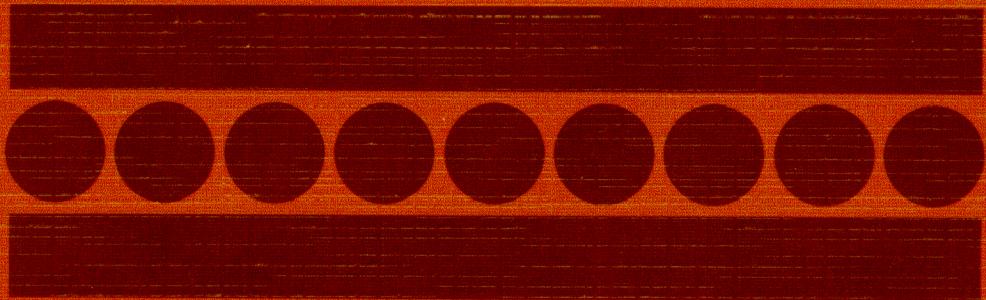


NAZIOARTEKO ESTATISTIKA  
MINTEGIA EUSKADIN

**1983**

SEMINARIO INTERNACIONAL  
DE ESTADÍSTICA EN EUSKADI



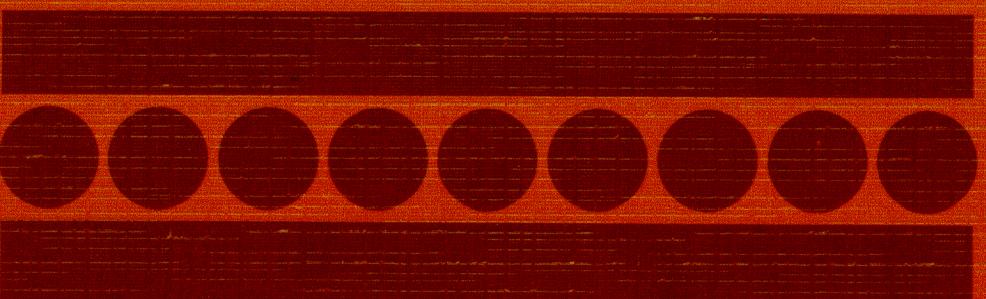
**INFERENTZIA ESTATISTIKO LINEALA**

**INFERENCIA ESTADÍSTICA LINEAL**

**INFERENCE STATISTIQUE LINEAIRE**

**LINEAL STATISTICAL INFERENCE**

**C.R. RAO**



EUSKO JAURLARITZA  
Estatística Zuzendaritza



GOBIERNO VASCO  
Dirección de Estadística

NAZIOARTEKO ESTATISTIKA  
MINTEGIA EUSKADIN

**1983**

SEMINARIO INTERNACIONAL  
DE ESTADISTICA EN EUSKADI

**INFERENTZIA ESTATISTIKO LINEALA**

●  
**INFERENCIA ESTADISTICA LINEAL**

●  
**INFERENCE STATISTIQUE LINEAIRE**

●  
**LINEAL STATISTICAL INFERENCE**

**C.R. RAO**

CUADERNO **1** KOADERNOA

EUSKO JAURLARITZA  
Estatística Zuzendaritza



GOBIERNO VASCO  
Dirección de Estadística

Zuzendaritza/Dirección: J. Pérez Vilaplana  
Lankidetza/Colaboración: Angeles Iztueta Azkue

© Eusko Jaurlaritza - Gobierno Vasco

Depósito Legal: S.S. 81-85  
ISBN: 84-7542-127-10. Obra Completa.  
ISBN: 84-7542-167-9  
Impreso en Itxaropena, S.A. - Errikobarra kalea, 2 - Zarautz.

## **AURKEZPENA**

Estatistikako Mintegi Internazionalak sustatzean, hainbat xederekin bete nahi luke Eusko Jaurlaritzaren Estatistika Zuzendaritzak, hala nola:

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

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

## **PRESENTACION**

Al promover los Seminarios Internacionales de Estadística, la Dirección de Estadística del Gobierno Vasco 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 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.

## **PRESENTATION**

La Direction de Statistique du Gouvernement Basque se propose d'atteindre plusieurs objectifs par la promotion des Séminaires Internationaux de Statistique:

- Encourager la collaboration avec l'université et spécialement avec les départements de statistique.
- Faciliter le recyclage professionnel des fonctionnaires, professeurs, élèves, et tous ceux qui pourraient être intéressés par la statistique.
- Inviter en Euskadi des professeurs mondialement renommés et des chercheurs de premier ordre en matière de Statistique avec tout ce que cela pourrait entraîner comme avantage dans les rapports et l'échange d'expériences.

En outre, il a été décidé de publier les exposés de ces rencontres afin d'atteindre le plus grand nombre de personnes et d'institutions intéressées, et pour contribuer ainsi à développer dans notre pays les connaissances sur cette matière. Dans chaque cas la langue d'origine du conférencier sera respectée.

## **PRESENTATION**

In promoting the International Seminars on Statistics, the Statistics Office of the Basque Government is attempting to achieve a number of objectives:

- Encourage joint working with the Basque University and, in particular, with its Department of Statistics.
- Facilitate the in-training of civil servants, teachers and students and of all those interested in the field of statistics.
- Bring to Euskadi distinguished academics and researchers in the front line of statistics work, at a world-wide level, with all the benefits that this will bring through direct contacts and the interchange of experiences and ideas.

As an additional step this year, it has been decided to publish in advance the papers to be presented at these courses, respecting the native language of the speaker, in each case. This is in order that as many interested people and institutions as possible are made aware. In this way we hope to contribute to the growth and awareness concerning this topic in our country.

Vitoria-Gazteiz, Diciembre 1984 Abendua

JOSE IGNACIO GARCIA RAMOS  
Estatistikako Zuzendaria  
Director de Estadística

## **SARRERA**

Liburu honek, Eusko Jaurlaritzaren Estatistika-Zuzendaritzak eta Euskal Herriko Unibertsitatearen, Matematika Aplikatuko Departamentuak antolaturik, C.R. RAOk Euskadiko Estatistikako I. Nazioarteko Mintegiaren barruan "Lineal Statistical Inference" gaiari buruz eman duen ikastaroa laburbiltzen du. I. Mintegi honek kontatzen du baita Txileko Santiagoko Unibertsitateko E. CANSADO Jaunaren partaidetzarekin ere, "Muestreo y Aplicaciones" gaiari buruzko ikastaro batekin, Sheffiel-eko Unibertsitateko V. BARNETT Jaunaren partaidetzarekin, "Statistical Education" gaiari buruzko ikastaroarekin, eta orobat Pariseko CREDOCeko P. CLAPIER Jaunaren partaidetzarekin, "Analyse des Données" gaiari buruzko ikastaroarekin.

## **INTRODUCCION**

Este cuaderno resume el curso que sobre "Lineal Statistical Inference" ha impartido C.R. RAO dentro del I Seminario Internacional de Estadística en Euskadi, organizado por la Dirección de Estadística del Gobierno Vasco y el Departamento de Matemática Aplicada de la Universidad del País Vasco. Este I Seminario cuenta con la participación de E. CANSADO de la Universidad de Santiago de Chile con un curso sobre "Muestreo y Aplicaciones", V. BARNETT de la Universidad de Sheffiel con un curso sobre "Statistical Education" y P. CLAPIER de CREDOC París con un curso sobre "Analyse des Données".

## **INTRODUCTION**

Ce cahier résume le cours que C.R. RAO a donné sur "Lineal Statistical Inference" dans le cadre du I Séminaire International de Statistique d'Euskadi, organisé par le Conseil de Statistique du Gouvernement Basque et par le Département de Mathématique Appliquée de l'Université du Pays Basque. Ce I Séminaire compte avec la participation de E. CANSADO de l'Université de Santiago de Chile avec un cours sur "Muestreo y Aplicaciones" ainsi que celle de V. BARNETT de l'Université de Shefield avec un cours sur "Statistical Education" et de P. CLAPIER du CREDOC (Paris) avec un cours sur "Analyse des Données".

## **INTRODUCTION**

This paper summarises the lecture given by R. RAO on "Lineal Statistical Inference" during the I International Statistics Seminar in Euskadi, organised by the Statistics Office of the Baque Government and by the Department of Applied Mathematics of the University of the Basque Country this I Seminar included the participation of E. CANSADO of the University of Santiago in Chile with his lecture on "Muestreo y Aplicaciones", "Statistical Education" and off P. CLAPIER from CREDOC in Paris with a lecture on "Analyse des Données".

