the subsampling is done using race/ethnicity data, this pairing should capture some of that feature of the design. Since this would result in a large number of variance strata, units are combined.

To increase the df for regional estimates, the combining is done within variance strata in a region not assigned to the NSR strata. For example, in region 1 the NSR units are assigned to variance strata 1 to 14 and the SR units are assigned to variance strata 15 to 70. As shown in Table 2, this results in approximately proportional allocation of the population to each variance stratum within a region.

The one exception is region 3 where the NSR units are assigned to variance strata 1-58 and this accounts for 83 percent of the 70 variance strata but only 48 percent of the population in the region. The variance estimates are more efficient if the population in the variance strata is closer to proportional, so some of the SR units were combined with the NSR units in this region. Figure 1 depicts the overall assignment of the units to the 70-variance strata for Method 3.

Figure 1. Assignment of VarStrat by region and type

Reg	Type	Variance Stratum (VarStrat)																																					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35			
1	NSR	1	2	3	4	5	6	7	8	9	10	11	12	13	14																								
1	SR																																						
2	NSR																																						
2	SR																																						
3	NSR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35			
3	SR																																						
4	NSR																																						
4	SR																																						
Reg	Type	Variance Stratum (VarStrat)																																					
		36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70			
1	NSR																																						
1	SR																																						
2	NSR	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52																					
2	SR																																						
3	NSR	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58															
3	SR																																						
4	NSR																		53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70			
4	SR																																						

Some of the advantages expected from the Method 3 approach to forming variance strata and replicates are: (1) only 70 replicates and replicate weights are used, (2) the variances should be more consistent than Method 1 because of improved handling of SR PSUs, and (3) the df of freedom for domains such as region, metropolitan status, and race/ethnicity should be relatively large.

#### 6.4 Properties of the Variance Estimators

The main criterion remaining for evaluating the three estimators is their variance. The full jackknife variance estimator reproduces the textbook variance estimate for linear statistics, and is used as the benchmark for comparing the three methods. Following the development in Rust (1986), the variance of the full jackknife variance estimator is given by equation (8). Combining strata decreases the precision of the estimator, as can be seen from the expression for the combined strata jackknife

$$V(v_{CJ}(\hat{\theta})) = \sum_{h=1}^H \frac{W_h^4 S_h^4 (\beta_h - 3)}{n_h^3} + 2 \sum_{g=1}^G \left( \frac{\sum_{h \in g} W_h^2 S_h^2}{l_g - 1} \right)^2 \quad (10)$$

where  $g$  denotes a combined stratum consisting of  $H_g$  original strata;  $W_h$  is the proportion of the population  $n$  in stratum  $h$ ;  $l_g$  is the number of PSUs in combined stratum  $g$  (in this example, it is always 2) and the other values were defined earlier.

This expression can be used to evaluate the precision of a variance estimator if the values of  $W_h$ ,  $S_h^2$ , and  $\beta_h$  are known. Generally, this is not the case and consistent sample estimates are substituted. Kish (1965) shows  $\beta$  is approximately 3 for large clusters in sampling human populations such as used in the NHIS. We assume  $\beta_h = 3$ . The value for  $W_h$  is estimated from the survey for each strata for each domain. The value of  $S_h^2$  is more difficult to estimate precisely. We estimated a design effect separately for the SR and NSR strata and assumed it was constant across all strata within these large classes. The value of  $S_h^2$  was then computed as a constant times the estimated design effect where the constant represents  $p(1-p)/n_h$  where  $p = .5$  and  $n_h = 2$ .

#### 6.5 Comparison of the Methods

To compare the three methods, estimated degrees of freedom were computed for each method. First, variance of the variance estimates were computed using formula (8) for the full jackknife and formula (10) for the combined strata jackknife. Estimates of the df for each method

were produced using these estimated variances and the ratio shown in formula (7). Methods 1 and 3 required the combined jackknife expression while Method 2 used the exact jackknife expression for variances.

As mentioned earlier, Method 1 while suitable for use with a replication approach does not yield a large number of df for some subnational estimates. Method 2 which does provide an adequate number of df for subnational estimates requires too many replicates to be a practical solution. The method we have proposed requires only 70 replicates, and has almost as many df for national estimates as the NCHS recommended Method 1 while providing adequate degrees of freedom for important subnational statistics. The df for a number of subnational statistics are given for all three methods in Table 3. Method 2 results are included for completeness, but comparisons are made only between Methods 1 and 3 as they are practical to implement in replication.

Table 3. Degrees of freedom for domains under three methods

Domain	Method 1	Method 2	Method 3
National	72	501	66
Region 1	7	169	61
Region 2	24	122	57
Region 3	51	140	59
Region 4	12	113	56
Poverty	53	154	52
Black	26	163	55
Hispanic	10	53	32
NonBlack	71	465	66
MSA	48	481	63
NonMSA	60	60	42

As the table reveals, Method 3 provides only 66 df for national estimates, contrasted with 72 df for Method 1 and 500 for Method 2. This difference in df between 1 and 3 should not cause a large increase in the variability of variance estimates. In almost all of the sub-national estimates, however, Method 3 provides more df than Method 1 often by a significant margin. Note the differences in df for several regional estimates (1 and 4) in which a fourfold increase in df result. Estimates for racial and ethnic sub-groups will have two or three times the df under Method 3.

Finally, note that Method 3 constructed using the principles of combining strata has an adequate df (i.e., >30) for all of the subnational domains listed in the table.

## 7. Summary

The paper began with a short description of the two predominant methods for estimating sampling errors for complex estimates based on complex sampling designs: linearization and replication. Section 1 reviewed some of the technical background behind the replication method and discussed a number of its strengths. Sections 2 through 5 discussed common sampling designs and provided guidance for adapting each design to a replication approach to variance estimation.

In addition to the straightforward 2 PSUs per stratum design, for which replication procedures are discussed in Wolter (1985), replication approaches were described for a number of more complex but frequently used designs. Among these were designs with: more than 2 PSUs, systematic sampling, and certainty PSUs. Sections 3 and 5 discussed methods for combining PSUs and combining strata to reduce the number of replicates required while still producing an asymptotically unbiased estimate of variance. The important topics of the role of the finite population correction and the effect of degrees of freedom were also examined in Section 3.

Finally, a practical example using the ongoing U.S. NHIS was presented. A replication approach which requires only 70 replicates was presented that was able to produce not only reasonably precise national estimates of sampling errors but subnational estimates, as well.

## REFERENCES

- Binder, D.A. (1996) Linearization methods for single phase and two-phase samples: a cookbook approach, *Survey Methodology*, **22**: 17-22.
- Brick, J.M., Kalton, G. (1996) Handling missing data in survey research, *Statistical Methods in Medical Research*; **5**: 215-38.
- Graubard, B.I., Korn, E.L. (1996) Survey inference for subpopulations, *American Journal of Epidemiology*, **144**: 102-106.
- Hinkins, S., Moriarity, C., Scheuren, F. (1996) Replicate variance estimation in stratified sampling with permanent random numbers, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Judkins, D. (1990) Fay's method for variance estimation, *Journal of Official Statistics* **6**, 223-240.
- Kish, L. (1965) *Survey Sampling*, New York: Wiley.
- Kish, L., Frankel, M. (1974) Inference from complex samples, *Journal of the Royal Statistical Society; B36*, 1-22.
- Kovar, J.G., Rao, J.N.K., Wu, C.F.J. (1988) Bootstrap and other methods to measure errors in survey estimates, *Canadian Journal of Statistics*; **16**: 25-46.
- Krewski, D., Rao, J.N.K. (1981) Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods, *Annals of Statistics*; **9**: 1010-19.
- McCarthy, P.J. (1966) Replication: an approach to the analysis of data from complex surveys, *Vital and Health Statistics, Series 2, No. 14*, National Center for Health Statistics, Public Health Service, Washington, D.C.
- Nixon, M., Brick, J.M., Kalton, G., Lee, H., Givens, J. (1998) Alternative variance estimation methods for the NHIS, to be published in *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Parsons, V.L., Casady, R.J. (1986) Variance estimation and the redesigned National Health Interview Survey, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Parsons, V.L., Chan, J., and Curtin, L.R. (1990) Analytic limitations to current National Health Surveys, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Rao, J.N.K., Shao, J. (1996) On balanced half-sample variance estimation in stratified random sampling, *Journal of the American Statistical Association*, **91**, 343-348.

- Rao J.N.K., Wu, C.F.J., Yue, K. (1992) Some recent work in resampling methods, *Survey Methodology*; **18**: 209-17.
- Rust, K. (1986) Efficient replicated variance estimation, *Proceedings of Survey Research Methods Section of the American Statistical Association*, 81-87.
- Rust, K.F., Rao, J.N.K. (1996) Variance estimation for complex surveys using replication techniques, *Survey Methods in Medical Research*; **5**: 283-310.
- Shah, B.V., Barnwell, B.G., Hunt, P.N., LaVange, L.M. (1992) *SUDAAN user's manual*, Release 6.0, Research Triangle Park, NC: Research Triangle Institute.
- Valliant, R. (1996) Limitation of balanced half-sampling, *Journal of Official Statistics*, 12, 225-240.
- Valliant, R. (1990) Comparisons of variance estimators in stratified random and systematic sampling, *Journal of Official Statistics*; **6**: 115-31.
- Westat, Inc. (1997) *A user's guide to WesVarPC*, Version 2.1, Rockville, MD: Westat, Inc.
- Westat, Inc. (1998) *WesVar complex samples 3.0 user's guide*, Chicago, IL: SPSS, Inc.
- Wolter, K.M. (1985) *Introduction to Variance Estimation*, New York: Springer.
- Yung, W. (1996) Contributions to poststratification in stratified multi-stage samples, PhD thesis, Carleton University.

# **APPENDIX**

Example 2.1 Creating 50 replicates from a sample of 50 patients (PSUs)

PSU	1	VarUnit	1
PSU	2	VarUnit	2
PSU	3	VarUnit	3
PSU	4	VarUnit	4
PSU	5	VarUnit	5
PSU	6	VarUnit	6
PSU	7	VarUnit	7
PSU	8	VarUnit	8
PSU	9	VarUnit	9
PSU	10	VarUnit	10
PSU	11	VarUnit	11
PSU	12	VarUnit	12
			<b>USE JK1</b>
PSU	13	VarUnit	13
PSU	14	VarUnit	14
PSU	15	VarUnit	15
	.	.	.
	.	.	.
	.	.	.
	.	.	.
PSU	45	VarUnit	45
PSU	46	VarUnit	46
PSU	47	VarUnit	47
PSU	48	VarUnit	48
PSU	49	VarUnit	49
PSU	50	VarUnit	50

Example 2.2. Stratified sample of schools, two sampled per strata from 50 strata

	SAMPLED PSUs
STRATUM 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
STRATUM 2	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
	PSU 11
	PSU 12
	PSU 13
STRATUM H	PSU 1
	PSU 2
	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

**USE BRR OR JK2**

Example 2.3. 100 establishments sampled from no more than 50 strata,  
each stratum having at least 2 PSUs sampled

SAMPLED PSUs		
STRATUM 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6 X
	PSU	7
	PSU	8 X
	PSU	9
	PSU	10
.		.
.		.
.		.
PSU		34
PSU		35
PSU		36 X
PSU		37
STRATUM 2	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	.	
	.	
	.	
	PSU	29
	PSU	30 X
PSU		31
.		.
.		.
.		.
STRATUM H	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X
	.	
	.	
	.	
	PSU	39
	PSU	40 X
PSU		41
PSU		42

USE JKn

Example 3.1. 50 establishments sampled from four strata

SAMPLED PSUs	
STRATUM 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6 X
	PSU 7
	PSU 8 X
	PSU 9
	PSU 10
	PSU 11 X
	PSU 12
	PSU 13 X
STRATUM 2	PSU 14
	PSU 15
	PSU 16 X
	PSU 17
	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
STRATUM 3	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
	PSU 11
	·
	·
	·
	PSU 1
	PSU 2
STRATUM 4	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12

Variant 3.1.1 Two PSU per strata design but significant fpc

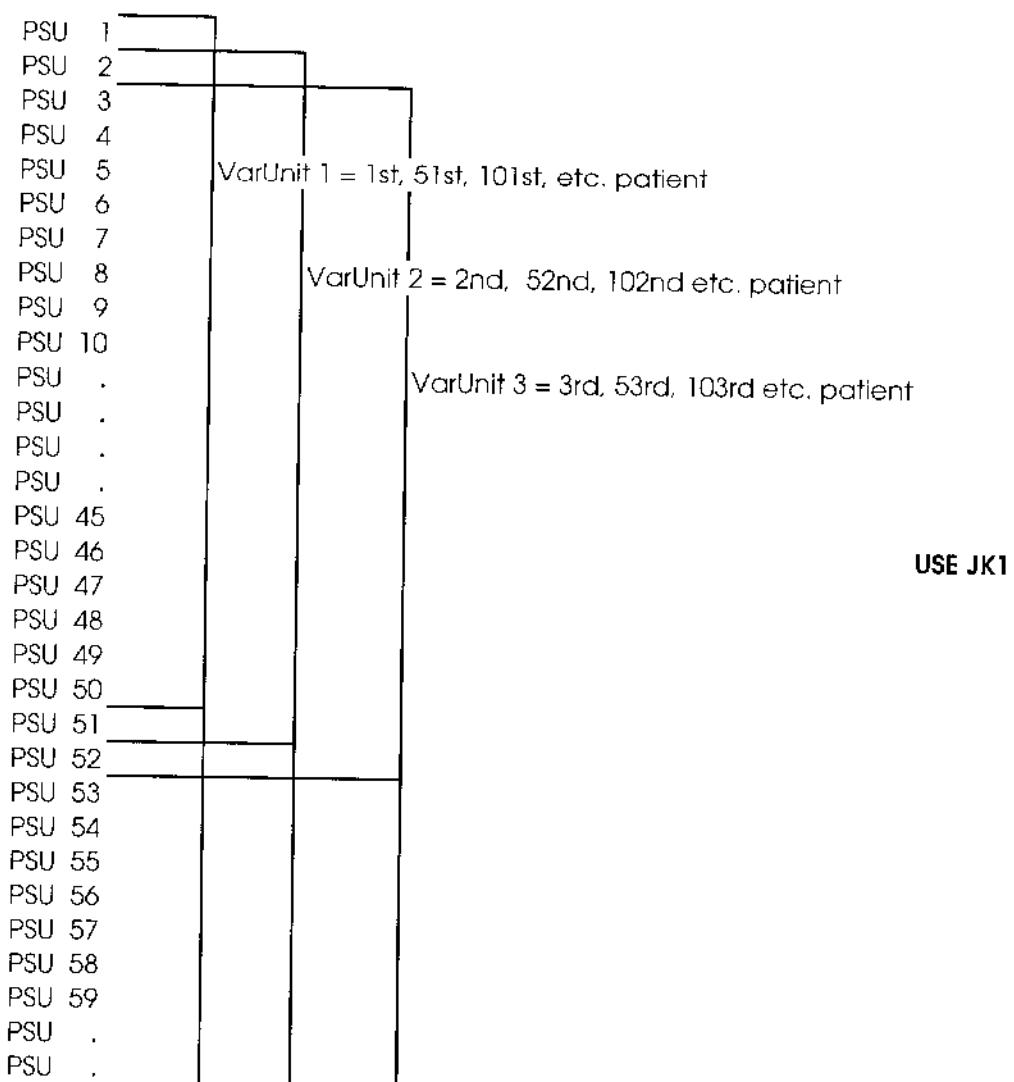
	SAMPLED PSUS	USE JK2 AND ATTACH FACTORS TO STRATA WITH SIGNIFICANT fpc	
STRATUM 1	PSU 1		
	PSU 2		
	PSU 3 X		
	PSU 4		
	PSU 5		
	PSU 6		
	PSU 7 X		
	PSU 8		
	PSU 9		
	PSU 10		
STRATUM 2	PSU 1	USE JK2 AND ATTACH FACTORS TO STRATA WITH SIGNIFICANT fpc	
	PSU 2		
	PSU 3 X		
	PSU 4		
	PSU 5		
	PSU 6		
	PSU 7 X		
	PSU 8		
⋮			
⋮			
STRATUM H	PSU 1	USE JK2 AND ATTACH FACTORS TO STRATA WITH SIGNIFICANT fpc	
	PSU 2		
	PSU 3		
	PSU 4		
	PSU 5 X		
	PSU 6		
	PSU 7		
	PSU 8		
	PSU 9		
	PSU 10 X		
	PSU 11		
	PSU 12		
	PSU 13		

Example 3.1.1. A single-stage stratified sample of high schools taken from 50 size strata, 2 PSUs taken per stratum. The fpc is significant in some strata.