## BIOGRAFIA

**C.R. Rao**, gaur egun, USAko Pittsburgh-eko Unibertsitatean irakasle dugu, eta irakasle baita Indiako Jawharlal Nehru Estatistikako Institutuan ere. Honoris Causa Doktore da honako universtitate hauetatik: SESBeko Leningradotik, Indiako Andhra, Deli eta Osmamiakoetatik, Greziako Atenaskotik, USAko Ohioko Columbus-ekotik, Peruko San Marcos Nazionaletik, Filipinetakotik eta Venezuelako Zentraletik.

Bilboko Estatistikako Erret Elkarteko Ohorezko Partaidea da, bestalde; USAko Nazioarteko Institutuko eta Arte eta Zientzien Amerikar Akademiako Ohorezko Bazkidea; B.H.ko Erret Elkarteko eta Indiako Zientzien Akademiako bazkidea, USAko Estatistika Matematikoko Institutuko, USAko Estatistikako Amerikar Elkarteko eta Ekonometriako Nazioarteko Elkarteko partaidea. Beste aldizkariren artean honakoen Argitalpen-Kontseiluko kidea dugu: Sankhya, Journal of Multivariate Analysis, Communications in Statistics eta Theory and Decision aldizkarietako alegia. 250 artikuluren egile dugu ikerketa estatistikoari buruz eta hainbat libururen autore, horien artean "Linear Statistical Inference and its Applications" liburua azpimarratu beharko litzatekeelarik, errusiera, alemaniera, japoniera, tsekiera, eta polskerara itzulia bera, eta estatistikako literaturan obrarik aipatuenetariko bat inongo ezpairik gabe.

## BIOGRAFIA

**C.R. RAO** es actualmente profesor de la universidad de Pittsburgh (USA) y profesor Jawaharlal Nehru del Instituto Indio de Estadística. Es doctor Honoris Causa por las universidades de Leningrado (URSS), Andhra, Delhi y Osmania (India), Atenas (Grecia), Ohio en Columbus (USA), Nacional de San Marcos (Perú), Filipinas y Central de Venezuela.

Es Honorary Fellow de la Real Sociedad de Estadística (U.K.). Miembro de Honor del Instituto Internacional de Estadística y de la Academia Americana de Artes y Ciencias (USA). Es miembro de la Real Sociedad (U.K.) de la Academia de Ciencias de la India, Fellow de Estadística Matemática (USA), de la Asociación Americana de Estadística (USA) y de la Sociedad Internacional de Econometría. Ha sido Presidente del Instituto Internacional de Estadística, del Instituto de Estadística Matemática (USA), de la Sociedad Internacional de Biometría y de la Sociedad Internacional de Econometría. Es miembro del Consejo Editorial de Sankhya, Journal of Multivariate Analysis, Communications in Statistics y Theory and Decision, entre otras revistas. Es autor de más de 250 artículos de investigación estadística y diversos libros, entre los que destaca "Linear Statistical Inference and its Applications" traducido al ruso, alemán, japonés, Checoslovaco y polaco, una de las obras más citadas en la literatura estadística.

## BIOGRAPHIE

**C.R. RAO** est à l'heure actuelle professeur à l'université de Pittsburgh (USA) et professeur Jawharlal Nehru de l'Institut Indien de Statistique. Il est Docteur Honoris Causa par les Universités de Leningrad (URSS), Andhra, Delhi, et Osmanie (Inde), Athènes (Grèce), Ohio à Columbus (USA), Nacional de San Marcos (Perou), Philippines et Central du Vénézuela.

Il est Honorary Fellow de la Société Royale de Statistique (UK), Membre d'honneur de l'Institut International de Statistique et de l'Académie Américaine d'Arts et de Sciences (USA). Il est membre de la Société Royale (UK), de l'Academie de Sciences de l'Inde. Fellow de l'Institut de Statistique Mathématique (USA), de l'Association Américaine de Statistique (USA) et de la Société Internationale d'Econométrie. Il est membre du Conseil Editeur de Sankhya, Journal of Multivariate Analysis, Communications in Statistics et Theory and Decision parmi d'autres revues. Il es l'auteur de plus de 250 articles de recherche statistique et de plusieurs livres parmi lesquels nous pouvons citer "Lineal Statistical Inference and its Applications" traduit au russe, allemand, japonais, tchèque et polonais. Il s'agit là de l'un des ouvrages les plus cités dans la littérature statistique.

## BIOGRAPHY

**C.R. RAO** is, at present, a professor at the University of Pittsburgh and Jawaharlal Nehru Professor of the Indian Institute of Statistics. Honorary Doctor of the Universities of Leningrad (USSR), Andhra, Delhi and Osmania (India) and Athens (Greece), of Columbus University, Ohio (USA), San Marcos University (Peru) and of the Central University (Venezuela).

Honorary Fellow of the Royal Society of Statistics (UK). Honorary Member of the International Institute of Statistics and of the American Academy of Arts and Sciences. Member of the Royal Society (UK), of the Academy of Sciences of India. Fellow of the Institute of Mathematical Statistics (USA) and of the International Society of Econometrics & of the American Association of Statistics. He has been President of the International Institute of Statistics, of the Institute of Mathematical Statistics (USA), of the International Society of Biometry and of the International Econometrics Society. Member of the Editorial Council of Sankhya, Journal of Multivariate Analysis, Communications in Statistics and Theory and Decision amongst other magazines. Author of more than 250 statistical research articles and several books amongst which "Linear Statistical Inference and its Applications" stands out and which has been translated into Russian, German, Japanese, Czechoslovak and Polish and which is one of the most quoted sources in statistics work.



## AURKIBIDEA / INDICE / INDEX

Aurkezpena/Presentación/Presentation/Presentation .....	5
Sarrera/Introducción/Introduction/Introduction .....	7
Biografia/Biografía/Biographie/Biography .....	9
Aurkibidea/Indice/Index .....	11
INFERENTZIA ESTATISTIKO LINEALA .....	13
LABURPENA .....	13
Estatistikariak, estatistikak eta politika publikoa .....	13
Banaketa binomial hiztunaren adibide natural bat .....	14
Eredu linealaren inferentzia ezarritako efektuekin - Oraindik emaitzak eta arazo zenbait .....	14
Erregresioari eta aurresanari buruzko zenbait gogoeta .....	14
Lineal Statistical Inference .....	15
STATISTICS, STATICIANS AND PUBLIC POLICY MAKING .....	15
1. Statistics .....	15
2. Staticians .....	16
3. Public Policy Making .....	17
4. Some new problems .....	18
5. Conclusions .....	19
A NATURAL EXAMPLE OF A WEIGHTED BINOMIAL DISTRIBUTION .....	21
1. An empirical result .....	21
2. Theoretical considerations - sampling proportional to size .....	21
3. A discussion on the mechanism of selection .....	22
INFERENCE FROM LINEAR MODELS WITH FIXED EFFECTS: RECENT RESULTS AND SOME PROBLEMS .....	24
1. Introduction .....	24
2. Unified approach to blue .....	24
2.1. Generalized projection .....	24
2.2. Least squares with restrictions .....	25
2.3. Inverse partitioned matrix (IPM) approach .....	25
3. Alternative criteria of estimation .....	26
4. Simultaneous estimation and prediction .....	26
4.1. Simultaneous estimation .....	26
4.2. Simultaneous prediction .....	28
4.3. Nonhomogeneous ridge estimator .....	28
5. Robustness of inference procedures .....	28
6. Canonical transformation .....	29
7. Model selection for prediction .....	31
8. Direct or inverse regression? .....	32
SOME THOUGHTS ON REGRESSION AND PREDICTION .....	34
1. Introduction .....	34
2. Disaggregation .....	35
3. Multicollinearity .....	36
4. Variables with a factor structure .....	40
5. Estimation of true measurement - an anomaly .....	42
ESTADISTICOS, ESTADISTICAS Y POLITICA PUBLICA .....	
1. Estadísticas .....	45
2. Estadísticos .....	46
3. Política Pública .....	47
4. Algunos problemas nuevos .....	48
5. Conclusiones .....	49
UN EJEMPLO NATURAL DE UNA DISTRIBUCION BINOMIAL .....	
1. Resultado empírico .....	52
1. Resultado empírico .....	52
2. Consideraciones teóricas-muestreo proporcional al tamaño .....	52
3. Discusión acerca del mecanismo de selección .....	53
INFERENCIA DEL MODELO LINEAL CON LOS EFECTOS ESTABLECIDOS: RESULTADOS RECIENTES Y DIVERSAS RESPUESTAS .....	
1. Introducción .....	55
2. Aproximación unificada a la estimación lineal insesgada óptima .....	55
2.1. Proyección generalizada .....	55
2.2. Mínimos cuadrados con restricciones .....	56
2.3. Enfoque de la matriz inversa particionada .....	57
3. Criterios alternativos de estimación .....	57
4. Predicción y estimación simultáneas .....	58
4.1. Estimación simultánea .....	58
4.2. Predicción simultánea .....	59
4.3. Estimador acanalado no homogéneo .....	59
5. Robustez de los procedimientos de inferencia .....	60
6. Transformación canónica .....	61
7. Selección del modelo de predicción .....	62
8. ¿Regresión directa o inversa? .....	62
ALGUNAS CONSIDERACIONES SOBRE PREDICCIÓN Y REGRESIÓN .....	
1. Introducción .....	64
2. Desagregación .....	65
3. Multicolinearidad .....	66
4. Variables con estructura factorial .....	70
5. Estimación de una medida verdadera-una anomalía .....	72

## **INFERENTZIA ESTATISTIKO LINEALA**

C.R.RAO  
Université de Pittsburgh  
U.S.A.

LABURPENA

### **ESTATISTIKARIK, ESTATISTIKAK ETA POLITIKA PUBLIKOA**

Estatistikarien erantzunkizunak ez dira datu eguneratu eta zehatzak ematearekin bakarrik bukatzen. Datuen analisia eta interpretazioan ere parte hartu behar dute. Datu-analisiak, atzeraelikadura ere permititzen du, datuok hobetzeko eta etorkizuneko premiak asetzeko.

Estatistikarien eta Zientzia Sozialako profesionarien artean lankidetza handiagoa behar da, eta hau/lor daiteke Unibertsitateetan ematen diren ikastaroen elkartruke egokiak eginez eta baita ere estatistikarien eta zientzi sozialako profesionarien artean behar den elkarrengina lortzeko gobernuko departamenduen jendehartze-politikaren bitartez.

Indikatzaile sozialei eta garapen-indizeei eman zaien garrantziarekin bide berriak ireki dira estatistika ofizialean, erantzunkizunak astunenak gobernuko estatistikarien gain geratuz.

Estatistikaren eragutza baliabide nazionala da, eta horrela ustiatu behar da, osoro eta efikazki.

Estatistika onak ezin daitezke estatistikari onen ordezko izan eta era berean, estatistikari onek ere/ez dituzte estatistika onak ordezkatzen. Biak dira beharrezko. Baino honekin ez dugu aski, estatistikarien

politikariekin batera lan egin behar dute erabakiak hartzekoan, baldin eta herrialdeak duen ezagutza estatistikoari ahalik eta onurariak handiena atera nahi badio behintzat.

Estatistika, ezagutzarako eta aurrerabiderako sarrera da, herrialde batek bere sistema estatisti-/koaren efizientziarekiko dependentzia bizia du.

#### **BANAKETA BINOMIAL HAZTATUAREN ADIBIDE NATURAL BAT**

Estatistika aplikatuez klasean ematen den adibide praktibo bat deskribatuko da, estatistika ikas-/ten hasten ari diren ikasleengan eragin izugarria izaten duena. Asmaketa-metodoei buruzko Fisher-en // (1.934) dokumentu aipagarri batean oinarritzen da.

#### **EREDU LINEALAREN INFERENTZIA EZARRITAKO EFETKUEKIN - (ORAINTSUKO EMAITZAK ETA ARAZO ZENBAIT)**

BLUEren eredu linealen teoria bateratua ematen da, errore-terminoen matrize diseinatua eta disper-tsio-matrizea, biak heinean defizienteak izan daitezkeenekoa. Eredu linealen zenbatespen eta aurresa-/te-arazo aldiberekoak, kuadratiko galeradunen funtzio komposatu baten menpean konsideratzen dira. Ketezko erregresioaren kalkulu jeneralizatua proposatzen da. Diseinuaren espezifikazio desegokitarako, / dispersio-matrizarako eta normaltasunetiko desbidazioetarako esangura-testak duten indarra azter-/tzen da. Nahikotasun linealaren eta gutxienezko nahikotasun linealaren kontzeptuk sartzen dira eta /// BLUEren klasea gutxienezko nahiko lineala dela erakusten da. Etorkizuneko behaketak aurresatetako eredu hautatzeko metodoak idarokitzen dira. Azkenik, Bayes-en zenbatespen empirikoen zenbait propietate az-/tertzen dira.

#### **ERREGRESIOARI ETA AURRESANARI BURUZKO ZENBAIT GOGOETA**

Erregresio-teknikaren aplikazio batzuk eta beronekin lotutako tranpa ahalgarriak aztertzen ditu // lan honek.

Erregresio-metodoak disagregazio-probleman okerreko emaitzetara eraman dezakeela demonstratzeko adi bideak ematen dira.

Multikolinealtasuna definitzerakoan, aldagai esplikatzaileen interes-eremu duen aurresatetako doitasunari egin zaio erreferentzia, eta ez aldagai hauen arteko interdependentzia handiari.

Aldagai esplikatzaileen bektorearen norabide batzutako aurresatetako oso ona izan daiteke // nahiz eta aldagaien arteko koerlazioa oso handia izan.

Multikolinealtasunaren efektuak xehetasunez estudiatu dira, aldagai esplikatzaileek behaketa-erroreak izan ditzaketen kasua ere kontuan izanik. Erregresio-funtzioaren zenbatespena konsideratua izan/da aldagai esplikatzaileek eta menpekoek eaitura faktoriala dutenean.

Bukatzeko, aldagai esplikatzaile ezezagun baten zenbatespena egiteko alderantzizko erregresioa zu-zeneko erregresioa baino baliagarriagoa den egoera bat deskribatzen da.

## **INFERENTZIA ESTATISTIKO LINEALA**

C.R.RAO

Université de Pittsburgh  
U.S.A.

### **STATISTICS, STATISTICIANS AND PUBLIC POLICY MAKING**

#### **I. STATISTICS**

*There cannot be a good plan for economic progress without adequate data and there cannot be adequate data / without a good plan for collecting them. P.C. Mahalanobis.*

*Any discussion on official statistics begins with an account of their defects in terms of gaps, economy, accuracy, presentation, priority phasing, personnel, interdepartmental coordination, etc., and ends with suggestions and hopes for improvements. Commissions have been appointed and committees are constituted, from time to time, to go into the problems of official statistics and make recommendations for improvement. There are, however, some inherent limitations in the development of efficient statistical systems. Some of the requirements like timeliness and economy conflict with demand for accuracy and more detailed information. The resources, expertise and personnel needed at any time to maintain a comprehensive statistical system to plan for the growing demand for information are often underestimated by the governments. Naturally, various types*

*of defects continue to exist in some form or other despite continuous efforts being made to remedy them. However, the present day official statistics in any country are in a much better shape than what they used to be. This is mainly due to the growing demand for statistics in formulating national plans for socio-economic development, which every developing country is trying to do to achieve quick improvements in the quality of life of the people.*

*Planning involves assessing the present against the background and experience of the past and setting goals for the future. Past trends in economy have to be examined and current inputs necessary to achieve future specified targets have to be determined. All these processes require the use of statistics in its dual aspects, viz., data and methodology. In fact, development of statistics is a prerequisite for the formulation of meaningful socio-economic / plans. The demand for statistics may increase in the future when planning done at the national level now is extended to regional and subregional levels. All these require the formulation of a broad plan for a coordinated national*

statistical system, and the creation of a special cell in the national statistical office with a group of experts / charged with the following responsibilities:

- To periodically review and suggest improvements / in the existing methods of collection and compilation of data in the traditional areas like agriculture and industry. As Claus Moser <sup>(1)</sup> says, / There is a built-in tendency in economic series to deteriorate unless enough care - and - maintenance work is done.
- To assess the data requirements for solving new / problems which continually arise in making policy decisions.
- To plan for collection of data through special / censuses and sample surveys in areas where it is not possible to acquire information through normal administrative channels.
- To lay down priorities in the compilation and publication of data.
- To organize <sup>(2)</sup> periodically seminars and / meetings with the participation of government / statisticians and outside specialists to discuss data improvements and to assess the future data / needs.
- To avoid duplication and coordinate the statistical activities of the different government agencies.