		SAMPLED PSUs
STRATUM 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6 X
STRATUM 2	PSU	1 X
	PSU	2
	PSU	3 X
	PSU	4
⋮		
⋮		
⋮		
STRATUM 50	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X

**USE BRR OR  
JK2 WITH fpc  
WHERE APPROPRIATE**

Example 3.3.1 Creating 50 replicates from a sample of 500 patients (PSUs)



Example 3.3.2. Creating 100 replicates from a systematic sample of 400 businesses

		SAMPLED PSUs	VarUnit
PSEUDO STRATUM 1	PSU	1	
	PSU	2	
	PSU	3 X	1
	PSU	4	
	PSU	5 X	2
	PSU	6	
	PSU	7 X	2
	PSU	8	
	PSU	9	
	PSU	10 X	1
PSEUDO STRATUM 2	PSU	11 X	1
	PSU	12	
	PSU	13 X	2
	PSU	14	USE BRR OR JK2
	PSU	15 X	
	PSU	16	
	PSU	17 X	
	PSU	18	
PSEUDO STRATUM 10C	PSU	4851	
	PSU	4852 X	2
	PSU	4853	
	PSU	4854	
	PSU	4855 X	1
	PSU	4856	
	PSU	4857	
	PSU	4858 X	1
	PSU	4859	
	PSU	4860 X	2
PSU	4861		
	4862		
	4863		

Example 3.3.3. 800 establishments sampled from 10 strata.  
A large sample taken from the 9th & 10th strata

		SAMPLED PSUs
STRATUM 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
	PSU	9
	PSU	10 X
	PSU	11
	PSU	12
	PSU	13

		USE JK <sub>n</sub>
(200 PSUs SAMPLED)	PSU	1
	PSU	2
	PSU	3
	PSU	4 X
	PSU	5
	PSU	6 X
	PSU	.
	PSU	X
	PSU	.
	PSU	X
	PSU	501
	PSU	502 X
	PSU	503
	PSU	504 X
(500 PSUs SAMPLED)	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5 X
	PSU	.
	PSU	X
	PSU	.
	PSU	X
	PSU	2101
	PSU	2102 X
	PSU	2103 X
	PSU	2104

Example 3.4. Combining strata to reduce the number of replicates

	SAMPLED PSUs	VarUnit
	PSU 1	
	PSU 2	
STRATUM 1	PSU 3 X	1
	PSU 4	
	PSU 5	
	PSU 6	
	PSU 7 X	2
	PSU 8	
	PSU 9	
	PSU 1	
	PSU 2	
STRATUM 2	PSU 3 X	1
	PSU 4	
	PSU 5	
	PSU 6	
	PSU 7 X	2
	PSU 8	
	PSU 9	
	.	
	.	
	PSU 1	
	PSU . X	
STRATUM 81	PSU .	
	PSU . X	
	PSU 13	
	PSU 1	
	PSU . X	
STRATUM 161	PSU .	1
	PSU . X	2
	PSU 13	
	.	
	.	
	PSU 1	
	PSU 2	
	PSU 3	
	PSU 4	
	PSU 5 X	2
STRATUM 200	PSU 6	
	PSU 7	
	PSU 8	
	PSU 9	
	PSU 10 X	1
	PSU 11	

USE BRR  
OR JK2

Example 4.1. Combining strata with a single PSU sampled

		SAMPLED PSUs	VarUnit
COMBINE THESE STRATA BECAUSE THEY ARE SIMILAR	PSU	1	
	PSU	2	
	STRATUM 1	PSU 3 X	1
		PSU 4	
		PSU 5	
		PSU 6	
		PSU 7	
		PSU 8	
		PSU 9	
		PSU 10	
		PSU 1	
STRATUM 2	PSU	2	2
	PSU	3 X	
	PSU	4	
	PSU	5	
	PSU	6	
	PSU	7	
	PSU	8	
	PSU	9	
	PSU	10	
	PSU	11	
	PSU	12	
	PSU	13	
	.		
STRATUM H	PSU	1	1
	PSU	2	
	PSU	3	
	PSU	4	
	PSU	5 X	
	PSU	6	
	PSU	7	
	PSU	8	
	PSU	9	
	PSU	10 X	
	PSU	11	
	PSU	12	
	PSU	13	

Example 4.2 Two schools (PSUs) are randomly sampled in each size class stratum

SAMPLED PSUs	
STRATUM 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
STRATUM 2	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
	PSU 11
	PSU 12
	PSU 13
.	
.	
.	
STRATUM H	PSU 1
	PSU 2
	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

Example 4.3. An even number, more than two, schools are sampled in a stratum

		SAMPLED PSUs	
STRATUM 1	PSU	1	
	PSU	2	
	PSU	3 X	
	PSU	4	COMBINE
	PSU	5 X	
	PSU	6	
	PSU	7 X	COMBINE
	PSU	8	
	PSU	9 X	
	PSU	10	
STRATUM 2	PSU	11 X	
	PSU	12 X	
	PSU	13	
	PSU	1	
	PSU	2	
	PSU	3 X	USE BRR, JK2 OR JKn
	PSU	4	
	PSU	5	
	PSU	6	
	PSU	7 X	
STRATUM H	PSU	8	
	PSU	9	
	PSU	10 X	
	PSU	11	
	PSU	12	
	PSU	13	
	PSU	1	
	PSU	2	
	PSU	3	
	PSU	4	

Variant 4.3.1. An odd number of units are sampled in some strata

PSUs	
STRATUM 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9 X
	PSU 10
RANDOMLY COMBINE	
USE BRR OR JK2	
STRATUM 2	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
	PSU 11
	PSU 12
	PSU 13
STRATUM H	PSU 1
	PSU 2
	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

Example 5.3.1. Multi-stage sample, 2 PSUs per stratum in all but one stratum with 2 SSUs in the SR PSU

	PSU	1 X	SSU	1
			SSU	2 X VarUnit 1
STRATUM 1			SSU	3
			SSU	4
			SSU	5
			SSU	6
			SSU	7
			SSU	8
			SSU	9
			SSU	10
			SSU	11 X VarUnit 2
			SSU	12
			SSU	13
			SSU	14
			SSU	15
			SSU	16
			SSU	17
			SSU	18
	PSU	1		
	PSU	2		
STRATUM 2	PSU	3 X		
	PSU	4		
	PSU	5		
	PSU	6		
	PSU	7 X		
	PSU	8		
	PSU	9		
	PSU	10		
	PSU	11		
	PSU	12		
	PSU	13		
	:			
	:			
	PSU	1		
	PSU	2		
	PSU	3		
	PSU	4		
	PSU	5 X		
STRATUM H	PSU	6		
	PSU	7		
	PSU	8		
	PSU	9		
	PSU	10 X		
	PSU	11		
	PSU	12		
	PSU	13		

Example 5.3.2. Multi-stage sample, two PSUs per stratum in all, but one stratum with many SSUs in the SR PSU

	PSU	1 X	SSU	1		
STRATUM 1			SSU	2 X	1	VarUnit 1
			SSU	3		
			SSU	4		
			SSU	5 X	2	VarUnit 2
			SSU	6		
			SSU	7		
			SSU	8 X	2	
			SSU	9		
			SSU	10		
			SSU	11 X	1	
			SSU	12		
			SSU	13 X	1	
			SSU	14		
			SSU	15		
			SSU	16 X	2	
			SSU	17		
			SSU	18		
STRATUM 2	PSU	1				
	PSU	2				
	PSU	3 X				USE BRR OR JK2
	PSU	4				COMBINE SSUs
	PSU	5				
	PSU	6				
	PSU	7 X				
	PSU	8				
	PSU	9				
	PSU	10				
	PSU	11				
	PSU	12				
	PSU	13				
STRATUM H	PSU	1				
	PSU	2				
	PSU	3				
	PSU	4				
	PSU	5 X				
	PSU	6				
	PSU	7				
	PSU	8				
	PSU	9				
	PSU	10 X				
	PSU	11				
	PSU	12				
	PSU	13				

## **EL MÉTODO DE REPLICACIÓN PARA LA ESTIMACIÓN DE ERRORES DE MUESTREO: LA CREACIÓN DE RÉPLICAS**

David Morganstein, J. Michael Brick, Pam Broene, y Mary G. Nixon  
Westat, Inc.

La replicación es un método muy útil para aproximarse a la varianza de los datos estadísticos no lineales en encuestas muestrales complejas. Proporciona resultados asintóticamente equivalentes al método alternativo de linealización, y presenta ventajas significativas en algunas de las situaciones más comunes que surgen en la práctica. En este estudio, explicamos cómo se puede implantar la teoría de replicación a los diferentes diseños de encuestas. Emplear la replicación nos brinda la oportunidad de alcanzar algunos objetivos importantes que con los métodos de linealización son intrínsecamente difíciles. Entre las ventajas de la replicación se encuentran: la facilidad con que se puede explicar la replicación a los usuarios de datos sin que tengan una formación especial en la estimación de varianza de encuestas muestrales; se puede aplicar a casi todos los datos estadísticos mediante el mismo procedimiento y sin modificar el software; puede manejar diferentes estrategias de estimación que incluyen diferentes tipos de ajustes de las ponderaciones; es un procedimiento constante para trabajar con subconjuntos de datos y datos ausentes; es un método simple y elegante para gestionar diseños de muestras longitudinales, y ofrece características de diagnóstico que resultan útiles para descubrir los problemas de ponderación y planificar las encuestas nuevas.

En el Apartado 1 hacemos una breve introducción a los diferentes métodos de estimación de errores de muestreo de las estimaciones de encuesta: métodos exactos; linealización y varios métodos de replicación. Las referencias incluidas indican que estos métodos aproximados son asintóticamente equivalentes. Por lo tanto, una elección entre métodos aproximados depende de otras características, tales como la facilidad de uso y de comprensión. El Apartado 1 incluye una revisión de la teoría de la replicación, y propone una lista de varias ventajas que no se aplican al método de linealización.

Los Apartados 2, 3, 4 y 5 explican los diseños comunes de muestreo e indican el modo de crear réplicas. El Apartado 2 abarca los diseños muestrales mono-etapicos, tanto con como sin la estratificación. El Apartado 3 describe un método de aproximación para la estimación de varianzas si sólo se selecciona una única unidad de muestreo. Asimismo, trata de varios temas importantes relacionados con el muestreo de la población finita, tales como la utilización de la corrección de población finita y el grado de libertad para estimar las varianzas. Se presenta el concepto de combinar las dos unidades de muestreo primario (UMP) y los estratos con el fin de reducir el esfuerzo de cálculo para diseños que de otra forma tendrían varias réplicas. El Apartado 4 profundiza en los temas del Apartado 2 para abarcar las contrapartidas de etapa múltiple. La creación de réplicas cuando existen unidades auto-representantes (certeza) se explica en el Apartado 5.

## 1. INTRODUCCION

Se han desarrollado varios métodos de aproximación con el fin de estimar los errores de muestreo de los datos estadísticos complejos provenientes de encuestas donde los diseños muestrales incluyen etapas múltiples, estratificación, conglomerados y probabilidades de selección desiguales. Dado que la ponderación de encuestas está sujeta a correcciones de no respuesta y baja respuesta, sería conveniente que el método del error de muestreo reflejara el efecto de estas correcciones de ponderación.

El Apartado 1.1 habla de los métodos más utilizados para estimar los errores de muestreo. El Apartado 1.2 proporciona el trasfondo estadístico que se encuentran tras el método de replicación. Una relación de las ventajas de la replicación se presenta en el Apartado 1.3, especialmente la capacidad de reflejar casi todas las correcciones de ponderación.

### 1.1 Métodos Alternativos

Cuando se realiza un encuesta que utiliza un diseño muestral complejo, el diseño debe tenerse en cuenta con el fin de producir estimaciones no sesgadas y errores standard de las estimaciones. La mayor parte del software estadístico standard como SPSS y SAS® System ignoran el diseño muestral y suponen que las observaciones son independientes y distribuidas de forma homogénea. Utilizando las características de ponderación, el software standard puede producir estimaciones puntuales no sesgadas, pero no es posible estimar los errores standard de forma correcta. De hecho, si el diseño muestral o el estimador son complejos, la expresión exacta de la varianza del estimador a menudo resulta insoluble y se requieren métodos de aproximación.

Una forma de aproximar la varianza de una estimación es utilizar los métodos de replicación que incluyen el jackknife (con las variaciones denominadas JK1, JK2 y JK<sub>n</sub>), la replicación de equilibrio repetido (RER), y métodos bootstrap. La elección de los métodos depende en gran parte de la naturaleza del diseño muestral, y esto se verá con más detalle más adelante. Sea cual sea la elección, cada enfoque requiere la división de la muestra completa en submuestras G, no necesariamente mutuamente exclusivas, conocidas como réplicas, para que cada muestra replicada, si se pondera de forma adecuada, proporcione una estimación de la característica de interés de la población. Principalmente, en este estudio trataremos las opciones jackknife y RER para ilustrar los puntos principales.

En el método JK1, apropiado para diseños sin estratificar, las réplicas se forman mediante la exclusión de una UMP (o grupos de UMP combinadas) de una submuestra replicada y ajustando las ponderaciones de las UMP restantes de forma que lleguen a la ponderación de la población. En general, las réplicas se forman de manera que reflejen el muestreo original de las UMP. (Históricamente, uno de los métodos originales de replicación, el método del grupo aleatorio, esencialmente es el método jackknife JK1 que se ha descrito aquí).

Se utilizan varios métodos para los diseños estratificados, incluyendo los JK2, JK<sub>n</sub>, y RER. Dos de estos métodos, JK2 y RER, son apropiados para los diseños que contienen exactamente dos UMP en cada estrato. En el RER tradicional, las réplicas se forman mediante la exclusión de una de las dos UMP muestreadas de cada estrato. Por lo tanto, sólo se utilizan la mitad de las UMP en la estimación de cada réplica. Si existen H estratos, entonces existen  $2^H$  posibles muestras de réplica RER. McCarthy (1.966) demuestra que un número mucho más reducido de "media-muestras" bien seleccionadas contiene la misma información que el conjunto completo de posibles muestras  $2^H$ . Este conjunto mínimo se llama una media-muestra equilibrada y puede construirse utilizando modelos ortogonales. (WesVar utiliza un conjunto de matrices de Hadamard ortogonales para crear media-muestras equilibradas cuando se solicitan réplicas RER). En una variación del RER de Fay (Judkins, 1996), las dos UMP de cada estrato están incluidas pero se les aplican ponderaciones diferentes. A una se le aplica una ponderación entre 0 y 1, por ejemplo k, mientras que a la otra UMP se le aplica una ponderación de 2-k. En todos los ejemplos que continúan donde el RER puede ser aplicado, el método Fay puede utilizarse.

En el JK2, sólo se excluye una UMP único de un estrato único para formar cada réplica. Un conjunto completo de réplicas K2 para un diseño de dos UMP por estrato requiere H réplicas, cada una corresponde a submuestras de réplicas que incluyen una UMP de un estrato. En cierto modo, la réplica JK2 única estima la varianza que forma un estrato simple.

El JK<sub>n</sub> se utiliza con diseños que incluyen más de dos UMP en algunos estratos. Es similar al planteamiento del JK2, pero requiere más réplicas. Si un estrato de muestreo contiene  $n_h$  UMP muestreadas, entonces  $n_h$  réplicas se forman mediante la omisión de cada una de las UMP muestreadas. Por lo tanto, en un estrato con dos UMP, el método JK<sub>n</sub> originaría dos réplicas mientras que el método JK2 nos proporcionaría sólo una réplica.

Un método alternativo es la linealización del estimador utilizando series de expansión de Taylor y posteriormente la utilización de métodos de estimación de varianza de encuesta standard para calcular la precisión del dato estadístico linealizado. Tanto el método de linealización como el de replicación se describen con más detalle en Wolter (1.985).

Muchos estudios teóricos y empíricos comparan los métodos de replicación y linealización para la estimación de varianza. En el aspecto teórico, los primeros trabajos de Krewski y Rao (1.981) muestran la equivalencia asintótica de estos métodos. Los estudios teóricos posteriores, que incluye el de Rao, Wu y Yue (1.992), concluyen que los métodos de linealización y replicación son asintóticamente equivalentes a un nivel razonablemente alto. En el sentido empírico, se han realizado estudios de simulación para comparar la precisión de las estimaciones que utilizan ambos métodos. Uno de los primeros estudios fue el de Kish y Frankel (1.974). Trabajos más recientes de Kovar, Rao y Wu (1.988), Rao, Wu y Yue (1.992) y Valliant (1.990) demuestran que los métodos por lo general originan estimaciones constantes. Las propiedades del

jackknife son muy similares a las de la linealización, mientras que las propiedades de los métodos RER y bootstrap se parecen mucho entre sí. La mayor parte de las comparaciones tienen en cuenta tanto la exactitud de las estimaciones de varianza como los intervalos de confianza derivados. Rust y Rao (1.996) repasan algunas de estas publicaciones y examinan otros aspectos de los dos métodos de aproximación de varianzas de las estimaciones de encuestas muestrales complejas.

En este estudio, revisamos las circunstancias bajo las cuales el método de replicación presenta ventajas significativas con respecto al método de linealización de estimación de varianzas en encuestas muestrales. Dado que ambos métodos proporcionan estimaciones que son asintóticamente equivalentes, los criterios de evaluación de las ventajas del método de replicación son factores diferentes al de la precisión de las estimaciones. Algunas de las características examinadas son el diseño muestral y las estrategias de estimación que pueden reflejarse más fácilmente con la replicación que con la linealización. Por supuesto, en algunas situaciones la aplicación de la linealización resulta más sencilla y más fácil, pero la mayoría de estas situaciones se analizan en otras publicaciones (Shah et al., 1.992).

Las ventajas de un método dependen enormemente del software disponible para llevar a cabo el cálculo. El paquete de software WesVar Complex Samples, basado en Windows para calcular las varianzas mediante la utilización de los métodos de replicación, se emplea para ilustrar las ventajas del método de replicación en las aplicaciones prácticas. El WesVar Complex Samples está diseñado para calcular las funciones de los totales ponderados y aplicar modelos a los datos de muestras complejas mediante la utilización de una variedad de técnicas de replicación (Westat, 1.998). El WesVar Complex Samples (versión 3) puede adquirirse de SPSS, Inc. (<http://www.spss.com>). (Puede disponerse de una versión anterior y menos potente, el WesVarPc, de forma gratuita del World Wide Web URL de Westat <http://www.westat.com>. La versión anterior y gratuita y su documentación puede bajarse de esta página).

## 1.2 Revisión de la Teoría de Replicación

La idea fundamental de la replicación es seleccionar submuestras repetidamente de la muestra completa, para calcular el dato de interés de cada submuestra, y posteriormente utilizar la variabilidad con estos datos estadísticos de submuestra o réplica con el fin de hacer la estimación de la varianza de los datos estadísticos de una muestra completa. Existen diferentes modos de crear submuestras de la muestra completa. Las submuestras se llaman réplicas y los datos estadísticos calculados de las réplicas se llaman estimaciones replicadas.

Estimadores de varianza replicada,  $v(\hat{\theta})$ , toman la forma

$$v(\hat{\theta}) = c \sum_{k=1}^G h_k (\hat{\theta}_{(k)} - \hat{\theta})^2 \quad (1)$$

donde

- $\theta$  es un parámetro arbitrario de interés,
- $\hat{\theta}$  es la estimación de  $\theta$  basada en la muestra completa,
- $\hat{\theta}_{(k)}$  es la estimación  $k$ -th de  $\theta$  basada en las observaciones incluidas en la réplica  $k$ -th,
- $G$  es el número de réplicas,
- $c$  es una constante que depende del método de replicación, y
- $h_k$  es una constante específico de estrato que sólo se requiere en ciertos diseños.

El valor  $c$  depende del método de replicación; para RER y bootstrap  $c = 1/G$ ; para el jackknife no estratificado (JK1),  $c = (G-1)/G$ , para el jackknife emparejado (JK2) y JKn,  $c = 1$ . El factor  $h_k$  se requiere para diseños jackknife con un número desigual de UMP por estrato (JKn) y puede utilizarse además para ajustar las correcciones de población finita de algunos diseños. Este tema no aparece más adelante en este estudio. Rust y Rao (1996) describen los métodos de replicación de forma más detallada.

WesVar es un software que se ha diseñado específicamente para acomodar los métodos de replicación. Los Apartados 2, 3, 4 y 5 describen cómo varios métodos de replicación pueden ser utilizados con el fin de desarrollar réplicas para diseños específicos de muestra. Diseñar las réplicas es una característica crítica con implicaciones importantes para la validez de las estimaciones de varianza. Para algunos diseños, la preparación de réplicas requiere mucho ingenio para poder producir estimaciones precisas de varianza sin la necesidad de cómputos extensivos. Para muchos diseños esto resulta muy sencillo.

### 1.3 Algunas Ventajas de la Replicación

#### Facilidad de Uso

La elección del método de estimación de varianza que se va utilizar puede ser examinada de diferentes maneras. Una dimensión importante es la capacidad de describir a los usuarios cómo se producen las estimaciones de varianzas sin la necesidad de formación avanzada en la estimación de varianzas en las encuestas muestrales. La replicación es un método muy conocido en el mundo de la estadística y muchos usuarios aceptan en primera instancia la premisa básica de la replicación. La división en submuestras de la muestra completa y la estimación de la variabilidad de submuestras replicadas en torno a la estimación de la muestra completa a menudo resulta ser la única explicación requerida.

Este aspecto de facilidad de uso se refuerza por el hecho de que se sigue el mismo procedimiento para casi todos los datos estadísticos: medias, totales, razones, y funciones más

complejas de los datos muestrales. Por ejemplo, la razón log-odds de una tabla de dos factores de una muestra puede expresarse así:

$$\hat{\theta} = \log\left(\frac{\hat{N}_{11}\hat{N}_{22}}{\hat{N}_{12}\hat{N}_{21}}\right) \quad (2)$$

donde la ‘^’ por encima de las cantidades indica que son estimaciones muestrales consistentes y aproximadamente no sesgadas de las características de la población. La varianza de la razón log-odds de la muestra puede ser calculada del mismo modo que la media muestral o total utilizando (1).

Esta facilidad contrasta con el método de linealización que requiere que se linealice cada dato estadístico no lineal antes de calcular la varianza. La forma linealizada del dato estadístico debe ser elaborada e incorporada al software. Mientras que la derivación de la forma linealizada a menudo no resulta difícil (Binder, 1996), el software de linealización software debe incluir ya el dato estadístico específico en su lista de datos estadísticos o debe aproximarse numéricamente al dato. Ambas opciones imponen limitaciones al usuario de datos.

Una ventaja muy importante de la replicación es que permite al usuario definir casi cualquier dato estadístico y calcular su varianza utilizando (1) sin la necesidad de modificar el software. WesVar hace esto de dos maneras. Primero, el usuario puede especificar un dato “Computado” como una función de los variables del grupo de datos, tal y como se ha explicado antes. Por ejemplo, algunos datos “Computados” que son fácilmente definidos en WesVar son ratios de ratios, el log de ratios, y el cuadrado de una diferencia de ratios. Un interface gráfico fácil de usar se utiliza para dar entrada al dato y el programa computa (1) para ese dato. Segundo, el usuario puede especificar un dato “Función” como una función de cantidades estimadas en las células de una tabla. Por ejemplo, el ejemplo del ratio log-odds ratio mencionado anteriormente puede calcularse con WesVar nombrando las cuatro células de la tabla de dos factores y especificando el log del ratio de los totales calculados en las células. Datos especificados por el usuario, tales como diferencias, contrastes, y logs de diferencias o ratios pueden ser calculados de esta manera. Una vez más, la sencillez es posible porque se calcula la varianza utilizando (1) para casi todo tipo de datos estadísticos.

### **El Manejo de Diferentes Métodos de Estimación**

En la mayoría de las encuestas muestrales, no basta sólo ponderar los datos observados por el reciproco de la probabilidad de la selección. Se emplean diferentes estrategias de estimación para reducir el sesgo y/o la varianza de las estimaciones. Por ejemplo, casi todas las encuestas están sujetas a la no-respuesta y las ponderaciones se ajustan para compensar esta no-respuesta (Brick y Kalton, 1996). Asimismo, las estrategias de estimación como la postestratificación, raking, o la de estimación de regresión generalizada se utilizan a menudo para aprovecharse de los datos auxiliares y reducir tanto el sesgo como las varianzas de las estimaciones. La replicación

ofrece un método sencillo y elegante de incorporar estos ajustes en las estimaciones de varianza (Yung, 1996).

Rust y Rao (1996) muestran que una manera sencilla de crear las estimaciones replicadas requeridas en (1) es crear ponderaciones replicadas y adjuntarlas a cada observación al igual que se adjunta normalmente la ponderación de la muestra completa a una observación. Las estimaciones replicadas pueden ser computadas utilizando las ponderaciones replicadas igual que las ponderaciones de muestra completa se utilizan para crear la estimación de muestra completa. Además destacan que, si la agencia responsable de la recogida de datos incluye las ponderaciones replicadas con los datos, proporciona tres ventajas importantes: primero, el usuario no necesita saber las características del diseño muestral para estimar las varianzas de muestra (se incluye toda la información en las ponderaciones replicadas); segundo, la agencia puede reflejar todas las etapas de estimación apropiadamente en las ponderaciones replicadas con el fin de reflejar el modelo completo de estimación utilizado en las ponderaciones de muestra completa; y, tercero, no es necesario modificar el software para la estimación de la varianza replicada para calcular las estimaciones de forma adecuada (se sigue empleando la ecuación (1)).

El software WesVar permite a los usuarios dar entrada a ponderaciones replicadas que han sido ajustadas para la no-respuesta, estimación de ratio, postestratificación, raking, estimación de calibración, o cualquier método de estimación utilizado. Por lo tanto, el usuario medio puede aprovecharse de los procedimientos avanzados de estimación empleados por la agencia que recoge los datos para reducir el sesgo y la varianza de las estimaciones, siempre que la agencia introduzca estas ponderaciones replicadas en el fichero de datos. Por ejemplo, si se utilizan los ajustes de no-respuesta en la creación de las ponderaciones definitivas, los mismos ajustes de no-respuesta pueden ser aplicados por separado a cada réplica (cada réplica tiene, por lo tanto, un ajuste de no-respuesta diferente) con el fin de reflejar este procedimiento en las estimaciones computadas. Además, WesVar proporciona una opción que permite al usuario crear ponderaciones replicadas con ajustes de postestratificación o de raking si los totales de control requeridos para estos procedimientos están disponibles y las ponderaciones replicadas todavía no incluyen los ajustes.

Esta sencillez contrasta con el método de linealización, donde cada estrategia de estimación requiere un desarrollo diferente. Por ejemplo, la postestratificación puede ser manejada en la teoría de linealización como un tipo de estimación de ratio, pero esto implica que el estimador es diferente para el estimador postestratificado y el tipo sencillo Horvitz-Thompson. Por lo tanto, el software de linealización requiere que el estimador específico linealizado postestratificado haya de incluirse en el software para calcular correctamente la varianza. Los ajustes de no-respuesta podrían ser manejados en el método de linealización, pero una vez más resulta ser una carga importante para el software que debe acomodar cada tipo de estimación. No conocemos un software de linealización que maneja los ajustes no-respuesta de manera sencilla. En la replicación, esto es muy sencillo, si la agencia que proporciona los datos incluye este componente

en las ponderaciones replicadas. El proceso es sencillo y debe llevarse a cabo sólo una vez cuando se crean las ponderaciones.

### **Gestionar los Subconjuntos de Datos y Datos Ausentes**

Como las ponderaciones replicadas llevan toda la información necesaria para estimar las varianzas, la replicación sirve muy bien para gestionar los datos de subconjunto o ausentes. Por ejemplo, si se selecciona una muestra de adultos, algunos analistas pueden querer estimar sólo las características de un subconjunto de los adultos, mujeres de entre 30 y 44 años por ejemplo. Este tipo de análisis es fácil de hacer con el método de replicación sin preocuparse de las implicaciones para la estimación de varianza.

Por ejemplo, en el software WesVar el subconjunto de datos puede ser identificado utilizando una sencilla frase de subconjunto y se crea un fichero que contiene sólo los registros de los adultos incluidos, junto con su ponderación de muestra completa y las ponderaciones replicadas. Se puede llevar a cabo un análisis detallado de sólo estos registros con el archivo subconjunto porque las ponderaciones replicadas contienen toda la información necesaria para estimar correctamente la precisión de las estimaciones.

Este proceso no es tan sencillo como en la mayor parte del software de linealización debido a que la información de diseño quizás ya no sea precisa si se utiliza únicamente el fichero de subconjunto. El problema básico es que el subconjunto puede no incluir las observaciones de todas las unidades de muestreo primario muestreadas y la mayor parte del software de linealización requiere que todas estas unidades se identifiquen en el fichero de datos empleado en el análisis. Graubard y Korn (1996) demuestran lo que podría ocurrir en esta situación e ilustran las consecuencias con dos ejemplos de encuestas reales. Una vez más, es viable manejar la situación con el método de linealización, pero los pasos necesarios son más complejos y más susceptibles al error que el método de replicación.

Otra circunstancia común en las encuestas muestrales, pero difícil de gestionar, son los datos ausentes. Algunos temas relacionados con la gestión de ajuste de no-respuesta de las ponderaciones que se emplean normalmente para corregir la no-respuesta han sido tratados anteriormente en este estudio. Ahora dirigimos nuestra atención a la no-respuesta de ítems y los datos ausentes de esta fuente.

Gestionar los datos ausentes en el análisis es un problema antiguo sin una solución totalmente aceptada y general. Es demasiado ambicioso esperar gestionar los datos ausentes en la estimación de varianza para satisfacer a todo el mundo cuando no se ha aceptado un método para gestionar las mismas estimaciones. No obstante, un objetivo más modesto sería gestionar los datos ausentes de forma constante para crear la estimación y estimar la varianza de las

estimaciones. En relación con esta cuestión, la replicación ofrece una ventaja considerable sobre los métodos de linealización.

Una vez que se ha determinado un procedimiento para la gestión de datos ausentes con el fin de crear una estimación, el mismo procedimiento puede ser aplicado para crear tanto la muestra completa como las estimaciones replicadas. WesVar lo hace; la muestra completa y las ponderaciones replicadas se emplean de la misma forma en relación con los datos ausentes para crear los componentes necesarios para el cálculo (1). Por ejemplo, al estimar un total con respuestas de datos ausentes, la estimación de la muestra completa y la estimación replicada pueden ser sesgadas a la baja debido a la no-respuesta (si no se realiza una imputación para este ítem), pero la varianza de la estimación se calcula correctamente para este procedimiento.

En el método de linealización, este tipo de consistencia es más difícil de lograr. De hecho, muchos de los paquetes de software de linealización requieren que todos los ítems tengan respuestas para poder calcular las varianzas. Otros paquetes de software de linealización permiten los datos ausentes, pero de una forma poco constante debido a que la varianza depende del diseño muestral y de la forma específica del estimador. Por ejemplo, el problema anterior con los subconjuntos puede ocurrir como resultado de los datos ausentes, pero puede ser muy difícil de detectar. Los métodos para gestionar los datos ausentes pueden hacer que las estimaciones de estratos se calculen de forma diferente a las estimaciones de la muestra completa.

Los procedimientos para manejar datos ausentes pueden tener consecuencias importantes. Por ejemplo, el software puede calcular las estimaciones de totales por el equivalente de imputar un valor cero para el dato ausente, mientras que la media puede ser calculada ignorando las observaciones con datos ausentes. El resultado es que diferentes datos estadísticos, medias y totales, no son tratados de igual forma. Un ejemplo muestra lo importante que esto puede resultar. Supongamos que hay 4 respondientes, 3 registrando valores de 99, 100, y 101, mientras que 1 tiene un valor ausente. En WesVar y los métodos de replicación, el total estimado de la muestra completa sería 300 y cada estimación replicada se acercaría entre 100 a 300 y por lo tanto la varianza estimada no sería muy elevada. Por otra parte, si el programa de linealización imputara un valor cero para el valor ausente, entonces la varianza estimada sería muy elevada (las 4 observaciones para calcular la varianza serían 99, 100, 101, y 0).

### **Manejar Diseños Longitudinales o Encuestas de Panel**

Las encuestas muestrales de muestra longitudinal son muy populares desde hace unas décadas, particularmente porque ofrecen la posibilidad de medir el cambio por el tiempo de forma más precisa que las encuestas independientes transversales. La correlación entre las unidades muestreadas por el tiempo produce un aumento en la precisión. No obstante, algunos de los beneficios de una estimación más precisa no pueden realizarse sin que estimaciones de precisión reflejen las respuestas correlacionadas.

Los métodos de replicación proporcionan un método bastante sencillo y elegante de incorporar la correlación. Las réplicas definidas para unidades muestreadas para la muestra línea base pueden ser utilizadas también para las recogidas de datos de seguimiento de forma que las estimaciones replicadas definidas en (1) contengan de manera natural la solapa apropiada y, como consecuencia, las estimaciones de precisión reflejen el diseño. Hinkins, et al. (1996) hablan de este tema con más detalle. Una vez más, WesVar puede ser utilizado sin modificación para producir estimaciones de precisión apropiadas para estimaciones, siempre que las ponderaciones replicadas para el diseño de seguimiento se produzcan correctamente.

Para el software de linealización, el problema es más complejo. Para algunas estimaciones es posible construir nuevas variables que son la diferencia entre los datos recogidos en las diferentes etapas de la encuesta, pero esto puede implicar una mayor carga sobre los usuarios de datos. En algunos casos, son necesarios procedimientos elaborados para resolver este problema.

### **El Diagnóstico**

Además de utilizarse en el cálculo de estimaciones de errores de muestreo para la estadística de encuestas, el conjunto de estimaciones replicadas juegan un papel diagnóstico. Los problemas de ponderación pueden ser detectados mediante un examen de las estimaciones replicadas y buscando un valor inusual, ‘aislado’. Los métodos JK2 y JK<sub>n</sub> proporcionan una situación donde las varianzas de estratos individuales pueden ser examinadas. El examen de las estimaciones de varianzas replicadas pueden ser de gran ayuda para valorar cómo los componentes de varianza se distribuyen por los estratos.