The last aspect, viz., coordination is extremely important specially if a country has a decentralized statistical system. But coordination is also a difficult task / and there may be administrative and technical bottlenecks which may not be easily resolved. <sup>(3)</sup>

## 2. STATISTICIANS

... the statistical science was the peculiar aspect of human progress which gave to the twentieth century its special character ... it is to the statistician that the present age turns for what is most essential in all its more important activities. R. A. Fisher.

What are the functions of a statistical office and / the responsibilities of government statisticians and the head of a statistical office? Is their job mainly to / collect and compile statistics for possible use (i.e., / statistical analysis and interpretation) by others? / Should they also be involved in statistical analysis of data and help policy makers in taking decisions? Opinions differ on these issues. I refer to a paper by Subra-

maniam <sup>(4)</sup> who quotes from the records of the British period in India to highlight two different opinions.

Commenting on the role of the Director of Statistics the Indian Industrial Commission of 1918 recorded:

The Director of Statistics should be a compiling officer only. His relation with statistics should be merely / arithmetical, that is, he should not comment on them.

The preparation of forecasts was a highly technical business. The existing practice whereby the Director of Statistics was responsible for amending the forecasts of / provincial officers was disapproved. Such a duty would / more appropriately be performed by a responsible officer having general experience of Indian agriculture.

On the other hand, the Imperial Statistical Conference held in London in 1920 noted:

The arithmetical work of a statistical officer constituted only the preparatory and less important stage of the work of such an office. If this principle be ignored, / the more important services which through statistical / work is capable of rendering the community will be sacrificed.

The above contrary views held more than sixty years ago still persist in spite of the emergence of statistics as a separate discipline and the recognition of statistics as an indispensable tool in scientific investigations, specially in the social sciences, and in decision making.

The distinction between suppliers and users of statistics is an unfortunate one. If it is the responsibility of the statistician to produce useful data which are timely, accurate and without gaps, he must have a good idea of the contemplated uses of data and also be personally involved in data analysis and problem solving. Usefulness of data cannot be thought of in absolute terms. There is nothing like fully accurate data, <sup>(5)</sup>, <sup>(6)</sup> or data without gaps. These issues are relevant only in / terms of their use in solving particular problems.

It is not suggested that government statistical offices should be turned into research departments for producing academic papers. But a statistical office should / have the necessary expertise to do the following:

- (i) To undertake research on methods of collecting data (types of schedules to be used, agency to

be employed to collect data, built in checks, ancillary information useful in detecting / errors and auditing the results).

- (ii) To do minimal statistical analysis which will help in judging the adequacy of given data for taking certain policy decisions, and assessing the errors involved in making forecasts and / predictions. This type of research should be oriented to data improvement, compilation of meaningful summary figures such as index numbers, input-output statistics, analysis of / time series, short term forecasts useful to the government for day-to-day administrative decisions.

The functions of a statistical office should extend beyond collection of data to its analysis which is necessary to hold a statistical system in place, to make / possible most communication with decision makers about their data needs, and to inform them of current statistical capability.

### 3. PUBLIC POLICY MAKING

There is no denying that many policy decisions / made by the government, which effect each one of us, depend crucially on their statistical base. Stretching the point further, one could argue that no / other profession today carries so far ranging a responsibility in public affairs as do we (statisticians). Claus Moser.

How best the government statisticians contribute to national efforts for the benefit of mankind? How can / they help the government, specially in developing countries, in meeting new challenging tasks of removing poverty and ensuring a desirable rate of economic growth?

The efficiency of a government depends on the nature of the decisions it takes. The quality of decisions in turn depends heavily on the availability of timely and accurate data and also, which is equally important, on proper statistical analysis of data. Compilation of data and statistical analysis require considerable skills for the exercise of which the statistician is specially / trained. Thus a statistician has the responsibility and an important role to play in national affairs. To argue that a statistician has no place at the policy making table and that he is merely a purveyor of pure facts is to ignore the whole body of the specialized branch of knowledge known as statistical methodology by which optimal

decisions are arrived at from given data and constraints to achieve desired goals.

Many countries have adequate facilities for training statisticians at various levels, and also well developed national statistical systems in relation to statistical data routinely collected in administration, as well as / in relation to special data collected through censuses and sample surveys. However, in spite of so much statistical expertise and competence, and so much statistical data becoming available, I must confess to a feeling of dissatisfaction that the statistical science and the profession are not giving their due in achieving and promoting a better quantitative understanding of the socio-economic problems and in suggesting and putting forward feasible solutions of the same. The reason for this / serious failure must be traced to lack of interaction between different groups of experts who seem to be / working independently of each other. We have the professional statisticians who collect data without being / aware of their uses, the academic statisticians who develop statistical theory for data analysis without relevance to current real problems and the social scientists / who formulate policy decisions based on data with which they may not be familiar, using statistical analysis / which may not be efficient.

How can we bring these experts together, coordinate their activities and use their specialized knowledge for achieving the best possible results? I am aware that / this can be only a gradual process. I suggest that / efforts should be made in several directions:

- The statistical departments in the universities have remained either allied to departments of mathematics or, where they are independent, they have tended to remain aloof from applied areas in general and social science in particular. In the social science departments the faculty and student body both not only / lack adequate mathematical training but are in fact generally weak in mathematical and quantitative reasoning. Hence a wide gap exists / between statistics and social sciences in the universities. To close this gap, serious / efforts should be made to reorganize the courses in statistics and the social sciences to produce professional statisticians with bias / towards applications and social scientists / with a quantitative bias and competence in / quantitative bias and competence in quantitative analysis.

- Provision should be made to recruit both statisticians and social scientists in statistical offices and policy making divisions of the government to provide a proper mix of personnel with different skills to interact and produce best possible results.
- The possibility of inviting experts from universities and research institutes to work in / government statistical offices and to send / government statisticians to teach in the universities on practical aspects of statistics / should be explored. Such an exchange of experts would have beneficial effects in improving the efficiency and usefulness of government statistical offices and also in the production of statisticians better suited to work in statistical offices after completion of their university education.

Public policy making is no longer a gamble with an unpredictable chance of success or a hit and miss affair. It is now within the realm of scientific techniques, / whereby optimal decisions can be taken on the basis of available evidence and the results continuously monitored for feedback and control.

#### 4. SOME NEW PROBLEMS

*... the stakes involved in economic and social / policies are enormous, and we (statisticians) have a good share of the responsibility.* Harold Wilson.

There are wide areas of social statistics which are neglected as their importance in the context of planning has not been fully understood. In a recent study (7), a comparison was made of the socio-economic profiles of six countries during the period 1950-1965. It was found that presence or absence of rigidities in social structure including divisions based on ethnic differences influence the economic growth. The consideration of social factors has introduced a new dimension to economic planning and places heavier responsibilities on statisticians for evolving an integrated system of economic and social statistics, and conducting studies on the influence of social factors on economic development.

While we have adequate economic statistics to compute magnitudes like the GNP, we do not have even basic data on social aspects such as income and expenditure of households, family size, health and educational status / of individuals, which are necessary for constructing social indicators. I understand that about two out of five

of professional statisticians in the U.K. are now working on social statistics. This ratio was very small five / years ago, which indicates the high priority given to social statistics, in the U.K. in recent years. The use of social indicators as objective functions besides the GNP is far more important for planning in developing countries.

For a long time there has been the dominant belief / that any type of education confers productive skills on the educated -- the only issue to be debated used to be whether the stream of additional returns accruing to the educated because of the acquired skills as compared to / the uneducated was commensurate with the costs of education. In short, in the orthodox view, acquiring education is like any other investment, only it is on the human / being. Thus we have the human capital theory.

Recently, because of the observed realities, an entirely new view contradictory to the human capital has emerged. This is the so called Screening Theory of education as expounded among others by the Nobel Laureate in Economics, Kenneth J. Arrow. An extreme version of this theory holds that education confers no skills - the productive contribution of the individual to the society being dependent only on his inherent ability and not on education. / on the other hand, those with greater ability have a / greater chance of success in obtaining a college degree. This being the case, the college degree serves merely as a filter or screen to separate those with potentially / greater inherent abilities from the rest. Thus, even / though from the society's point of view the college / degree confers no productive advantage and any use of resources in providing college education is a complete / waste, there is obviously positive gain to individuals to acquire a college degree and thereby providing the information to the employers that they have greater inherent ability and thus obtain a higher wage. The advantage to employers is also obvious - they are able to acquire employees with greater inherent abilities by hiring those with college degrees. It is not surprising, therefore, / that there is a substantial and increasing demand for education that governments are forced to meet. The rapid / expansion of education, particularly higher education is a pointer in this direction. Of course, this is an extreme view - some types of education namely technical and vocational education obviously confers skills. But, bulk of the expansion in our higher education took the form of / providing a liberal arts or a general science degree and it is quite likely that the screening theory is relevant here.