## **2. DISEÑOS MUESTRALES MONO-ETAPICOS**

En este apartado, tratamos los diseños mono-etapicos y algunos métodos para construir réplicas para estos diseños. Es importante notar que los métodos dados son útiles pero hay otros alternativos que también son válidos. Los diseños que contienen dos o más etapas se describen en el Apartado 4. El apartado comienza hablando de los diseños sin estratificación y después examina el tema de los diseños estratificados.

Los diseños no-estratificados no son muy corrientes en los EEUU. Quizás el diseño más típico de esta categoría es un muestra sistemática de una lista ordenada. El modelo JK1 es el recomendado para el diseño no-estratificado mono-etápico.

El tema de los diseños estratificados se subdivide por el número de UMP seleccionados en cada estrato. Cuando exactamente dos UMP son muestreadas de cada estrato, cualquier de los dos métodos, RER o JK2, puede utilizarse. Cuando se seleccionan más de dos UMP, se puede aplicar el método JK<sub>n</sub> directamente, siempre que el número de réplicas no sea tan elevado que acabe en

un esfuerzo excesivo de cálculo. Si el esfuerzo de cálculo que resulta del modelo JK<sub>n</sub> fuera demasiado grande, las UMP pueden ser combinadas para reducir el número de réplicas requeridas. La combinación de UMP para reducir el número de réplicas se describe en el Apartado 3.3.

En los restantes apartados se tratan dos términos. En cualquier diseño muestral hay unidades muestrales primarias, que podrían ser las únicas unidades de muestreo además de, por ejemplo, una muestra mono-etápica. En la mayoría de las muestras, los estratos se utilizan también. En muchos casos, las UMP muestreadas y los estratos muestrales se emplearán para crear réplicas. En otros casos, las UMP muestreadas y los estratos muestrales se combinarán o se agruparán (la diferencia entre los dos sistemas será explicada más adelante) con el fin de reducir el número de réplicas y el correspondiente esfuerzo de cálculo. Los conglomerados que se excluyen de las réplicas, que pueden ser UMP o UMP combinadas o agrupadas, serán clasificados como VarUnit. Los estratos que se utilizan en la estimación de varianza se clasificarán como VarStrat. (Estos términos son empleados también en el software WesVar.).

## 2.1 Muestreo sin Estratificación

Dos diseños mono-etápicos, sin estratificación que se emplean normalmente son una muestra aleatoria simple y una muestra sistemática, posiblemente con un marco de muestreo ordenado que emplea una o más variables ordenadas para lograr una estratificación implícita. En cualquiera de los casos, la solución recomendada es la misma, el método JK1.

### **DISEÑO 2.1: Diseños mono-etápicos donde las UMP son muestreadas sin estratificación.**

En este diseño, se utiliza el método de replicación JK1. Se define cada UMP como una VarUnit.

#### **Ejemplo 2.1: Una simple muestra aleatoria de 50 pacientes se selecciona de un hospital para una investigación. Los pacientes muestreados son identificados por una variable de identificación del paciente con valores de 1 a 50.**

En esta situación, se utiliza el JK1 y la variable de identificación del paciente ID como la VarUnit (véase el Anexo Ejemplo 2.1).

#### **Variación 2.1: Se toma una muestra sistemática de las UMP, posiblemente de una lista ordenada.**

Aunque esta variación es más común que el DISEÑO 2.1, no hay un método no sesgado para estimar los errores de muestreo de una muestra sistemática única. Este diseño puede

considerarse como la selección de un conglomerado único de unidades. Con sólo una selección única, las varianzas no pueden ser estimadas de forma no sesgada.

Un método común es dividir la muestra sistemática en un conjunto de submuestras mutuamente exclusivas y exhaustivas, intercaladas y sistemáticas. El error de muestreo se estima como la varianza entre estas menores muestras sistemáticas. Por ejemplo, si la muestra sistemática consta de 1.000 unidades, las réplicas pueden ser 50 submuestras sacadas sistemáticamente, cada una de 20 unidades. La primera réplica contiene las unidades 1, 51, 101, ... 951 (cada unidad 50<sup>a</sup> comenzando con la primera unidad muestreada). La segunda submuestra contiene las unidades 2, 52, 102, ... 952 (cada unidad 50<sup>a</sup> comenzando con la segunda unidad muestreada). El método JK1 se emplea con las 50 réplicas como la VarUnit.

## 2.2 Muestreo estratificado

El caso del muestreo estratificado se examina en dos partes. Comenzamos con la situación más sencilla, la de dos UMP muestreadas de cada estrato (DISEÑO 2.2). Este caso está tratado en detalle en la literatura especializada. Los métodos RER y JK2 se desarrollaron para la estimación de varianza con estos diseños.

A continuación examinamos el caso de más de dos, pero no un número elevado de UMP muestreados de uno o más estratos, pero no de un número elevado (DISEÑO 2.3). Con ‘no de un número elevado,’ queremos decir una situación donde el número de réplicas necesarias para una aplicación directa del método JK<sub>n</sub> requeriría un tiempo aceptable de cálculo.

Más adelante se describen otras variaciones de diseños estratificados. Cuando se muestrea un número elevado de UMP, o se utilizan un número elevado de estratos, quizás resulte deseable combinar las UMP o estratos para reducir el número de réplicas y, como consecuencia, el esfuerzo de cálculo. La combinación de UMP y estratos son explicados en los Apartados 3.3 y 3.4. El caso de una selección única aleatoria de un estrato, otro diseño en el cual las estimaciones de errores de muestreo no sesgados no son posibles, se ve en el Apartado 3.5. Se hace un examen de la estimación del error de muestreo cuando se selecciona una o más UMP de certeza en el Apartado 5.

### DISEÑO 2.2: Muestreo estratificado mono-etápico: Dos UMP son muestreadas de cada estrato.

En este diseño, se utiliza el método RER (o el de Fay) o el JK2. Se definen los estratos muestrales como el VarStrat y las UMP muestreadas definen la VarUnit. Se les asignan de forma aleatoria valores de 1 o 2 dentro de cada estrato. Para ser más breve, ya no mencionamos el método Fay; no obstante, es una variación del método RER y sirve siempre cuando se recomienda el método RER.

**Ejemplo 2.2: Los colegios públicos de un estado son estratificados por tamaño y el número de alumnos en el colegio. Se muestran dos colegios de forma aleatoria en cada estrato de tamaño de clase.**

En este ejemplo, se utiliza el método RER o el JK2. La variable que identifica el tamaño de la clase de los colegios es el VarStrat. La VarUnit es la variable que se emplea para identificar el colegio muestreado dentro del VarStrat. La variable empleada debe tener dos valores únicos, tales como 1 y 2, en cada VarStrat. (véase Anexo Ejemplo 2.2).

### **DISEÑO 2.3: Más de dos UMP se muestran de un número reducido de estratos.**

Para este diseño, se utiliza el método de replicación JK<sub>n</sub>. Se definen los estratos muestrales como el VarStrat y las UMP como la VarUnit.

**Ejemplo 2.3: Se selecciona una simple muestra aleatoria y estratificada de 100 negocios de cuatro estratos. Los negocios podrían ser muestreados con probabilidad igual o tamaños desiguales.**

En este ejemplo, se utiliza el método JK<sub>n</sub>. Se clasifican los estratos de muestreo como el VarStrat. Los negocios son la VarUnit. El número de réplicas que se crea es 100 (véase el Ejemplo 2.3 del Anexo).

## **3. TEMAS RELACIONADOS CON LAS APLICACIONES DE MUESTREO FINITO**

Esta sección incluye varios temas importantes. Primero examinamos los métodos de reflejar el efecto del factor de la corrección de población finita (cpf) para reducir el error de muestreo en los diseños muestrales sin reemplazamiento. En segundo lugar, debido a que la variabilidad de las estimaciones de varianza se captura en grados de libertad (gl), examinamos cómo evaluar el gl en el Apartado 3.2. El Apartado 2 hablaba de diseños estratificados en los cuales un método directo de construir réplicas conllevaría un esfuerzo de cálculo extensivo. Los Apartados 3.3 y 3.4 tratan de la combinación de las UMP y la combinación de estratos como métodos para reducir el número de réplicas necesarias para estimar las varianzas. El último apartado habla del método de estimar las varianzas cuando se selecciona una UMP única de un estrato.

### **3.1 Corrección de Población Finita**

La cpf es un factor que reduce la varianza porque se hace una muestra de una gran proporción de la población. (Si se incluyera toda la población, habría un error de muestreo cero. No examinamos el caso común, relacionado pero ambiguo de la estimación de variabilidad de datos de un censo con no-respuesta.). En general, la cpf sólo es aplicable a las muestras mono-

etápicas con muestreo equiprobable. Cuando se incorpora el efecto de la cpf, normalmente se supone que todas las unidades muestreadas respondan.

- Para SRS  $fpc = (1 - n/N)$ ; (3)

donde  $n$  es el número de unidades muestreadas y que responden, seleccionadas de una población de unidades  $N$ .

- Para STSRS  $fpc_h = (1 - n_h/N_h)$ ; (4)

donde  $n_h$  es el número de unidades muestreadas y que responden, seleccionadas de una población estrato de unidades  $N_h$ .

La versión estratificado es difícil de aplicar con el método RER si no hay una cpf común para todos los estratos. (véase Wolter 1.985 para un enfoque que revisa las ponderaciones de los que responden). Por otra parte, la cpf se aplica fácilmente utilizando JK2 o JKn porque las réplicas son definidas por los estratos. Recomendamos este método en las siguientes muestras.

### **DISEÑO 3.1: Muchas UMP son muestreadas de un número reducido de estratos (y la cpf es significativa).**

Siempre que el número de réplicas no sea excesivo, se utiliza el método de replicación JKn. Se definen los estratos muestrales como VarStrat y las UMP como la VarUnit. Si la cpf es significativa para cualquier VarStrat, debe ser contabilizada multiplicando la contribución de la réplica por el factor de reducción de la varianza apropiada, (4). (WesVar puede contabilizar una cpf significativa mediante la pantalla Adjuntar Factores -*Attach Factors*- donde los factores cpf se introducen como si fuesen los términos  $h_k$  en (1).)

#### **Ejemplo 3.1: Se selecciona una muestra simple estratificada aleatoria de 50 negocios de cuatro estratos.**

En este ejemplo, se utiliza el método JKn. Los estratos muestrales se clasifican como el VarStrat y los negocios como VarUnits. El número de réplicas creadas es 50. Para cada uno de los 50 VarStrat, se proporciona la cpf. (Cuando se emplea el software WesVar, utilicen la pantalla Adjuntar Factores -*Attach Factors*- donde  $fpc_h = 1 - n_h/N_h$ ) (véase Ejemplo 3.1 Anexo).

#### **Variación 3.1.1: La muestra sólo tiene dos UMP por estrato pero la fracción muestral en algunos de los estratos es elevada (es decir, mayor de 20).**

En esta variación, se utiliza el método JK2;. Se emplean factores cpf únicos para cada réplica o para conjuntos de réplicas con factores cpf similares. (Si se utiliza el Programa WesVar,

se emplea la pantalla Adjuntar Factores -*Attach Factors*- para especificar la corrección cpf para cada VarStrat). Para estratos con las cpf cercanas al 1, quizás no sea necesario incluir la cpf (véase el Anexo Variación 3.1.1).

**Ejemplo 3.1.1: Se selecciona una muestra mono-etápica de institutos en un estado.**  
**Los institutos son estratificados en 50 categorías de tamaño y dos institutos son seleccionados por estrato. Supongamos que el porcentaje de institutos muestreado en los cinco estratos de categoría de tamaño más grandes son: 50% en el más grande, 30% en el siguiente más grande, 20% en el tercero, y 10% en el cuarto y quinto más grandes. En todos los estratos restantes, menos del 5% de los institutos son muestreados.**

Se utiliza la variable categoría de tamaño como el VarStrat. Las variables de identificación del instituto (clasificados 1 y 2 dentro del tamaño) son la VarUnit. Si los 5 valores más elevados de VarStrat corresponden a las categorías de tamaño más grandes, se adjuntan factores de 0.5, 0.7, 0.8, 0.9, y 0.9 a estas réplicas. Las fracciones de muestreo en el VarStrat restante son tan reducidas que un factor de 1.0 es razonable para cada una.

### 3.2 Grados de Libertad de la Estimación de Varianza

Los grados de libertad de la estimación de varianza se emplean principalmente para calcular la amplitud del intervalo de confianza, I.C., para el dato estadístico estimado, a menudo mediante una constante obtenida de la distribución-t (9 o el equivalente para el test de hipótesis). El I.C. para el parámetro  $\theta$  se obtiene mediante:

$$\hat{\theta} \pm t_{df, \alpha/2} \sqrt{v(\hat{\theta})} \quad (5)$$

La amplitud del I.C. es más grande porque la varianza es estimada. Cuanto más precisa sea la estimación de la varianza, más pequeña será el I.C. La cantidad del aumento es la razón del dato estadístico-t a un dato estadístico-Z que se emplearía si se conociese la varianza:

$$\frac{t_{df, \alpha/2}}{Z_{\alpha/2}} \quad (6)$$

La tabla 3-1 muestra esta razón para los diferentes gl. Evidentemente, la razón disminuye según aumenta el gl. Así, es recomendable que la estimación de varianza tenga un número adecuado de grados de libertad. Una indicación aproximada del gl para la estimación de réplicas es  $\sum n_h - L$ , el número de UMP menos el número de estratos. Por lo tanto, para dos UMP por diseño de estrato, el gl será aproximadamente el número de réplicas. No obstante, en algunas situaciones esto puede ser engañoso y el número de réplicas será una sobreestimación del verdadero gl. La tabla indica que existe muy poco error en el aumento de la amplitud I.C. con gl de 30 o más, y apenas existe error para 60 gl o más.

Tabla 3-1. El incremento en amplitud de un 95% I.C. debido al gl de la estimación de varianza

gl	t	t/z
1	12.71	6.48
3	3.18	1.62
5	2.57	1.31
10	2.23	1.14
15	2.13	1.09
20	2.09	1.06
25	2.06	1.05
30	2.04	1.04
35	2.03	1.04
40	2.02	1.03
50	2.01	1.02
60	2.00	1.02
70	1.99	1.02
80	1.99	1.01
90	1.99	1.01
100	1.98	1.01
Infininito	1.96	1.00

Si se dispone de o se puede estimar la información necesaria, se puede hacer una aproximación del gl utilizando la fórmula siguiente:

$$df \approx \frac{2Var(\hat{\theta})}{Var(v(\hat{\theta}))} \quad (7)$$

donde  $Var$  es la varianza y  $v(\hat{\theta})$  es la varianza estimada.

Cuando se trata de un diseño estratificado y la estimación de totales de población, el denominador de esta expresión es aproximadamente:

$$Var(v(\hat{\theta})) = \sum_h \frac{N_h^4 S_h^4}{n_h^3} \left( \beta_h - \frac{n_h - 3}{n_h - 1} \right) \quad (8)$$

donde  $\beta_h$  es la kurtosis y  $S_h$  es la desviación standard de la UMP en el estrato h.

Cuando  $\beta_h = 3$  y todos los  $n_h$  son elevados, entonces el  $df \approx \sum n_h - L$ ; cuando la asignación de muestra es diferente, el número posible de gl puede ser mucho menor a  $\sum n_h - L$ .

### 3.3 La combinación de UMP para la Estimación de Varianza

Si se ha muestreado mucho más de 100 UMP, entonces para utilizar los métodos jackknife o RER tal y como se ha descrito arriba, serían necesarias más de 100 réplicas. El esfuerzo de

cálculo para repetir el proceso de estimación un gran número de veces, especialmente para los análisis iterativos como la regresión logística, puede resultar inaceptable. Además, puede ser innecesario, como se explica a continuación. Por supuesto, con un coste de cálculo mucho más reducido que en el caso anterior, el problema es mucho menor que cuando los equipos informáticos y el tiempo de cálculo eran más costosos. Para reducir el esfuerzo de cálculo, no obstante, es posible combinar las UMP sin una reducción significativa del gasto para la estimación de varianza mientras se reduce el número de réplicas a un número más aceptable en términos de cálculo. El objetivo es reducir los requisitos de cálculo necesarios para un diseño de muestreo mediante la combinación de UMP a la vez que se retiene una estimación no sesgada de la varianza.

Los métodos para combinar las UMP se describen a continuación. Para el método JK1, se crean grupos G de UMP, para que probablemente cada grupo proporcione una buena estimación de la característica de población. Para los métodos RER, JK2, y JKn, dentro de cada estrato, se crean grupos  $G_h$  (donde  $G_h = 2$  para los RER y JK2), así que probablemente cada grupo proporcionará una buena estimación de la característica de población del estrato.

**DISEÑO 3.3.1: Un número elevado de UMP se muestrea como un SRS. Definir cada UMP como una VarUnit en JK1 conllevaría más réplicas que las deseadas.**

En esta situación, se utiliza el método de replicación JK1. Se determina el número de réplicas deseado, digamos G; y se combinan las UMP en G unidades VarUnit para que cada VarUnit sea aproximadamente un subconjunto aleatorio de la muestra completa.

**Ejemplo 3.3.1: Se selecciona una muestra simple aleatoria de 500 pacientes de un hospital, pero sólo se requieren 50 réplicas.**

En este ejemplo, se utiliza el método JK1 y se crea una variable UMP mediante la asignación aleatoria de valores de 1 a 50 para que se asigne 10 pacientes a cada UMP. Se emplea esta variable UMP como la VarUnit (véase el Anexo Ejemplo 3.3.1).

**DISEÑO 3.3.2: Se muestrea un elevado número de UMP de forma sistemática sin una estratificación explícita.**

Esta situación es similar al DISEÑO 2.1, salvo que se ha muestreado un elevado número de UMP. En este caso, funcionará el método recomendado anteriormente. Además, proporcionamos una alternativa.

También funcionará una solución que utiliza el método JK2. El número de réplicas deseado se indica como G. Se combinan las UMP en seudo-estratos G. Al combinar las UMP, se tiene en cuenta cualquier muestreo sistemático. Aunque no había estratos reales, se definen los seudo-

estratos como unidades que son adyacentes en la lista sistemática, y se les asigna al mismo VarStrat.

**Ejemplo 3.3.2: Se muestrea una muestra de 400 negocios de forma sistemática con la probabilidad proporcional al cuadrado del número de empleados. Cada uno de los 400 negocios es una UMP. Se forman 100 réplicas.**

En este ejemplo, se utiliza el método de replicación JK2 y se hace una lista de los 400 negocios con el orden del muestreo. Se asignan los negocios 1-4 al VarStrat 1, se identifican dos como VarUnit 1 y dos como VarUnit 2. Se asignan los negocios 5-8 al VarStrat 2, se identifican dos como VarUnit 1 y dos como VarUnit 2. Se asignan los negocios de esta manera hasta que los negocios 397-400 se asignen al VarStrat 100 (véase Anexo Ejemplo 3.3.2).

**DISEÑO 3.3.3: En una muestra estratificada, el número de UMP muestreadas en unos pocos estratos es tan elevado que definir cada UMP como una VarUnit conllevaría más réplicas de las deseadas.**

En este diseño, se utiliza el método de replicación JK<sub>n</sub>; no obstante, se combinan las UMP en la VarUnit dentro de los estratos utilizando los métodos descritos anteriormente.

**Ejemplo 3.3.3: Hay 800 negocios muestreados de 10 estratos con una probabilidad proporcional al tamaño (PPT) donde la medida del tamaño es el total de las ventas anuales. VarStrat identifica los 10 estratos muestrales. En los primeros ocho VarStrat, el número de UMP (negocios) muestreadas es relativamente reducido. En los últimos dos VarStrat se seleccionan muchas UMP (se muestrean 200 en el estrato 9; se muestrean 500 en el estrato 10).**

Para la velocidad de cálculo, no son deseables más de 200 réplicas. Las UMP en el VarStrat 9 y 10 deben ser combinadas. En el VarStrat = 9, se hace una lista de los 200 negocios muestreados en el orden de muestreo y se asignan 1, 51, 101, y 151 como VarUnit = 1. Se asignan 2, 52, 102, y 152 como VarUnit = 2, etc.

Se hace lo mismo para los 500 negocios muestreados del VarStrat 10. Esto proporciona alrededor de 50 grados de libertad para VarStrat 9 y 10 y el número máximo para los demás (véase Anexo Ejemplo 3.3.3).

### **3.4 Combinar los Estratos para la Estimación de Varianza**

Al igual que con un número elevado de UMP, si se han utilizado muchos más de 100 estratos en el diseño muestral, entonces para emplear los métodos jackknife o RER, harán falta más de 100 réplicas. De la misma forma que las UMP pueden ser combinadas con el fin de

reducir el esfuerzo de cálculo, también se pueden combinar los estratos. La combinación de estratos es distinto al colapso, un método que se requiere para manejar los diseños en los cuales se muestrea una UMP única de un estrato. El colapso se explica en el Apartado 3.5.

Al igual que cuando se combinan las UMP, el objetivo es reducir el esfuerzo de cálculo, sin introducir un sesgo, a la vez que se retiene un elevado número de grados de libertad. Asimismo, al igual que antes, no se utilizan los datos de la muestra para tomar decisiones sobre la combinación de estratos, porque esto puede conllevar sesgos, en general subestimaciones de los errores de muestreo. Combinar los estratos se denomina "equilibrio parcial" en el diseño RER. La elección de los estratos que se van a combinar debería seguir el principio de que los estratos combinados proporcionan una buena estimación de la población de los mismos estratos.

Un ejemplo muy simplificado puede ayudar a aclarar el significado de la combinación de estratos. Pongamos como ejemplo un diseño muestral con cuatro regiones de un país (noreste, noroeste, sudeste y sudoeste). Se tratarán a estas regiones como estratos con dos UMP seleccionadas de cada. Una solución JK2 directa implicaría cuatro réplicas, y cada una de ellas excluiría una de las dos UMP en un estrato. La combinación de los cuatro estratos en dos, se puede obtener una estimación no sesgada (aunque más variable) del error de muestreo mediante la utilización de un total de sólo dos réplicas.

Una estimación combinada con un único grado de libertad puede formarse de la manera siguiente. Se creará una variable VarStrat con valor 1 para cualquier UMP muestreada de las regiones noreste o sudoeste (estratos) 1 o 2. La variable VarUnit sigue siendo el número UMP (1 o 2) dentro del estrato muestreado. Se formará una única réplica en la cual se excluya una UMP de las regiones noreste y sudoeste. Se puede hacer lo mismo en los otros dos estratos (noroeste y sudeste) para generar la segunda de las dos réplicas de estratos combinados. Este procedimiento hace una estimación de la varianza contribuida mediante el muestreo de ambos estratos de una manera no sesgada, aunque generará una estimación de varianza más variable que el método directo que se recomendó en el párrafo anterior.

Si no se combinan los dos estratos, entonces harían falta dos réplicas y conllevaría una estimación con aproximadamente dos grados de libertad. En una réplica se excluiría una de las UMP del estrato noreste y la otra se ponderaría para representar el estrato total. En la segunda réplica se aplica el mismo procedimiento a las dos UMP del estrato sudoeste.

Como contraste, una estimación de un estrato agrupado que este sesgada puede ser producida mediante la creación de la misma variable VarStrat con un valor 1 para las UMP tanto del noreste como sudoeste pero con una variable VarUnit que se numera de 1 a 4 para las cuatro UMP muestreadas en estos dos estratos. En esta situación, se utilizaría el JK<sub>n</sub> y se producirían cuatro réplicas en cada estrato agrupado; no obstante, la estimación resultante sobrecalcula el

error de muestreo del diseño porque crea un componente entre estratos que no corresponde al diseño muestral.

#### **Ejemplo 3.4: Hay 400 UMP muestreadas de 200 estratos, con dos UMP por estrato.**

Si se asignasen los 200 como VarStrat, habría 200 réplicas con JK2 o RER. Para mejorar la velocidad de cálculo, se reduce el número de réplicas a  $G = 80$  mediante la combinación de 200 estratos en 80 VarStrat. Se utiliza el método de replicación RER o JK2. Se hace una lista de los 200 estratos para que los estratos adyacentes sean lo más parecidos. Se asignan estratos 1, 81, y 161 a VarStrat = 1. Se asignan estratos 2, 82, y 162 a VarStrat = 2. Se sigue hasta que a los estratos 80 y 160 se les asigna VarStrat = 80. Se deja que el VarUnit sea la variable UMP con valores de 1 y 2 que se asignaron antes de combinar los estratos (véase Anexo Ejemplo 3.4).

#### **3.5 Agrupar Estratos para la Estimación de Varianza**

Cuando se muestrea una UMP única de un estrato, no existe manera no sesgada de estimar la variabilidad de muestreo dentro del estrato. El método tradicional para tratar esta situación es agrupar los estratos con unidades únicas de muestreo con otro estrato. Cuando se emplea este método, no obstante, es probable que el método de los ‘estratos agrupados’ sobreestime la varianza del diseño de una UMP por estrato. Utilizando los datos de marco, se agrupan para que las UMP en estratos agrupados sean lo más parecidas posible.

#### **Ejemplo 3.5: Se muestrea una muestra de 200 explotaciones agrícolas con una explotación seleccionada por estrato.**

Se utiliza el método RER o JK2. Se ordenan los 200 estratos para que sean lo más parecidos posible en relación con las estimaciones clave que se van a producir (región, clase de explotación, tamaño, etc.). Se agrupan los estratos 1 y 2 en VarStrat=1; 3 y 4 en VarStrat=2; etc. Se asigna cada explotación de forma aleatoria en un VarStrat para que VarUnit=1 para una explotación y VarUnit=2 para la otra.

### **4. DISEÑOS MUESTRALES MULTI-ETAPICOS**

La mayoría de los procedimientos ya explicados para diseños muestrales mono-etápicos se pueden aplicar a los muestreos multi-etápicos. No obstante, hay que tener cuidado en algunas circunstancias especiales. Por ejemplo, cuando hay unidades auto-representantes, o cuando el factor de corrección de población finita (cpf) es de alguna importancia. Estas circunstancias especiales se describen en los Apartados 4.1 y 5. El Apartado 4.1 explica el factor de corrección de población finita (cpf) y otros temas del muestreo multi-etápicos. El Apartado 4.2 describe cómo crear ponderaciones replicadas para dos tipos de diseño: (1) exactamente dos UMP muestreadas en cada estrato, y (2) más de dos UMP por estrato.

## 4.1 Aspectos de los Muestreos Multi-Etápicos

En general, el uso del factor de corrección de población finita es inapropiado e innecesario con el muestreo multi-etápico. Cuando las fracciones de muestreo en la primera etapa del muestreo son pequeñas, menores del 20%, por ejemplo, en general no es necesario incorporar la cpf cuando se calcula la varianza. La cpf puede ser inapropiada como medio de ajustar la estimación de varianza si el muestreo en la primera etapa se realiza con probabilidades desiguales de selección.

El sesgo que resulta al ignorar la cpf en el muestreo multi-etápico es (Wolter, 1985, eqn. 24.16):

$$Bias = 2Var(\hat{\theta}_{WR}) - Var(\hat{\theta}_{WOR})$$

donde  $Var(\hat{\theta}_{WR})$  es la varianza con reemplazamiento, y  $Var(\hat{\theta}_{WOR})$  es la varianza sin reemplazamiento. El sesgo sólo ocurre en el componente UMP media de la estimación de la varianza. Si se aplica una cpf general a la estimación de varianza con reemplazamiento, afectará tanto el componente medio como el interior, así que la cpf en general no es apropiada.

Los procedimientos para combinar las UMP y combinar los estratos explicados para los muestreos mono-etápicos en los Apartados 3.3 y 3.4 también pueden aplicarse a los muestreos multi-etápicos. Las UMP deben ser combinadas de forma que la nueva UMP combinada proporcione una estimación no sesgada de la población o estrato. Asimismo, los estratos combinados deberían proporcionar una estimación no sesgada de la población de UMP en el conjunto de estratos seleccionados para la combinación. No se deben combinar estratos hasta tal punto que las unidades combinadas aporten una fracción elevada a la varianza.

### DISEÑO 4.1: Una UMP muestreada de varios estratos.

Ninguna estimación no sesgada de la varianza es posible cuando se muestra sólo una UMP por estrato. El método de los estratos agrupados a menudo es apropiado en esta situación. Con este método, los estratos con sólo una UMP muestreada se combinan con otros estratos similares. Entonces se aplica el método de replicación más apropiado a los estratos agrupados.

### Ejemplo 4.1: Se selecciona un muestreo multi-etápico de una UMP por estrato. Los estratos se ordenan de forma útil.

Este diseño muestral puede ser manejado mediante la reducción de la muestra a una muestra estratificada con dos UMP seleccionadas por estrato, a continuación se utilizan los

métodos RER o JK2. Primero, se agrupan las parejas de estratos muestrales que están adyacentes en la lista ordenada. Los estratos agrupados se denominan VarStrat. Cada VarStrat debería tener dos UMP muestradas que se clasifiquen aleatoriamente o VarUnit=1 o 2 (véase el Anexo Ejemplo 4.1).

## 4.2 Procedimientos para algunos Diseños Muestrales

### DISEÑO 4.2: Dos UMP muestradas de cada estrato.

En esta situación, se puede utilizar el método RER (incluyendo la variación de Fay) o el JK2. Ambos métodos requieren exactamente dos VarUnits en cada VarStrat. Los estratos de muestreo de primera etapa se definen como el VarStrat y se definen las UMP como la VarUnit mediante la asignación aleatoria de valores de 1 o 2 dentro de cada estrato.

**Ejemplo 4.2:** Los colegios se estratifican por el número de alumnos. Se muestran dos colegios PPT en cada estrato de tamaño de clase. Se submuestran clases dentro de colegios y alumnos dentro de clases.

La variable del tamaño de clase es el VarStrat, porque los colegios se estratifican por tamaño. La VarUnit es la variable que identifica el colegio muestrado dentro del VarStrat (véase el Anexo Ejemplo 4.2).

### DISEÑO 4.3: Se muestrea un número par de UMP de algunos estratos.

En este diseño, se puede utilizar el método RER (incluyendo la variación de Fay), JK2 o el JKn. Los estratos de primera etapa se definen como el VarStrat. Para el RER y el JK2, se agrupan las UMP para formar dos VarUnit dentro de cada VarStrat. La variable VarUnit se asigna primero agrupando todas las UMP pares en el VarStrat en un grupo y todas las UMP impares en el segundo grupo. A continuación se asigna de forma aleatoria a cada grupo de UMP un valor 1 o 2 dentro de cada VarStrat. Este método no es apropiado si se muestran muchas UMP en un número reducido de estratos porque conllevaría insuficientes grados de libertad para la estimación de varianzas. Para el JKn, sencillamente se fija la VarUnit igual al identificador UMP. Si esto da como resultado demasiadas réplicas, se agrupan las UMP primero antes de asignar la VarUnit.

### Ejemplo 4.3: Un estrato tiene seis UMP muestradas.

Se puede utilizar los métodos RER (incluyendo la variación de Fay), JK2, o JKn. Para crear la VarUnit para el método RER o JK2, se asignan de forma aleatoria tres de las UMP a VarUnit = 1 y las otras tres a VarUnit = 2. Si las UMP se muestran en orden sistemático, (es decir, se ordenan dentro de un estrato, entonces se muestran utilizando el muestreo sistemático), después se asignan la primera, tercera y quinta UMP a VarUnit=1 y las otras tres a VarUnit=2 (véase el

Anexo Ejemplo 4.3). Para utilizar el JK<sub>n</sub>, se asigna cada UMP a su propio VarUnit. Esto tendrá como resultado la creación de 6 réplicas para este estrato.

**Variación 4.3.1: Se muestran dos UMP en la mayoría de los estratos, pero se muestran tres UMP en algunos estratos.**

Suponiendo que el número de estratos sea elevado y ningún estrato aporte una fracción elevada a la varianza de las estimaciones, un método simple es combinar de forma aleatoria dos de las UMP en una para aplicar el modelo standard de RER o JK2 (véase el Anexo Variación 4.3.1).

**Ejemplo 4.3.1: Una muestra tiene 100 estratos. Los estratos 1 a 99 tienen dos UMP muestreadas. El estrato 100 tiene tres UMP muestreadas, denominadas *a*, *b*, y *c*.**

Para utilizar RER o JK2, se seleccionan dos UMP de forma aleatoria en el estrato 100, y se denominan *a* y *b*. Se utiliza la variable del estrato como el VarStrat, y la recién creada variable UMP como la VarUnit.

Si este procedimiento tiene como resultado grandes sobre-estimaciones de la varianza, entonces se pueden crear dos o más réplicas para los estratos con tres UMP. Una manera sencilla de hacer esto es utilizar el JK<sub>n</sub>. Con el JK<sub>n</sub>, se crea una réplica por cada VarUnit de forma que se creen dos réplicas por cada estrato con dos UMP, y se creen tres réplicas para estratos con tres UMP.

Otra opción es crear dos réplicas para los estratos con tres UMP. En la primera réplica, se ponderan dos de las tres UMP (por ejemplo *a* y *b*) por 1.5 y *c* se pondera 0. En la segunda réplica, se pondera otro grupo de dos UMP (*a* y *c*) por 1.5 y *b* se pondera 0.

**DISEÑO 4.4: Se muestrea un número impar de UMP en algunos estratos.**

Cuando se muestrea un número impar de UMP (tres o más) en algunos estratos, se debería utilizar el método JK<sub>n</sub>. El número de réplicas que resultan será igual al número de UMP muestreadas. Si este número es demasiado grande, el número de réplicas puede ser reducido mediante el colapso de UMP antes de que se asigne la VarUnit.