It is of great importance to know whether education confers skills or not and what types of education does / confer useful skills. Though attempts have been made in the past to estimate statistically the contribution of / education to the observed differential in the earnings of the educated as compared with the uneducated, most of these calculations are unsatisfactory because it is by no means simple to avoid confounding of the effects of inherent ability and that of education. Any one who can develop statistical tools and techniques to isolate the effect of education will provide a basis for rational / educational policy for a number of developing and poor nations.

Another problem of current interest is the estimation of undernourished persons in the developing countries. The only available data for each country are indirect estimates of per capita food consumption (with an unknown magnitude of error) and empirical income (expedite) distributions with all its unreliabilities. Even if one had reliable estimates of these quantities, it is not possible to obtain proper estimates of the undernourished persons in each country without making some assumptions on calorie intake-income elasticity, distribution models and unknown variability of individual / calorie requirements. Recent estimates of the number of undernourished persons made by the FAO and World Bank / differed considerably since they are based on different sets of assumptions. It is difficult to justify one or the other of the estimates in view of the extreme inadequacy of the data base and the *judgmental assumptions* made by these two organizations. I believe FAO is contemplating a world wide survey to collect data for assessing the extent of undernourishment in the developing / countries. Here is an interesting and extremely important problem where the statisticians and the social / scientists can collaborate in making a useful contribution.

## 5. CONCLUSIONS

In conclusion let me say the following:

The responsibilities of statisticians do not end with

just prveying timely and accurate data. They must have a deep involvement in the analysis and interpretation of data. Analysis of data also enables feedback for data improvement and future data needs.

There should be greater interface between statisticians and social scientists, which can be brought out by suitable changes in the courses given at the universities and in the recruitment policy of government departments / to provide the necessary interaction between statisticians and social scientists.

With emphasis on social indicators as indices of development new dimensions of official statistics have / opened up placing heavier responsibilities on government statisticians.

Statistical knowledge is a national resource and as such should be fully and efficiently exploited.

Good statistics is not a substitute for a good statistician and similarly a good statistician is not a substitute for good statistics. We must have both. But this is not sufficient. It is also necessary that statisticians work alongside policy makers in decision making if the / country is to get full benefits of the available statistical knowledge and skills.

In an address to the delegates of the International Statistical Conference held in 1951, Jawaharlal Nehru, / the then Prime Minister of India, stressed the importance of statistics in understanding the problems of the country and seeking for solutions. He said:

*I find in India, and possibly in other parts of / the world too, an attempt being made to answer questions before framing the questions. That is an extraordinary thing, of course, that everybody is answering questions without knowing what the questions are. In other words, everybody is finding some remedy with out knowing what the malady is.*

Statistics is a gateway to knowledge and the progress a country makes is vitally dependent on the efficiency of its statistical system.

- (1) Claus Moser (1976). The role of the Central Statistical Office in assisting public policy makers. *The American Statistician*, / 30, 59-66.
- (2) In India, the Econometric Society holds periodical seminars on Data Base of Indian Economy and issues proceedings of the papers presented on dataimprovements and needs together with / recommendations for consideration by the government.
- (3) Reference may be made to the article, "A report and comments / on improving the federal statistical system", *The American Statistician*, 35, 183-209 dated November 1981.
- (4) Subramanian, S. (1969), Statistical Wisdom - old or new, *J. / Roy. Statist. Soc. A*, 132, 229-234.
- (5) Weirus, a German physician of the sixteenth century, a time / when most Europe was gripped by the fear of demons and witches calculated that exactly 7.405.926 demons inhabited the earth./ Most people believed that the figure 7.405.926 sounded " about right " for to them demons were a reality and Weirus was a / learned man.
- (6) The story is told about a man who, when asked about the age of a certain river replied that it was 3.000.004 years old. When asked how he could give such an accurate figure, his answer / was that four years ago it was reported that the particular / river was three million years old.
- (7) Report by the U. N. Research Institute for Social Development Geneva, entitled "Levels of Living and Economic Growth".

### A NATURAL EXAMPLE OF A WEIGHTED BINOMIAL DISTRIBUTION

#### 1.- AN EMPIRICAL RESULT

You ask each male student in a class to write down the number of brothers including himself and sisters he has and raise the following question. Let  $k$  be the number of male students asked, and  $B$  and  $S$  be the total numbers of brothers and sisters they have. What would be the approximate value of  $B/(B + S)$ , the ratio of brothers to the total number of children? The general tendency is to say that  $B/(B + S)$  is close to one half or slightly over one half. The answer is, of course, wrong. Surprisingly when the number of boys in the class is not very small you can make a fairly accurate prediction of the relative magnitudes of  $B$  and  $S$  and the ratio  $B/(B + S)$ . This may be stated in the form of an empirical theorem.

**Empirical Theorem:** Let  $k$  boys observed in a classroom or in any gathering have a total number of  $B$  brothers (INCLUDING THEMSELVES) and a total number of  $S$  sisters. Then the following predictions can be made:

- (i)  $B$  is much greater than  $S$ .
- (ii)  $B - k$  is approximately equal to  $S$ .
- (iii)  $(B - k)/(B + S - k)$  is close to one half.
- (iv)  $B/(B + S)$  is larger than one half;  $\left[B/(B + S)\right] - \left[k/(B + S)\right]$  is close to one half.

Before I conduct the experiment I write out the theorem on a piece of paper and give it to one of them in a closed envelope, which is opened after the data are collected for verification. The roles of  $B$  and  $S$  are reversed in the Empirical Theorem if you obtain the data from a number of girls in a classroom or any gathering.

Table 1 gives a summary of data collected from male students in the age range 19-24 years taking courses in statistics in universities in different cities and countries. The agreement with predicted values seems to be good as indicated by the figures in the last three columns, the values in the last two columns being close to one half. Further  $S$  and  $B - k$  are in the ratio 1:1.

A brief outline of the model on which the predictions are based is given in section 2.

#### 2.- THEORETICAL CONSIDERATIONS-SAMPLING PROPORTIONAL TO SIZE

Let us start with a binomial distribution  $b(n, \pi)$  with index  $n$  and probability of success  $\pi$ . The probability of  $r$  successes is

$$\binom{n}{r} \pi^r (1-\pi)^{n-r}, \quad r = 0, 1, \dots, n \quad (2.1)$$

Let us construct another model where the probability of the event  $r$  is altered to

$$cp_r \binom{n}{r} \pi^r (1-\pi)^{n-r}, \quad r = 0, \dots, n \quad (2.2)$$

where  $p_r$  are assigned non-negative numbers and  $c$  is the standardizing constant. Note that the distribution (2.2) is applicable if we carry out what sample survey statisticians call p.p.s. sampling from (2.1) i.e.; giving unequal probabilities to events with different values of  $r$ . In a previous paper (Rao, 1965), the distribution (2.2) was referred to as a weighted binomial distribution. Subsequent developments on such distributions are given in a recent paper by Patil and Ord (1975) and in the references given in their paper.

The case  $p_r \propto r$  is of some interest. In this case, (2.2) reduces to

$$\binom{n-1}{r-1} \pi^{r-1} (1-\pi)^{n-r}, \quad r = 1, \dots, n \quad (2.3)$$

which is again binomial with index  $n$  and observation each reduced by unity. If the model appropriate for our sampling is (2.3), then  $E(r-1) = (n-1)\pi$  giving

$$E(r) = n\pi + (1-\pi) \quad (2.4)$$

and not  $n\pi$  as in the case of (2.1).

Let us interpret the data collected on boys in terms of the new distribution (2.3) with the special choice of  $p_r \propto r$  in (2.2). The distribution of the number of male children in families with a given number of children, say  $n$ , is  $b(n, \frac{1}{2})$ , i.e., binomial with  $\pi = \frac{1}{2}$  (at least approximately). Suppose we choose a family according to some mechanism which does not give equal probabilities to families but probabilities proportional to the numbers of their male children. Then the appropriate model for our observations is (2.3) and not (2.1).

TABLE 1

Verification of the predictions based on data collected on male students in the age range 19-24 years.