**Ejemplo 4.4.1: Se muestran 60 colegios en tres estratos. Los tamaños de la muestra de estratos son 15, 20, y 25, y las fracciones de muestreo son pequeñas. Los profesores se muestrean dentro de los colegios.**

En esta situación, se puede utilizar el JK<sub>n</sub> con el VarStrat fijado igual al estrato del colegio y la VarUnit fijada igual a la variable que identifica el colegio. Sesenta ponderaciones replicadas JK<sub>n</sub> se crearán para cada colegio y cada profesor mediante la omisión de un colegio cada vez.

**Ejemplo 4.4.2: Se muestrean 1.000 colegios en 8 estratos. Los colegios se ordenan por matrícula dentro de cada estrato antes del muestreo. Los tamaños de la muestra del estrato UMP varían de 100 a 255 y las fracciones de muestreo del colegio son pequeñas. Se submuestrean los alumnos dentro de los colegios.**

Si se utilizara el JK<sub>n</sub> aquí con el VarStrat asignado al estrato de colegio y la VarUnit asignada a la variable que identifica el colegio, esto conllevaría 1.000 réplicas. Con el fin de reducir el número de réplicas, se agrupan los colegios antes de asignar la VarUnit. Por ejemplo, si se desea tener 100 réplicas, entonces se agrupan los colegios en 100 VarUnit. Para hacer esto, se ordenan los colegios dentro de cada estrato en el orden en que fueron seleccionados. Después, se numeran de 1 a 100 repetidamente, empezando de nuevo con uno hasta alcanzar el final de cada estrato. En cada estrato, se fija la VarUnit igual al número de secuencia asignado al colegio. Debido a que los números de secuencia van de 1 a 100, esto tendrá como resultado 100 VarUnit y 100 ponderaciones replicadas para cada colegio y alumno. En un estrato que contenga 255 colegios muestreados, esto es equivalente a asignar los colegios 1, 101, 201 a VarUnit = 1; colegios 2, 102, 202 a VarUnit = 3; colegios 3, 103, 203 a VarUnit = 3, y así sucesivamente.

## 5. UNIDADES AUTO-REPRESENTANTES

En una muestra mono-etápica, las unidades de certeza o auto-representantes (AR) no contribuyen al error de muestreo. Se pueden manejar fácilmente si se fija el peso de la réplica igual al de la muestra completa. Si la muestra es multi-etápica con submuestreo, se trata a cada unidad AR como un estrato, y se definen las réplicas utilizando las unidades de la segunda etapa. Si hay muchas unidades de segunda etapa, utilicen los métodos para combinarlas. Si la unidad AR es más grande que el estrato medio, se le asignan más réplicas mediante la creación de scudo-estratos.

### 5.1 Diseños Mono-Etápicos

**DISEÑO 5.2: El diseño es una muestra mono-etápica estratificada con un estrato todo incluido (AR).**

Se fijan las ponderaciones de las réplicas igual a la ponderación de la muestra completa para las unidades del estrato AR. Así, el estrato AR no contribuirá a las varianzas estimadas. En los estratos restantes, se utiliza el RER, JK<sub>2</sub>, o JK<sub>n</sub> para crear réplicas, según el número de UMP de los estratos.

**Ejemplo 5.2:** Se selecciona una muestra mono-etápica de universidades en una provincia. Se seleccionan las universidades con certeza y otras se estratifican por tamaño con dos muestreadas por cada 10 estratos. La variable de estrato toma el valor 1 para universidades y los valores 2 a 11 corresponden a los 10 estratos de tamaño. Dentro de los 10 estratos de tamaño, se seleccionan las universidades con probabilidades iguales.

Se puede utilizar el método JK2 porque se muestran dos universidades muestreadas en cada uno de los 10 estratos. Cada universidad muestreada en VarStrat 2 a 11 tiene una variable con la VarUnit igual a 1 o 2 que se asigna de forma aleatoria. En VarStrat 1, se fijan todas las ponderaciones replicadas iguales a la de la muestra completa. Si la fracción muestral es grande en cualquier de los VarStrat restantes, se incluye la cpf.

## 5.2 Diseños Multi-Etápicos

**DISEÑO 5.3:** Se muestrea una o más UMP con certeza (AR) en una muestra multi-etápica.

Una muestra multi-etápica de las UMP tiene algunas UMP AR. En todas las UMP AR, se muestran las unidades de segunda etapa (USE). Se trata a cualquier UMP AR como un estrato separado de varianza (VarStrat). Se tratan a las USE como UMP muestreadas y son las VarUnit.

Si no hay USE en algunas de las UMP AR, estas UMP AR no contribuyen al error de muestreo de estimaciones. Con el fin de asegurar que sus respuestas se incluyan en la estimación pero no contribuyan a la estimación de error de muestreo, sus ponderaciones replicadas deben ser iguales a la ponderación de la muestra completa.

**Ejemplo 5.3.1:** Se selecciona una muestra multi-etápica de dos UMP de zona por estrato, salvo que en un estrato se selecciona una UMP AR única. Esta UMP AR se divide en bloques y se muestrean dos bloques.

Se puede utilizar el método RER o el JK2, porque se muestran dos bloques en la UMP AR y se muestrean dos UMP en los estratos restantes. Se denomina la UMP AR como un VarStrat y se denomina cada una de los bloques en la UMP AR como la VarUnit (véase el Anexo Ejemplo 5.3.1). En los estratos restantes, se asigna el VarStrat al estrato y la VarUnit a la UMP que identifica.

**Variación 5.3.2:** El número de unidades de segunda etapa muestreadas dentro de la UMP AR es grande.

Primero, se determina el número de VarStrat que se va a asignar a cada UMP AR. Si la UMP AR es aproximadamente del mismo tamaño que el otro VarStrat, entonces se le asigna a un

VarStrat. Si es mucho más grande, se le asigna un número más grande de VarStrat. Después de crear el número deseado de VarStrat, se trata a la USE dentro de la UMP AR como VarUnit. Si esto tiene como resultado demasiadas réplicas, entonces se combinan las USE. Hay que tener en cuenta que si se incluyera una muestra estratificada en la UMP AR con  $m$  estratos, entonces se crearían  $m$  réplicas de la UMP AR si no se combinasesen los estratos.

**Ejemplo 5.3.2: Se muestrean bloques de una ciudad de forma sistemática, sin estratificación, en una ciudad grande que es una UMP AR. Supongamos que la ciudad es aproximadamente cuatro veces la población del VarStrat medio.**

En este ejemplo, se supone que el JK<sub>n</sub> conllevaría demasiadas réplicas. Por lo tanto, se utilizará el método RER o JK<sub>2</sub> con la creación de VarStrat y VarUnit de la forma siguiente. Se asignan cuatro VarStrat a la UMP AR organizando los bloques con un orden muestreado y denominando la primera cuarta parte de los bloques como VarStrat = 1, el segundo cuarto de bloque como VarStrat = 2, etc. Dentro del VarStrat, se asignan la mitad de los bloques de forma aleatoria a VarUnit = 1 y la otra mitad a VarUnit = 2 (véase Anexo Ejemplo 5.3). Alternativamente, se puede asignar la VarUnit mediante la ordenación de los bloques dentro de cada VarStrat con un orden serpenteante como 1,2,2,1,1,2,2,1. No sería apropiado asignar la VarUnit como 1,2,1,2 porque esto podría crear un sesgo. Por ejemplo, si los bloques se ordenasen por tamaño, la VarUnit 1 tendería a contener bloques más pequeños que la VarUnit 2.

## **6. LA CREACIÓN DE REPLICADAS PARA LA NHIS**

### **6.1 Introducción a la NHIS**

Utilizando los métodos explicados en los Apartados 2, 3, 4 y 5, esta sección describe la creación de REPLICADAS para la estimación de varianza en la Encuesta Nacional sobre la Salud -National Health Interview Survey (NHIS), una encuesta doméstica de gran escala sobre la población civil, no-institucionalizada de los EEUU. La Encuesta Nacional sobre la Salud comenzó en 1.957 y se han modificado el contenido y el diseño de muestreo cada 10 a 15 años aproximadamente. Consta de un cuestionario principal y varios cuestionarios especializados y suplementarios. Debido a que la NHIS proporciona estimaciones de características importantes de la salud de la población de los EEUU, hacen falta medidas fiables de precisión para las estimaciones.

La NHIS emplea un diseño muestral complejo que incluye la estratificación y varias etapas muestrales. Se ajustan las ponderaciones muestrales para la no-respuesta y se post-estratifican los totales de población. El diseño muestral complejo y los procedimientos de estimación complican la estimación de varianza. Además, la cuestión del secreto estadístico conlleva la supresión de la información completa en el estrato y los identificadores UMP en los ficheros de uso público. Como resultado, el Centro Nacional de Estadística Sanitaria - National Center for Health Statistics (NCHS) ofrece propuestas para simplificar la estimación de varianza de la NHIS que ha enmascarado ligeramente el estrato y las variables UMP.

La documentación que acompaña al fichero público de 1.995 de la NHIS explica dos métodos para estimar las varianzas. El primer método trata la muestra de la NHIS como un diseño de dos UMP por estrato y es apropiado para el software de replicación. El segundo método incorpora más información sobre el diseño muestral; no obstante, tal y como se presenta conlleva un número muy elevado de REPLICADAS. Los dos métodos permiten al usuario especificar el estrato y las UMP variables enmascaradas para cada ficha en el fichero público, y por lo tanto se puede utilizar cualquier software apropiado para la estimación de varianza de un diseño complejo.

Esta sección presenta una tercera alternativa a las dos estrategias descritas en la documentación de 1.995 de la NHIS que utilizan los métodos explicados en apartados anteriores.

### **6.2 Visión global del Diseño Muestral y la Estimación de la NHIS**

El diseño muestral de la NHIS es multi-etápico, con estratificación y conglomerado en varias etapas. Se hace una partición del universo en aproximadamente 1.900 UMP, y se incluyen muchas grandes en la muestra con certeza (UMP AR). Se estratifican las demás UMP por regiones y características UMP dentro del Estado. Se seleccionan dos UMP en cada estrato no-autorepresentativo (NAR) con la probabilidad proporcional a la población de la UMP. Las SSU

se forman dentro de cada UMP seleccionada y se estratifican sobre la concentración de población negra y hispana. Se muestran las SSU con tasas diferentes, y se pueden muestrear domicilios dentro de las SSU muestreadas. Asimismo, se pueden submuestrear personas dentro de los domicilios muestreados.

Las ponderaciones de la NHIS dan cuenta de las varias etapas del muestreo, se ajustan para la no-respuesta y se post-estratifican a los totales de edad/sexo/raza o etnia. Se realizaron ajustes de ponderación adicionales a los datos de 1.995 para compensar los datos perdidos como resultado del cierre gubernamental de tres semanas a fin de año.

### **6.3 Estimación de varianza de la NHIS**

Dos métodos de estimación de varianzas son explicados en la documentación que acompaña al fichero de datos de uso público del año 1.995. Al principio se desarrollaron estos métodos con el fin de permitir a los usuarios calcular las varianzas utilizando los métodos de replicación o de linealización sin necesidad de tener conocimientos profundos de diseño muestral y los procedimientos de estimación de la NHIS. Aunque se prepararon para un diseño anterior de la NHIS, fueron adaptados al nuevo diseño. La base racional y los procedimientos se encuentran en los estudios de Parsons y Casady (1986) y Parsons, Chan, y Curtin (1990). Primero se describen los dos métodos y a continuación se presenta un método alternativo.

El método 1 sirve para el software de replicación que requiere exactamente dos UMP en cada estrato. Se hace una partición de la muestra en 187 estratos de varianza y cada estrato de varianza contiene exactamente dos seudo-UMP. La documentación indica que los estratos y las UMP se basan lo más posible en el diseño muestral para los estratos NAR donde dos UMP se muestran por estrato (aunque el colapso es necesario cuando sólo se muestrea una UMP en un estrato). Se tratan las UMP AR como estratos y se identifican todas las SSU dentro de una UMP AR como una de dos seudo-UMP.

El primer procedimiento lleva 187 estratos de varianza, cada uno con dos UMP. Si se emplea la BRR, entonces un diseño totalmente equilibrado requerirá 188 REPLICADAS y ponderaciones de REPLICADAS. Para facilitar la comparación con otros métodos, empleamos el método de replicación emparejado - jackknife (JK2) estratificado, que requiere 187 REPLICADAS y ponderaciones de REPLICADAS. Los métodos reales que se utilizaron con esta evaluación son algo más complejos que lo que se describe aquí. Los detalles completos se encuentran en Nixon et al. (1998).

El método 2 es la alternativa propuesta para usar con el software de linealización. Este método maneja los estratos NAR de casi la misma forma que el método 1, con dos UMP para cada estrato NAR y el colapso según sea necesario para los estratos en los cuales sólo se muestrea una UMP por estrato. La principal diferencia se encuentra en el manejo de las UMP AR. En el

método 2, se tratan todas las SSU en una unidad AR como UMP en vez de emparejarlas en seudo-UMP tal y como se hace en el método 1.

El método 2 se considera más eficaz en términos estadísticos, porque incorpora más información de diseño para las UMP AR. Sin embargo, el método conlleva 4,079 UMP en la UMP AR y 259 UMP en los estratos NAR. Aunque el método 2 debería producir estimaciones de varianza más estables que el método 1, el método 2 es computacionalmente intensivo si se utilizan métodos de replicación porque conllevaría miles de REPLICADAS y ponderaciones de REPLICADAS. Aunque el método 2 resulta más estable que el método 1, sería enormemente impreciso suponer que el número de gl para las estimaciones de varianza calculadas con el método 2 podría estimarse mediante la diferencia entre el número de UMP y el número de estratos (Apartado 3.2). Los gl según este y otros métodos se describen a continuación.

Esta sección describe una tercera alternativa, denominada método 3, que aprovecha mejor la información del diseño en las UMP AR que el método 1 y sin necesidad de un elevado número de REPLICADAS como ocurre con el método 2. El objetivo principal al diseñar la alternativa es producir estimaciones de varianza estables y fiables para estimaciones de ámbito nacional a la vez que reduce el número de REPLICADAS a un nivel razonable. Esto se consigue utilizando la información de diseño del método 1 para los estratos NAR, la información de diseño del método 2 para las UMP AR, y combinando las unidades con el fin de reducir el número de REPLICADAS tal y como se describe en los apartados 3.3 y 5.2. La estrategia de combinación de unidades tiene en cuenta los dominios de análisis importantes y la estabilidad de las estimaciones de varianza.

Se pueden utilizar las distribuciones de la NHIS para evaluar cuántas REPLICADAS hagan falta. Las primeras columnas de la Tabla 2 muestran el número de UMP del Método 2 y los totales de población estimados para cada región y tipo de UMP (AR o NAR). Debido a que los estratos NAR son menos metropolitanos (un dominio de análisis importante) y más grandes en la población total que las UMP AR individuales, se limita la cantidad de colapsos de estratos. El número máximo de estrato de varianza para los estratos NAR en regiones 1, 2, y 4 son 14, 38, y 18, respectivamente. Al no combinar los estratos NAR, el número de gl de los estratos NAR es elevado, pero sólo hay 70 ( $14 + 38 + 18$ ) REPLICADAS entre los tres estratos. Para mantener el número total de REPLICADAS de las cuatro regiones NAR en 70, los estratos NAR en la región 3 deben ser combinados con estratos NAR de las otras regiones, pero esto sigue generando un máximo de gl para las estimaciones de la región 3.

Tabla 2. Método 3 Asignación de 70 VarStrat para la NHIS

Región	Tipo	Número de UMP	Población %	Región %	Número de VarStrat	% en VarStrat	VarStrat asignado
1	NAR	29	3,5%	17,9%	14	20,0%	01 a 14
	AR	994	16,1%	82,1%	56	80,0%	15 a 70
	Total		19,6%				
2	NAR	76	10,2%	42,9%	38	54,3%	15 a 52
	AR	766	13,6%	57,1%	32	45,7%	01 a 14/53 a 70
	Total		23,8%				
3	NAR	117	16,7%	47,5%	58	82,9%	01 a 58
	AR	1.210	18,5%	52,5%	70	*	59 a 70,01 a 59
	Total		35,3%				
4	NAR	37	5,2%	24,3%	18	25,7%	53 a 70
	AR	1.109	16,1%	75,7%	52	74,3%	01 a 52
	Total		21,3%				
Total	NAR	259	35,6%				
	AR	4.079	64,4%				
	Total	4.338	100,0%				

\*En esta categoría, las SSU se asignan primero a VarStrat 59 a 70 hasta llegar al mismo tamaño aproximadamente que el VarStrat NAR, a continuación se distribuyen por todos los VarStrat.

Se combinan las UMP AR, mediante el muestreo sistemático de las SSU en la Encuesta Nacional sobre la Salud. Una primera aproximación es emparejar las SSU consecutivas y tratarlas como selecciones emparejadas de un estrato dentro de la UMP AR. Debido a que se realiza el submuestreo mediante datos de raza/etnia, este emparejamiento debería captar algo de esa característica del diseño. Dado que esto conllevaría un número elevado de estratos de varianza, se combinan las unidades.