Source and date of enquiry		k	B	S	B - K	$\frac{S}{B + S}$	$\frac{2B - k}{2(B + S)}$	$\frac{B - k}{B - k + S}$
Teheran (Iran)	April '75	21	65	40	44	.619	.519	.524
Ispahan (Iran)	April '75	11	45	32	34	.584	.513	.515
Tokyo (Japan)	June '75	50	90	34	40	.726	.524	.540
Delhi (India)	Feb. '75	29	92	66	63	.582	.490	.488
Calcutta (India)	Aug. '63	104	414	312	310	.570	.498	.498
Waltair (India)	June '69	39	123	88	84	.583	.491	.488
Ahmedabad (India)	Aug. '75	29	84	49	55	.632	.523	.529
Bangalore (India)	Aug. '75	55	180	127	125	.586	.496	.497

Thus if  $b$  and  $s$  are numbers of brothers and sisters in a family as reported by a boy, then

$$E(b - 1/b + s) = \frac{b + s - 1}{2} \quad (2.5)$$

where  $E(x/y)$  denotes conditional expectation of  $x$  given  $y$ . If  $(b_i, s_i)$ ,  $i = 1, \dots, k$  are observations on  $k$  boys then, with  $B = \sum b_i$ ,  $S = \sum s_i$ ,

$$E(B - k/B + S) = \frac{B + S - k}{2} \quad (2.6)$$

which implies

$$E\left[\frac{B - k}{B + S - k}\right] = \frac{1}{2} \quad (2.7)$$

and

$$E\left[\frac{B}{B + S} - \frac{k}{2(B + S)}\right] = \frac{1}{2}$$

The two results in (2.7) form the basis of the statements (i)-(iv) of the Empirical Theorem, using the usual limit theorems. We may state the second result in (2.7) in the form

$$\frac{B}{B + S} \text{ is approximately } \left(\frac{1}{2} + \frac{1}{2f}\right) \quad (2.8)$$

where  $f = (B + S)/k$ , the average family size. Thus  $B/(B + S)$  overestimates the true probability by the order of  $(\frac{1}{2}f)$ .

### 3.- A DISCUSSION ON THE MECHANISM OF SELECTION

We proved the empirical theorem on the assumption // that the probability of selection of family is propor-

tional to the number of male children. This is true if our sampling frame is the list of all male children and a family is selected when a male child belonging to it is observed. But our sampling frame is slightly different, // being male children of a certain age group.

Let us designate a male child in this age group by  $M$  and suppose that there can be only one  $M$  in a family. // Then, owing to the nature of our sampling frame, the probability that the selected family has  $r$  male children // out of  $n$  is

$$P(r/M, n) = cP(M/r, n) \binom{n}{r} \pi^r (1 - \pi)^{n-r} \quad (3.1)$$

where  $c = 1/P(M/n)$  and  $P(M/r, n) = P_r$ , say, is the probability that a family with  $r$  male children out of  $n$  has // an  $M$  ( $r = 1, 2, \dots, n$ ). Suppose that we select a male // child at random from the collection of all male children who belong to families having  $r$  male children out of  $n$ . // The probability that the selected child is an  $M$  is  $r^{-1}P_r$ . // If it is reasonable to believe that  $r^{-1}P_r = p_1$  then //  $p_r = rp_1$  in which case model (2.3) is applicable (se- // tting  $p_0 = 0$ ). In the data reported in Table 1. such a mechanism seems to apply.

I have collected similar data from graduate students and faculty members of statistics departments in the age group 25-60 years in three other universities. Pennsylvania State University at State College in the U.S.A., Agricultural University at Poznan (Poland) and University of the Basque Country at Bilbao. In addition, data were also obtained from the delegates to a conference on the // TEACHING OF STATISTICS at Warsaw (Poland). The data and the predicted values according to the Empirical // Theorem are reported in Table 2 for males ////

TABLE 2.

Summary of data collected on male graduate students and faculty members (25-60 years)

Source and date of enquiry		k	B	S	B - k	$\frac{B}{B - S}$	$\frac{2B - k}{2(B + S)}$	$\frac{B - k}{B - k - S}$
Warsaw (Poland)	Ausg. '75	18	41	21	23	.660	.525	.523
Poznan (Poland)	Sept. '75	24	50	17	26	.746	.567	.605
State College (U.S.A.)	Aug. '75	30	87	39	47	.690	.571	.546
Bilbao	Sept. '83	24	95	56	71	.629	.549	.559

TABLE 3.

Summary of data on female graduate students and faculty members (20-40 years)

Source and date of enquiry		k	S	B	S - k	$\frac{S}{B + S}$	$\frac{2S - k}{2(B + S)}$	$\frac{S - k}{B + S - k}$
Poznan (Poland)	Sept. '75	20	42	15	22	.737	.562	.594
State College	Aug. '75	12	34	16	22	.680	.560	.579

and Table 3 for females. It is seen that the values in the last two columns of Tables 2 and 3 are somewhat larger than one half indicating the possibility that sampling probabilities for families are larger than the size ratios in the case of graduate students and faculty/members. This may be expected as the graduate students/ and faculty members are subject to a two stage selection, once when they enter the school and again when they are. At the first stage the probability of selecting a boy from a family with boys out of  $n$  may be proportional to  $x$ . This gives rise to a weighted binomial distribution discussed in section 2. If research fellows and teachers are recruited from those who have completed school and taken a degree from a University, there is likely to be a further selection in favour of families with a larger number of boys. It will be of interest to determine what selection probabilities operate at the two stages.

## REFERENCES

- 1.- FISHER, R.A. (1934): The effect of the methods of ascertainment upon the estimation of frequencies. / Ann. Eugen. London, 6, 13-25.
- 2.- PATIL, G.P. and ORD. J.K. (1975): On size-biased / sampling and related form invariant wighted distributions. Sankhya. Series A, 38.
- 3.- RAO, C. RADHAKRISHNA (1965): On discrete distributions arising out of methods of ascertainment. Classical and Contagious Distributions. G.P. Patil(ed.) S.P.S. Calcutta, pp. 320-332. Also reprinted in Sankhya A. 27, 311-324.

INFERENCE FROM LINEAR MODELS WITH FIXED EFFECTS:  
RECENT RESULTS AND SOME PROBLEMS

## 1. INTRODUCTION

Consider the general Gauss-Markoff model

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 V) \quad (1.1)$$

where the matrices  $X$  and  $V$  may be deficient in rank.// The paper reviews some recent work on inference from the model (1.1) when  $\beta$  and  $\sigma^2$  are unknown, and presents some problems which are of interest in practical applications. Section 2 deals with unified approaches to BLUE (best linear unbiased estimation) applicable to the general case where  $X$  and  $V$  are possibly deficient in rank. Section 3 presents criteria for estimation as alternatives to unbiasedness and minimum variance. Section 5 is devoted to robustness of some inference procedures. Section 6 introduces concepts of linear sufficiency and completeness in linear models without making any distributional assumptions. Section 7 discusses model selection for prediction purposes. Section 8 examines an undesirable feature of empirical Bayes estimators. Topics not explicitly covered are admissibility of linear estimators (Rao, 1976; Klonicki, 1979; La Motte, 1980), robustness of estimators for misspecification of the  $X$  and  $V$  matrices (Rao, 1967a; Zyskind, 1967), and estimation of  $X\beta$  when  $V$  has a specified structure with unknown parameters (Rao and Kleffe, 1980).

The following notations are used throughout the lecture.

- (i)  $p(A)$  denotes the rank of a matrix and  $R(A)$ , the vector space generated by the columns of  $A$ .
- (ii)  $A^-$  is a g-inverse of  $A$  (defined by  $AA^-A = A$  with properties discussed in Rao, 1962).
- (iii)  $Z$  is a matrix of full rank such that  $X'Z = 0$  where  $X$  is the design matrix in the linear model (1.1). We also write  $Z = X^\top$ . Note that  $R(X) \cap R(VZ) = 0$  // which implies
- $$p(V:X) = p(X:VZ), \quad p(VZ) = p(V:X) - p(X) = p(Z'VZ)$$
- (iv)  $P_{A/B}$  is defined as the generalized projection operator on  $R(A)$  along a disjoint subspace  $R(B)$  // although  $R(A) \cup R(B)$  may not be the whole space if

where  $y_A$  is the unique component of  $y \in R(A:B)$  // in the decomposition  $y = y_A + y_B$ ,  $y_A \in R(A)$  and  $y_B \in R(B)$  [see Rao (1974) and Rao and Yanai (1979)]. The  $P_{A/B}$  in (1.3), need only satisfy the conditions not necessarily unique,

$$PA = A, \quad PB = 0. \quad (1.4)$$

The following facts about the model (1.1) are used in the discussion of various results.

(1)  $Y \in R(V:X)$  with probability 1.

(2)  $\beta_1 \neq \beta_2 \Leftrightarrow X\beta_1 \neq X\beta_2$  iff  $X$  is of full rank, // This implies that  $\beta$  is not identifiable if  $X$  is deficient in rank. A linear parametric function  $p'\beta$  is said to be identifiable if  $p'\beta_1 \neq p'\beta_2 \Rightarrow X\beta_1 \neq X\beta_2$  for which a necessary and sufficient condition is // that  $p' \in R(X')$ . This condition is the same as that // for unbiased estimability of  $p'\beta$  by a linear function of the observations. Of course, it does not make sense to estimate nonidentifiable parametric functions.

## 2.- UNIFIED APPROACH TO BLUE

### 2.1.- GENERALIZED PROJECTION

Let  $(Y, X\beta, \sigma^2 V)$  be the linear model where  $X$  and  $V$  may be deficient in rank. From (iii), Section 1,  $R(X) \cap R(VZ) = 0$  so that there exists a projection operator  $P$  on  $R(X)$  along  $R(VZ)$  satisfying the equations

$$PX = X, \quad PVZ = 0. \quad (2.1)$$

Note that, by solving (2.1),

$$P = X(X'V^{-1}X)^{-1}X'V^{-1}, \quad (\text{when } V \text{ is nonsingular}) \quad (2.2)$$

$$= X(X'X)^{-1}X', \quad [\text{when } V = I, \text{ see Rao (1967b)}] \quad (2.3)$$

$$= X(X'TX)^{-1}X'T, \quad (\text{in general}) \quad (2.4)$$

where  $T = (V+XUX')^-$  for any g-inverse, with  $U$  such that  $p(V+XUX') = p(V:X)$ . We have the following general proposition, which answers a question raised by Kempthorne (1976) about projection operators when  $V$  is singular.

**Proposition 2.1** Let  $P$  be a projection operator as defined by (2.1), and  $L$  be any matrix such that  $LX = X$ . Then

$E(PY-X\beta) (PY-X\beta)' \leq E(LY-X\beta) (LY-X\beta)'$   
i.e.,  $PY$  is the minimum dispersion unbiased estimator / (MDUE) of the identifiable vector parameter  $\theta = X\beta$ .

**Proof** Set  $M = L-P$  so that  $MX = 0$  or  $M = AZ'$  for some  $A$ . Then

$$\begin{aligned} E[L(Y-X\beta) (Y-X\beta)' L'] &= (M+P) V (M+P)' \\ &= MVM' + PVP' \geq PVP' = \\ &= E[P(Y-X\beta) (Y-X\beta)' P'] \end{aligned}$$

since  $MVP' = AZ'VP' = 0$  by (2.1). The proposition is proved, and provides the following results.

**Note 1**  $PY$  is unique with probability 1 for any  $P$  satisfying (2.1) since  $Y \in R(X:VZ)$  with probability 1.

**Note 2** If  $p'\beta$  is an identifiable parametric function, then  $p'\beta = \lambda'X\beta$  and  $\lambda'PY$  is MVUE of  $p'\beta$ . This is a simple consequence of  $PY$  being MDUE.

**Note 3** Choosing  $P$  as in (2.4)

$$\lambda'PY = \lambda'X(X'TX)^{-1}X'Ty = p'\hat{\beta}$$

defining  $\hat{\beta} = (X'TX)^{-1}X'Ty$

**Note 4**  $\hat{\beta}$  as defined in Note 3 is a stationary value / of the quadratic form

$$(Y-X\beta)' T(Y-X\beta) \quad (2.5)$$

thus providing  $\hat{\beta}$  as a least squares estimator with respect to  $T$ . Of course, if  $V$  is nonsingular then  $T = V^{-1}$  is one obvious choice, but  $T$  could also be  $(V+XUX')^{-1}$  // where  $U$  is any nonnegative definite matrix.

**Note 5** An unbiased estimator of  $\sigma^2$  is

$$\begin{aligned} \hat{\sigma}^2 &= f^{-1}(Y-X\hat{\beta})' T(Y-X\hat{\beta}) \\ &= f^{-1}Y' T(I-P)Y, \quad f = \rho(V:X)-\rho(X). \end{aligned}$$

**Note 6** With  $\hat{\beta} = (X'TX)^{-1}X'TY$ ,

$$\begin{aligned} V(p'\hat{\beta}) &= \sigma^2 p' [(X'TX)^{-1}U] p \\ \text{Cov}(p'\hat{\beta}, q'\hat{\beta}) &= \sigma^2 p' [(X'TX)^{-1}U] q. \end{aligned}$$

**Note 7** It is shown in Rao (1973b) that the  $T$  in the quadratic form (2.4) must be of the form  $(V+XUX')^{-1}$  if the stationary values provide BLUE-s.

The Proposition 2.1 and the notes provide the geometry of estimation through the concept of generalized projection and the algebra of estimation through a least // squares type technique without making any assumption on

$\rho(X)$  snf  $\rho(V)$ .

It may be mentioned that the case  $\rho(X) < m$  and  $\rho(V) = n$  was first considered by R.C. Bose who showed that in such a case some parametric functions are not unbiasedly estimable.

## 2.2.- LEAST SQUARES WITH RESTRICTIONS

The case  $\rho(V) < n$  was first considered by Goldman/ and Zelen (1964) who reduced the problem to one of least squares theory with restrictions on parameters. A slightly simpler formulation was given by Mitra and Rao (1968).

**Proposition 2.2** Let  $N$  be a matrix of rank  $s = n-\rho(V)$  such that  $N'V = 0$  which implies that  $N'X\beta = N'X\hat{\beta}$  with / probability 1, and  $V^-$  be any g-inverse of  $V$ . Further let  $\hat{\beta}$  such that

$$\min (Y-X\beta)' V^- (Y-X\beta) = (Y-X\hat{\beta})' V^- (Y-X\hat{\beta}) = R_0^2. \quad (2.6)$$

Then:

- (i)  $p'\hat{\beta}$  is the BLUE of  $p'\beta$ ,  $p \in R(X')$
- (ii)  $f^{-1}R_0^2$  is an unbiased estimator of  $\sigma^2$ , where  $f = \rho(V:X) - \rho(X)$ .
- (iii)  $R_0^2 \sim \sigma^2 X^2(\delta)$   $\quad Y \sim N_n(X\beta, \sigma^2 V)$
- (iv) If  $Y \sim N_n(X\beta, \sigma^2 V)$  and  $K'\beta = \omega$  is a specified consistent hypothesis on  $\beta$ , then  $R_1^2 - R_0^2 \sim \sigma^2 X^2(h)$  independently of  $R_0^2 \sim \sigma^2 X^2(\delta)$  where  $h = \rho[D(K'\hat{\beta})]$  and,

$$R_1^2 = \min (Y-X\beta)' V^- (Y-X\beta). \quad (2.7)$$

$$N'X\beta = N'Y$$

$$K'\beta = \omega$$

**Note 1** Proposition 2.2 is essentially least squares theory with restrictions on parameters. Also  $K'\beta = \omega$  is a consistent hypothesis iff  $[D(K'\hat{\beta})]^- [D(K'\hat{\beta})]^- (K'\hat{\beta} - \omega) = K'\hat{\beta} - \omega$ .

**Note 2** Attempts have been made to show that the same values  $R_0^2$  and  $R_1^2$  can be obtained by dropping the // restrictions  $N'X\beta = N'Y$  in the expressions (2.6) and / (2.7) and choosing a special g-inverse of  $V$ . In Rao // (1978), it is shown that this is not possible.

## 2.3.- INVERSE PARTITIONED MATRIX (IPM) APPROACH

Another unified approach called the IPM was developed in Rao (1971) for the general case where  $V$  and //  $X$  may be deficient in rank.