Para aumentar el gl para las estimaciones regionales, se realiza la combinación dentro de los estratos de varianza de una región no asignada a los estratos NAR. Por ejemplo, en la región 1 se asignan las unidades NAR a los estratos de varianza 1 a 14 y las unidades AR se asignan a los estratos de varianza 15 a 70. Tal y como se ve en la Tabla 2, esto conlleva la asignación proporcional aproximada de la población a cada estrato de varianza dentro de una región.

La única excepción es la región 3 donde se asignan las unidades NAR a los estratos de varianza 1-58 y esto cubre el 83% de los 70 estratos de varianza pero sólo el 48% de la población de la región. Las estimaciones de varianza son más eficaces si la población en los estratos de varianza es lo más proporcional posible, así que se combinaron algunas de las unidades AR con las unidades NAR en esta región. El Gráfico 1 muestra la asignación global de las unidades a los 70 estratos de varianza para el método 3.

Gráfico 1. Asignación de VarStrat por región y tipo

		Estrato de Varianza (VarStrat)																																				
Reg.	Tipo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
1	NAR	1	2	3	5	6	7	8	10	11	12	13	14																									
1	AR																																					
2	NAR															15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
2	AR																																					
3	NAR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
3	AR																																					
4	NAR																																					
4	AR																																					
		Estrato de Varianza (VarStrat)																																				
Reg.	Tipo	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70		
1	NAR																																					
1	AR																																					
2	NAR	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52																				
2	AR																																					
3	NAR	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58														
3	AR																																					
4	NAR																		53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70		
4	AR																																					

Algunas de las ventajas previstas del método 3 para formar estratos de varianza y REPLICADAS son: (1) sólo se emplean 70 REPLICADAS y ponderaciones de REPLICADAS, (2) las varianzas serán más constantes que con el método 1 como consecuencia de la mejora en el manejo de las UMP AR, y (3) los gl para dominios como región, categoría metropolitana, y raza/ctnia serán relativamente numerosos.

#### 6.4 Propiedades de los Estimadores de Varianza

El criterio principal para evaluar los tres estimadores es su varianza. El estimador de varianza jackknife reproduce la estimación de varianza clásica para los datos lineales, y se emplea como referencia para la comparación de los tres métodos. Siguiendo el proceso descrito en Rust (1.986), la varianza del estimador de varianza jackknife se consigue por ecuación (8). La combinación de estratos disminuye la precisión del estimador, tal y como se ve en la expresión de los estratos jackknife combinados

$$V(v_{CJ}(\hat{\theta})) = \sum_{h=1}^H \frac{W_h^4 S_h^4 (\beta_h - 3)}{n_h^3} + 2 \sum_{g=1}^G \left( \frac{\sum_{h \in g} \frac{W_h^2 S_h^2}{n_h}}{l_g - 1} \right)^2 \quad (10)$$

donde  $g$  representa un estrato combinado que consiste en los estratos originales  $H_g$ ;  $W_h$  es la proporción de la población  $n$  en el estrato  $h$ ;  $l_g$  es el número de UMP en un estrato  $g$  combinado (en este ejemplo, siempre es 2) y los otros valores se definieron antes.

Se puede utilizar esta expresión para evaluar la precisión de un estimador de varianza si se conocen los valores de  $W_h$ ,  $S_h^2$ , y  $\beta_h$ . En general, este no es el caso y se sustituyen las estimaciones muestrales constantes. Kish (1.965) muestra  $\beta$  como aproximadamente 3 para conglomerados grandes en el muestreo de poblaciones humanas tal y como se emplea en la NHIS. Suponemos que  $\beta_h = 3$ . El valor de  $W_h$  se estima de una encuesta para cada estrato de cada dominio. El valor de  $S_h^2$  es más difícil de estimar con precisión. Estimamos el efecto del diseño por separado para los estratos AR y NAR y supusimos que era constante para todos los estratos dentro de estos grupos grandes. El valor de  $S_h^2$  se calculó como una constante por el efecto estimado del diseño donde la constante representa  $p(1-p)/n_h$  donde  $p = .5$  y  $n_h = 2$ .

## 6.5 Comparación de los métodos

Para comparar los tres métodos, se calcularon los grados de libertad estimados para cada método. Primero, se calcularon las estimaciones de varianza utilizando la formula (8) para el jackknife completo y la formula (10) para los estratos jackknife combinados. Se produjeron las estimaciones de gl para cada método utilizando estas varianzas y el ratio mostrados en la formula (7). Los métodos 1 y 3 requirieron la expresión jackknife combinada mientras que el método 2 utilizaba la expresión jackknife exacta para las varianzas.

Tal y como se ha explicado arriba, mientras que el método resulta apropiado para la replicación, no proporciona un número elevado de gl para algunas estimaciones subnacionales. El método 2 sí proporciona un número adecuado de gl para estimaciones subnacionales pero requiere demasiadas REPLICADAS como para ser una solución práctica. El método que proponemos sólo requiere 70 REPLICADAS, y tiene casi tantos gl para estimaciones nacionales como el método 1 recomendado por la NCHS a la vez que proporciona suficientes grados de libertad para datos estadísticos subnacionales importantes. Se presenta el gl de varios datos subnacionales para los tres métodos en la Tabla 3. Se incluyen los resultados del método 2 para completar, pero sólo se comparan los métodos 1 y 3 porque son útiles para la replicación.

Table 3. Grados de libertad para dominios según los tres métodos

Dominio	Método 1	Método 2	Método 3
Nacional	72	501	66
Región 1	7	169	61
Región 2	24	122	57
Región 3	51	140	59
Región 4	12	113	56
Pobreza	53	154	52
Raza negra	26	163	55
Hispanos	10	53	32
No raza negra	71	465	66
MSA	48	481	63
No MSA	60	60	42

La tabla muestra que el método 3 solo proporciona 66 gl para las estimaciones nacionales, comparado con 72 gl para el método 1 y 500 para el método 2. Esta diferencia de gl entre los métodos 1 y 3 no debería causar un aumento importante en la variabilidad de las estimaciones de varianza. En la mayoría de las estimaciones subnacionales, sin embargo, el método 3 proporciona más gl que el método 1, y a menudo la diferencia es significativa. Hay que destacar la diferencia de gl para varias estimaciones regionales (1 y 4) en las cuales se da un aumento por cuatro en gl. Las estimaciones para los sub-grupos de raza y etnia tendrán dos o tres veces la cantidad de gl con el método 3.

Por último, conviene resaltar que el método 3, que se construye utilizando los principios de combinación de estratos, tiene suficientes gl (es decir., >30) para todos los dominios subnacionales de la tabla.

## 7. Conclusión

Este estudio comenzó con una breve descripción de los dos métodos predominantes para estimar los errores de muestreo para estimaciones complejas basadas en diseños muestrales complejos: linealización y replicación. El Apartado 1 incluye una revisión de algunas de las características técnicas tras el método la replicación y explica varias de las ventajas. Los Apartados 2, 3, 4 y 5 explican los diseños muestrales comunes y proporcionan ideas para adaptar cada diseño a un método de replicación para la estimación de varianza.

Además del sencillo diseño de dos UMP por estrato, cuyos los procedimientos se describen en Wolter (1.985), los métodos de replicación se presentaron para varios diseños más complejos

pero frecuentemente utilizados. Esto incluye los diseños con más de dos UMP, muestreo sistemático, y UMP de certeza. Los Apartados 3 y 5 explican los métodos para combinar las UMP y combinar los estratos con el fin de reducir el número de REPLICADAS necesarias y a la vez producir una estimación de varianza asintóticamente no sesgada. Los temas importantes sobre el papel de la corrección de población finita y el efecto de los grados de libertad se examinaron también en el Apartado 3.

Por último, se presentó un ejemplo práctico utilizando la actual NHIS de los EEUU. Se presentó un método de replicación que sólo requiere 70 REPLICADAS y es capaz de producir no sólo estimaciones de errores de muestreo nacionales con bastante precisión, sino también estimaciones subnacionales.

## BIBLIOGRAFIA

- Binder, D.A. (1996) Linearization methods for single phase and two-phase samples: a cookbook approach, *Survey Methodology*, **22**: 17-22.
- Brick, J.M., Kalton, G. (1996) Handling missing data in survey research, *Statistical Methods in Medical Research*; **5**: 215-38.
- Graubard, B.I., Korn, E.L. (1996) Survey inference for subpopulations, *American Journal of Epidemiology*, **144**: 102-106.
- Hinkins, S., Moriarity, C., Scheuren, F. (1996) Replicate variance estimation in stratified sampling with permanent random numbers, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Judkins, D. (1990) Fay's method for variance estimation, *Journal of Official Statistics* **6**, 223-240.
- Kish, L. (1965) *Survey Sampling*, New York: Wiley.
- Kish, L., Frankel, M. (1974) Inference from complex samples, *Journal of the Royal Statistical Society; B36*, 1-22.
- Kovar, J.G., Rao, J.N.K., Wu, C.F.J. (1988) Bootstrap and other methods to measure errors in survey estimates, *Canadian Journal of Statistics*; **16**: 25-46.
- Krewski, D., Rao, J.N.K. (1981) Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods, *Annals of Statistics*; **9**: 1010-19.
- McCarthy, P.J. (1966) Replication: an approach to the analysis of data from complex surveys, *Vital and Health Statistics, Series 2, No. 14*, National Center for Health Statistics, Public Health Service, Washington, D.C.
- Nixon, M., Brick, J.M., Kalton, G., Lee, H., Givens, J. (1998) Alternative variance estimation methods for the NHIS, to be published in *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Parsons, V.L., Casady, R.J. (1986) Variance estimation and the redesigned National Health Interview Survey, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Parsons, V.L., Chan, J., and Curtin, L.R. (1990) Analytic limitations to current National Health Surveys, *Proceedings of the Section of Survey Research Methods of the American Statistical Association*.
- Rao, J.N.K., Shao, J. (1996) On balanced half-sample variance estimation in stratified random sampling, *Journal of the American Statistical Association*, **91**, 343-348.
- Rao J.N.K., Wu, C.F.J., Yue, K. (1992) Some recent work in resampling methods, *Survey Methodology*; **18**: 209-17.

- Rust, K. (1986) Efficient replicated variance estimation, *Proceedings of Survey Research Methods Section of the American Statistical Association*, 81-87.
- Rust, K.F., Rao, J.N.K. (1996) Variance estimation for complex surveys using replication techniques, *Survey Methods in Medical Research*; **5**: 283-310.
- Shah, B.V., Barnwell, B.G., Hunt, P.N., LaVange, L.M. (1992) *SUDAAN user's manual*, Release 6.0, Research Triangle Park, NC: Research Triangle Institute.
- Valliant, R. (1996) Limitation of balanced half-sampling, *Journal of Official Statistics*, 12, 225-240.
- Valliant, R. (1990) Comparisons of variance estimators in stratified random and systematic sampling, *Journal of Official Statistics*; **6**: 115-31.
- Westat, Inc. (1997) *A user's guide to WesVarPC*, Version 2.1, Rockville, MD: Westat, Inc.
- Westat, Inc. (1998) WesVar complex samples 3.0 user's guide, Chicago, IL: SPSS, Inc.
- Wolter, K.M. (1985) *Introduction to Variance Estimation*, New York: Springer.
- Yung, W. (1996) Contributions to poststratification in stratified multi-stage samples, PhD thesis, Carleton University.

# **APENDICE**

Ejemplo 2.1 Creación de 50 réplicas de una muestra de 50 pacientes (UMP)

PSU	1	VarUnit	1
PSU	2	VarUnit	2
PSU	3	VarUnit	3
PSU	4	VarUnit	4
PSU	5	VarUnit	5
PSU	6	VarUnit	6
PSU	7	VarUnit	7
PSU	8	VarUnit	8
PSU	9	VarUnit	9
PSU	10	VarUnit	10
PSU	11	VarUnit	11
PSU	12	VarUnit	12
			<b>SE UTILIZA JK1</b>
PSU	13	VarUnit	13
PSU	14	VarUnit	14
PSU	15	VarUnit	15
.	.	.	
.	.	.	
.	.	.	
.	.	.	
PSU	45	VarUnit	45
PSU	46	VarUnit	46
PSU	47	VarUnit	47
PSU	48	VarUnit	48
PSU	49	VarUnit	49
PSU	50	VarUnit	50

Ejemplo 2.2. Muestra estratificada de colegios, dos muestrado por estrato de 50 estratos

MUESTREADOS

PSUs

	PSU	1
	PSU	2
ESTRATO 1	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
	PSU	9
	PSU	10
	PSU	1
	PSU	2
ESTRATO 2	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
	PSU	9
	PSU	10
	PSU	11
	PSU	12
	PSU	13
	PSU	1
	PSU	2
ESTRATO H	PSU	3
	PSU	4
	PSU	5 X
	PSU	6
	PSU	7
	PSU	8
	PSU	9
	PSU	10 X
	PSU	11
	PSU	12
	PSU	13

SE UTILIZA BRR O JK2

Ejemplo 2.3. 100 establecimientos muestreados de no más de 50 estratos. cada estrato tiene al menos dos UMP muestreadas

SAMPLED PSUs		
ESTRATO 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6 X
	PSU	7
	PSU	8 X
	PSU	9
	PSU	10
.		.
.		.
.		.
PSU		34
PSU		35
PSU		36 X
PSU		37
ESTRATO2	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	.	
	.	
	.	
	PSU	29
	PSU	30 X
PSU		31
.		.
.		.
.		.
ESTRATO H	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X
	.	
	.	
	.	
	PSU	39
	PSU	40 X
PSU		41
PSU		42

Ejemplo 3.1. 50 establecimientos muestreados de cuatro estratos

SAMPLED PSUs		
ESTRATO 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6 X
	PSU	7
	PSU	8 X
	PSU	9
	PSU	10
	PSU	11 X
	PSU	12
	PSU	13 X
	PSU	14
	PSU	15
	PSU	16 X
	PSU	17
ESTRATO 2	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
	PSU	9
	PSU	10
	PSU	11
ESTRATO 4	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X
	PSU	6
	PSU	7
	PSU	8
	PSU	9
	PSU	10 X
	PSU	11
	PSU	12

Variante 3.1.1 Dos PSU por diseño de estrato pero con fpc significativo

	SAMPLED PSUS	
ESTRATO 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
	PSU	9
	PSU	10
ESTRATO 2	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6
	PSU	7 X
	PSU	8
<hr/>		
.		
.		
.		
ESTRATO H	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X
	PSU	6
	PSU	7
	PSU	8
	PSU	9
	PSU	10 X
	PSU	11
	PSU	12
	PSU	13

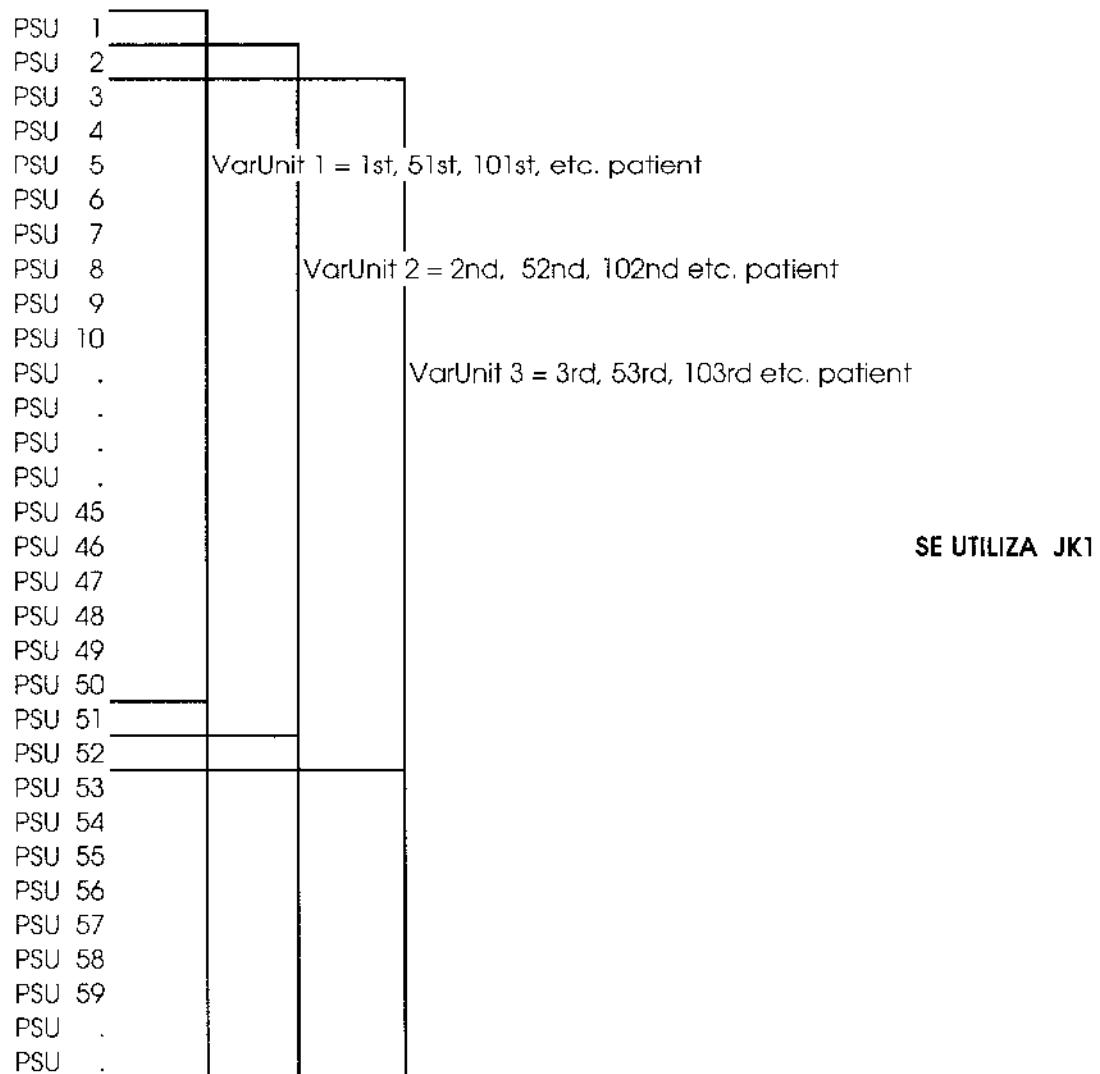
**SE UTILIZA JK2  
Y SE ADJUNTAN FACTORES  
A LOS ESTRATOS CON  
fpc SIGNIFICATIVO**

Ejemplo 3.1.1. Una muestra mono-etápica estratificada de institutos de 50 estratos de tamaño, 2 UMP tomados por estrato. El fpc es significativo en algunos estratos.

		SAMPLED PSUs
ESTRATO 1	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5
	PSU	6 X
ESTRATO 2	PSU	1 X
	PSU	2
	PSU	3 X
	PSU	4
.		
.		
.		
ESTRATO 50	PSU	1
	PSU	2
	PSU	3
	PSU	4
	PSU	5 X

**SE UTILIZA BRR O  
JK2 CON fpc  
DONDE CONVIENE**

Ejemplo 3.3.1 Creación de 50 réplicas de una muestra de 500 pacientes (UMP)



Ejemplo 3.3.2. Creación de 100 réplicas de una muestra sistemática de 400 negocios

		SAMPLED PSUs	VarUnit
SEUDO ESTRATO 1	PSU	1	
	PSU	2	
	PSU	3 X	1
	PSU	4	
	PSU	5 X	2
	PSU	6	
	PSU	7 X	2
	PSU	8	
	PSU	9	
	PSU	10 X	1
SEUDO ESTRATO 2	PSU	11 X	1
	PSU	12	
	PSU	13 X	2
	PSU	14	
	PSU	15 X	2
	PSU	16	
	PSU	17 X	1
	PSU	18	
<hr/>			
.			
.			
SEUDO ESTRATO 100 PSU	PSU	4851	
	PSU	4852 X	2
	PSU	4853	
	PSU	4854	
	PSU	4855 X	1
	PSU	4856	
	PSU	4857	
	PSU	4858 X	1
	PSU	4859	
	PSU	4860 X	2
<hr/>			
PSU			
PSU			
PSU			
<hr/>			

SE UTILIZA BRR O JK2

Ejemplo 3.3.3. 800 establecimientos muestreados de 10 estratos.  
 Se toma una muestra grande de los estratos N° 9 y 10

	SAMPLED PSUs
STRATUM 1	1
	2
	3 X
	4
	5
	6
	7 X
	8
	9
	10 X
	11
	12
	13

SE UTILIZA JK <sub>n</sub>		
ESTRATO 9 (200 UMP)	PSU	1
	PSU	2
	PSU	3
	PSU	4 X
	MUESTREADA	5
	PSU	COMBINE
	PSU	6 X
	PSU	.
	PSU	X
	PSU	.
	PSU	X
	PSU	501
	PSU	502 X
	PSU	503
	PSU	504 X
ESTRATO 10 (500 UMP)	PSU	1
	PSU	2
	PSU	3 X
	PSU	4
	PSU	5 X
	MUESTREADA	COMBINE
	PSU	.
	PSU	X
	PSU	.
	PSU	X
	PSU	2101
	PSU	2102 X
	PSU	2103 X
	PSU	2104

Ejemplo 3.4. Combinar estratos para reducir el número de réplicas

	SAMPLED PSUs	VarUnit
	PSU 1	
	PSU 2	
ESTRATO 1	PSU 3 X	1
	PSU 4	
	PSU 5	
	PSU 6	
	PSU 7 X	2
	PSU 8	
	PSU 9	
	PSU 1	
	PSU 2	
ESTRATO 2	PSU 3 X	1
	PSU 4	
	PSU 5	
	PSU 6	
	PSU 7 X	2
	PSU 8	
	PSU 9	
.		.
.		.
.		.
ESTRATO 81	PSU 1	2
	PSU . X	
	PSU .	
	PSU . X	1
	PSU 13	
.		.
.		.
ESTRATO 161	PSU 1	1
	PSU . X	
	PSU .	
	PSU . X	2
	PSU 13	
.		.
.		.
.		.
STRATUM 200	PSU 1	
	PSU 2	
	PSU 3	
	PSU 4	
	PSU 5 X	2
	PSU 6	
	PSU 7	
	PSU 8	
	PSU 9	
	PSU 10 X	1
	PSU 11	

SE UTILIZA BRR  
O JK2

Ejemplo 4.1. Combinar estratos con una UMP muestreada

		<u>SAMPLED PSUs</u>	VarUnit
SE COMBINAN ESTOS ESTRATOS PORQUE SON SIMILARES	PSU	1	
	PSU	2	
	ESTRATO 1		1
	PSU	3 X	
	PSU	4	SE UTILIZA BRR O JK2
	PSU	5	
	PSU	6	
	PSU	7	
	PSU	8	
	PSU	9	
	PSU	10	
ESTRATO 2	PSU	1	
	PSU	2	
	PSU	3 X	2
	PSU	4	
	PSU	5	
	PSU	6	
	PSU	7	
	PSU	8	
	PSU	9	
	PSU	10	
	PSU	11	
	PSU	12	
	PSU	13	
<hr/>			
<hr/>			
<hr/>			
STRATUM H	PSU	1	
	PSU	2	
	PSU	3	
	PSU	4	
	PSU	5 X	1
	PSU	6	
	PSU	7	
	PSU	8	
	PSU	9	
	PSU	10 X	2
	PSU	11	
	PSU	12	
	PSU	13	

Ejemplo 4.2 Dos colegios(UMP) se muestran aleatoriamente en cada estrato de tamaño de clase

SAMPLED PSUs	
ESTRATO 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
<hr/>	
ESTRATO 2	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
	PSU 9
	PSU 10
	PSU 11
	PSU 12
	PSU 13
<hr/>	
<hr/>	
<hr/>	
ESTRATO H	PSU 1
	PSU 2
	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

Ejemplo 4.3 Se muestrean en un estrato un número par, más de dos, de colegios.

		SAMPLED PSUs	
ESTRATO 1	PSU	1	
	PSU	2	
	PSU	3 X	
	PSU	4	COMBINE
	PSU	5 X	
	PSU	6	
	PSU	7 X	
	PSU	8	
	PSU	9 X	
	PSU	10	
ESTRATO 2	PSU	11 X	
	PSU	12 X	
	PSU	13	
	PSU	1	
	PSU	2	
	PSU	3 X	SE UTILIZA BRR, JK2
	PSU	4	o JKn
	PSU	5	
	PSU	6	
	PSU	7 X	
STRATUM H	PSU	8	
	PSU	9	
	PSU	10 X	
	PSU	11	
	PSU	12	
	PSU	13	
	PSU	1	
	PSU	2	
	PSU	3	
	PSU	4	
	PSU	5 X	
	PSU	6	
	PSU	7	

Variante 4.3.1. Se muestrea un número impar de unidades en algunos estratos

PSUs	
ESTRATO 1	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
	PSU 7 X
	PSU 8
ESTRATO 2	PSU 9 X
	PSU 10
	PSU 1
	PSU 2
	PSU 3 X
	PSU 4
	PSU 5
	PSU 6
ESTRATO H	PSU 7 X
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

RANDOMLY COMBINE

SE UTILIZA BRR O JK2

ESTRATO H	PSU 1
	PSU 2
	PSU 3
	PSU 4
	PSU 5 X
	PSU 6
	PSU 7
	PSU 8
	PSU 9
	PSU 10 X
	PSU 11
	PSU 12
	PSU 13

Ejemplo 5.3.1. Muestra multi-etápica, 2 UMP por estrato en todos los estratos menos uno con 2 SSU  
in the SR PSU

	PSU	1 X	SSU	1	
			SSU	2 X VarUnit 1	
ESTRATO 1			SSU	3	
			SSU	4	
			SSU	5	
			SSU	6	
			SSU	7	SE UTILIZA BRR O JK2
			SSU	8	
			SSU	9	
			SSU	10	
			SSU	11 X VarUnit 2	
			SSU	12	
			SSU	13	
			SSU	14	
			SSU	15	
			SSU	16	
			SSU	17	
			SSU	18	
ESTRATO 2	PSU	1			
	PSU	2			
	PSU	3 X			
	PSU	4			
	PSU	5			
	PSU	6			
	PSU	7 X			
	PSU	8			
	PSU	9			
	PSU	10			
	PSU	11			
	PSU	12			
	PSU	13			
	:				
ESTRATO H	PSU	1			
	PSU	2			
	PSU	3			
	PSU	4			
	PSU	5 X			
	PSU	6			
	PSU	7			
	PSU	8			
	PSU	9			
	PSU	10 X			
	PSU	11			
	PSU	12			
	PSU	13			

Ejemplo 5.3.2. Muestra multi-etápica, dos UMP por estrato en todos menos uno de los estratos con muchas SSU en la UMP AR

	PSU	1 X	SSU	1		
ESTRATO 1			SSU	2 X	1	VarUnit 1
			SSU	3		
			SSU	4		
			SSU	5 X	2	VarUnit 2
			SSU	6		
			SSU	7		
			SSU	8 X	2	
			SSU	9		
			SSU	10		
			SSU	11 X	1	
			SSU	12		
			SSU	13 X	1	
			SSU	14		
			SSU	15		
			SSU	16 X	2	
			SSU	17		
			SSU	18		
ESTRATO 2	PSU	1				
	PSU	2				
	PSU	3 X				<b>SE UTILIZA BRR O JK2</b>
	PSU	4				<b>SE COMBINAN LAS SSU</b>
	PSU	5				
	PSU	6				
	PSU	7 X				
	PSU	8				
	PSU	9				
	PSU	10				
	PSU	11				
	PSU	12				
	PSU	13				
<hr/>						
<hr/>						
ESTRATO H	PSU	1				
	PSU	2				
	PSU	3				
	PSU	4				
	PSU	5 X				
	PSU	6				
	PSU	7				
	PSU	8				
	PSU	9				
	PSU	10 X				
	PSU	11				
	PSU	12				
	PSU	13				

**ORAIN ARTE ARGITARATUTAKO GAIAK  
BOOKS PUBLISHED UP TO DATE  
CURSOS PUBLICADOS HASTA LA FECHA**

- 1.- LINEAL STATISTICAL INFERENCE  
(Inglés, español)  
C.R.RAO  
1983
- 2.- MUESTREO Y APLICACIONES  
(Español)  
E. CANSADO  
1983
- 3.- STATISTICAL EDUCATION  
(Inglés, Español)  
V. BARNETT  
1983
- 4.- ANALYSE DES DONNÉES  
(Francés, Español)  
P. CLAPIER  
1983
- 5.- DESIGN OF EXPERIMENTS  
(Inglés, Español)  
D.J.FINNEY  
1984
- 6.- ASPECTOS DE TEORIA Y APLICACIONES EN EL MUESTREO  
(Español)  
F.AZORIN POCH  
1984
- 7.- CURSO BASICO INTENSIVO DE MUESTREO  
(Español)  
J.L.SANCHEZ-CRESPO  
1985

- 8.- ANALYSE DES SERIES CHRONOLOGIQUES: LES INDICES STATISTIQUES  
(Francés, Español)  
J. FOURASTIE  
1985
- 9.- COMPENSATING FOR MISSING SURVEY DATA.  
G. KALTON  
1985  
ARGITARATU GABE. NO ESTÁ PUBLICADO.
- 10.- METHODOLOGY AND TREATMENT FOR NON-RESPONSE  
(Inglés, Español)  
R.PLATEK  
1986
- 11.- STATISTICAL OPERATION BY SAMPLING  
(Inglés, Español)  
L.KISH  
1986
- 12.- ANALISIS DE SERIES TEMPORALES: ALGUNAS TECNICAS DE PREDICCION  
(Español)  
I.GALLASTEGI  
1986
- 13.- BASES DE DATOS  
(Español)  
F.SALTOR  
1987
- 14.- METODOS ESTADISTICOS PARA LA INVESTIGACION EPIDEMIOLOGICA  
(Español)  
L.C.SILVA  
1987
- 15.- SAMPLING AND NON-SAMPLING ERRORS IN SURVEYS  
(Inglés)  
A.MARTON  
1988
- 16.- LES ENQUÊTES TELEPHONIQUES  
(Francés)

V.SALVY  
1988

17.- GENERALIZED LINEAR MODELS IN EPIDEMIOLOGY

(Inglés)  
J.C.DUFFY  
1989

18.- NEW TECHNOLOGIES IN COMPUTER ASSISTED SURVEY PROCESSING

(Inglés)  
J.G.BETHLEHEM AND W.J.KELLER  
1989

19.- EVALUATION OF QUESTIONNAIRE DESIGN EFFECTS

(Inglés)  
G.NATHAN  
1990

20.- PROCEDIMIENTO DE DEPURACION DE DATOS ESTADISTICOS

(Español)  
I.VILLAN CRIADO Y M.S.BRAVO CABRIA  
1990

21.- THE X11 ARIMA/88 SEASONAL ADJUSTMENT METHOD

(Inglés)  
E.BEE DAGUM  
1990

22.- RENTA Y DISTRIBUCION DE LA RIQUEZA, DESIGUALDAD Y POBREZA: TEORIA,

MODELOS Y APLICACIONES  
(Español)  
C.DAGUM  
1991

23.- LA CONTABILIDAD NACIONAL COMO MARCO DE LAS ESTIMACIONES DE VARIABLES

ECONOMICAS  
(Español)  
V.ANTON VALERO  
1991

24.- MACRO-EDITING. METHODS FOR RATIONALIZING THE EDITING OF QUANTITATIVE

DATA  
(Inglés)

L.GRANQUIST  
1991

25.- METHODOLOGICAL ISSUE IN FAMILY EXPENDITURE SURVEYS

(Inglés, Español)

M.KANTROWITZ  
1992

26.- QUALITY CONTROL IN STATISTICS FROM ADMINISTRATIVE REGISTERS AND RECORDS

(Inglés, Español)

HANS PETTERSSON  
1992

27.- ANALYSE DES DONNÉES ET CLASSIFICATION AUTOMATIQUE NUMÉRIQUE ET

SYMBOLIQUE

(Francés)

E.DIDAY  
1992

28.- METHODOLOGIE DE LA RECHERCHE EXPERIMENTALE

(Francés, Español)

ROGER PHAN TAN LUU  
1993

29.- STRUCTURAL EQUATION MODELING WITH LISREL

(Español)

KARL G. JÖRESKOG  
1993

30.- METODOS ESTADISTICO-ECONOMETRICOS PARA EL ANALISIS DE LA COYUNTURA

ECONOMICA

(Español)

ANTONI ESPASA  
1994

31.- ESTADISTICAS DE EMPLEO Y PARO EN LA UNION EUROPEA

(Español)

BERNARD GRAIS, ALOIS VAN BASTELAER, ANDRÉ PERSENAIRE  
1994

32.- EL SISTEMA EUROPEO DE ESTADISTICAS INTEGRADAS DE PROTECCION SOCIAL

(SEEPROS)

(Inglés, Español)

ALFONSO BARRADA, MERCEDES ALCALDE, COR N.GORTER  
1995

33.- CAPTURE-RECAPTURE MODELS: AN OVERVIEW

(Inglés, Español)  
KENNETH H. POLLOCK  
1995

34.- RESIDENTS REGISTRATION SYSTEMS AND STATISTICS

(Inglés, Español)  
PEKKA MYRSKYLÄ, WOLFGANG MOHR  
1995

35.- TECNICAS ESTADISTICAS EN LOS PROCESOS DE GESTION DE CALIDAD TOTAL EN LAS EMPRESAS Y ORGANIZACIONES DE SERVICIOS

(Español)  
A. PRAT BARTES, P. GRIMA CINTAS  
1997

36.- METODOS DE MUESTREO EN AUDITORIA E INTERVENCION

(Español)  
R. ESCUDER VALLES  
1997

37.- THE REPLICATION METHOD FOR ESTIMATING SAMPLING ERRORS

(Inglés, Español)  
DAVID MORGANSTEIN  
1998

38.- METHODOLOGY FOR A SATELLITE ACCOUNT OF HOUSEHOLD PRODUCTION: PAID AND UNPAID WORK

(Inglés, Español)  
JOHANNA VARJONEN  
1998