*Proposition 2.3* Let

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$$

for any g-inverse. Then the following hold:

- (i) [Use of  $C_2$  and  $C_3$ ] The BLUE of any identifiable / function  $p'\beta$  is  $\hat{p}'\hat{\beta}$  where

$$\hat{\beta} = C_2 Y \text{ or } C_3 Y \text{ (which may not be the same).}$$

The condition for estimability of  $p'\beta$  is

$$p'C_3 X = p' \text{ or } p'C_2 X = p'$$

- (ii) [Use of  $C_4$ ] The dispersion matrix of  $\hat{\beta}$  is  $\sigma^2 C_4$  in the sense that

$$V(p'\hat{\beta}) = \sigma^2 p'C_4 P, \text{ cov}(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'C_4 q = \sigma^2 q'C_4 p.$$

- (iii) [Use of  $C_1$ ] An unbiased estimator of  $\sigma^2$  is

$$f^{-1} Y' C_1 Y \text{ where } f = R(V:X) - R(X).$$

*Proposition 2.4* Let  $\hat{S}'\hat{\beta}$  be the vector of BLUE's of  $k$  estimable parametric functions  $S'\beta, R_0^2 = Y' C_1 Y$  and  $f$  be as defined in (2.6). If  $Y \sim N_n(X\beta, \sigma^2 V)$ , then:

- (i)  $\hat{S}'\hat{\beta}$  and  $R_0^2$  are independently distributed with

$$S'\hat{\beta} \sim N_k(S'\beta, \sigma^2 G), R_0^2 \sim \sigma^2 X^2(f)$$

where  $G = S'C_4 S$ .

- (ii) Let  $S'\beta = \omega$  be a null hypothesis is consistent / iff

$$GG^- u = u, u = p'\hat{\beta} - \omega$$

in which case the test statistic

$$F = \frac{u'G^- u}{h} + \frac{R_0^2}{f}, h = p(G)$$

has central F distribution on  $h$  and  $f$  degrees of freedom.

It may be noted that a general approach to minimum / variance unbiased estimation when  $X$  and  $V$  are possibly / deficient in rank is also given by Zyskind(1967) through certain linear transformations of the model. In Section 6, we develop a more detailed transformation of this // kind and discuss the estimation of parameters through / the concept of linear sufficiency.

### 3.- ALTERNATIVE CRITERIA OF ESTIMATION

Questions have been raised about the appropriateness of the conditions of unbiasedness and minimum variance / in linear estimation. We may consider alternative criteria which may have an intuitive appeal. For this purpose let  $X$  and  $V$  be of full rank (the general case is discussed in Section 6). Let  $P\Delta Q'$  be the singular value // decomposition of  $V^{-1/2}X$ , where  $P$  is  $n \times m$  semi orthogonal matrix. If  $R$  of order  $n \times (n-m)$  is the orthogonal complement of  $P$ , then the model  $(Y, X\beta, \sigma^2 V)$  can be transformed to

$$y_1 = P'V^{-1/2}Y = \Delta Q'\beta + P'V^{-1/2}\varepsilon, \quad (3.1)$$

$$y_2 = R'V^{-1/2}Y = R'V^{-1/2}\varepsilon. \quad (3.2)$$

Writing  $\theta = \Delta Q'\beta$ , the model  $(Y, X\beta, \sigma^2 V)$  is equivalent to two uncorrelated models

$$(y_1, \theta, \sigma^2 I_m) \text{ and } (y_2, 0, \sigma^2 I_{n-m}). \quad (3.3)$$

Under the assumptions on the linear model  $(Y, X\beta, \sigma^2 V)$ ,  $y_2$  has no information on  $\theta$  either in its expectation or // through its correlation with  $y_1$ , and is thus redundant / for the estimation of  $\theta$ . We may lay down the following / criteria, for the estimation of an identifiable parametric function  $p'\beta$ . Note that  $p'\beta$  can be equivalently written as a linear function of  $\theta$ , say  $q'\theta$ .

- (i) The estimator of  $q'\theta (= p'\beta)$  is a function of  $y_1$  only / say  $f(y_1)$ .

- (ii) The estimator  $f(y_1)$  is MC (method consistent to the concept of Fisher consistency), i.e., if  $y_1$  has no / error then  $f(y_1) = q'\theta$ , i.e., the estimator reduces to the true value.

The condition (ii) implies that  $f(\theta) = q'\theta \forall \theta \in R^m$ , i.e.,  $f(y_1) = q'y_1$ , which is the least squares estimator of  $q'\theta = p'\beta$ . Thus, we arrive at least squares estimators without appealing to linearity of estimator, unbiasedness or minimum variance. (A solution on the above lines was / given in Example 4, p. 309 of Rao, 1973a).

### 4.- SIMULTANEOUS ESTIMATION AND PREDICTION

#### 4.1.- SIMULTANEOUS ESTIMATION

Let us consider  $k$  linear models

$$Y_i = X\beta_i + \varepsilon_i, i = 1, \dots, k \quad (4.1)$$

where  $y_i$  is an  $n$ -vector,  $X$  is a given  $n \times m$  matrix, and  $(\beta_i, \varepsilon_i)$  are independent and identically distributed random variables such that

$$E \begin{bmatrix} \beta_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}, \quad D \begin{bmatrix} \beta_i \\ \varepsilon_i \end{bmatrix} = \begin{bmatrix} F & 0 \\ 0 & \sigma^2 V \end{bmatrix} \quad (4.2)$$

where  $\beta$ ,  $F$  and  $\sigma^2$  are unknown. The problem is one of // simultaneous prediction (estimation) of  $p'\beta_i$ ,  $i = 1, \dots, k$ , for purposes of comparison. An early example of such a problem of estimation (which may be called empirical / Bayes) is the construction of selection index in genetics suggested by R.A. Fisher and developed by Fairfield Smith (1936), Henderson (1950), Panse (1946) and Rao // (1952, 1953).

**Proposition 4.1** Let  $\beta_i^{(l)}$  be the least squares estimator of  $\beta_i$  from the  $i$ -th model and denote  $U = (X'V^{-1}X)^{-1}$ . If  $\sigma^2$ ,  $\beta$  and  $F$  are known, then the optimum linear predictor of  $p'\beta_i$  is  $p'\beta_i^{(b)}$  (i.e., for which mean square/prediction error (MSPE) is a minimum) where

$$\beta_i^{(b)} = \beta_i^{(l)} - \sigma^2 U(F + \sigma^2 U)^{-1}(\beta_i^{(l)} - \beta) \quad (4.3)$$

with the MSPE  $p'Q^{(b)}p$  where

$$Q^{(b)} = \sigma^2 U - \sigma^4 U(F + \sigma^2 U)^{-1}U \quad (4.4)$$

**Note 1**  $\beta_i^{(b)}$  is the Bayes estimator of  $\beta_i$  under the assumptions (4.2).

**Note 2** . The mean dispersion error for  $\beta_i^{(b)}$ ,  $i=1, \dots, k$ , is

$$K^{-1} \sum_1^k E(\beta_i^{(b)} - \beta_i)(\beta_i^{(b)} - \beta_i)' = \sigma^2 U - \sigma^4 U(F + \sigma^2 U)^{-1}U \quad (4.5)$$

which is smaller than  $Q^{(l)} = \sigma^2 U$ , the mean dispersion // error for  $\beta_i^{(l)}$ ,  $i = 1, \dots, k$ .

When  $\sigma^2$ ,  $\beta$  and  $F$  are unknown we may estimate // them by

$$k\beta_* = \sum_1^k \beta_i^{(l)}, \quad (4.6)$$

$$k(n-m)\sigma_*^2 = \sum_1^k (Y_i'V^{-1}Y_i - Y_i'V^{-1}X\beta_i^{(l)}) = W, \quad (4.7)$$

$$(k-1)(F_* + \sigma_*^2 U) = \sum_1^k (\beta_i^{(l)} - \beta_*)(\beta_i^{(l)} - \beta_*)' = B. \quad (4.8)$$

substituting a constant multiple of these estimators for  $\sigma^2$ ,  $\beta$  and  $F$  in (4.3), we obtain an empirical Bayes estimator of  $\beta_i$ .

$$\beta_i^{(e)} = \beta_i^{(l)} - c W U B^{-1} (\beta_i^{(l)} - \beta_*), \quad i = 1, \dots, k. \quad (4.9)$$

We determine  $c$  by minimizing

$$E \sum_1^k (\beta_i^{(e)} - \beta_i)(\beta_i^{(e)} - \beta_i)' \quad (4.10)$$

under the assumption that  $\varepsilon_i$  and  $\beta_i$  have independent multivariate normal distributions. The following proposition is proved in Rao (1975).

**Proposition 4.2** Let  $\beta_i$  and  $\varepsilon_i$  have multivariate normal distributions with parameters as in (4.2). Then  $B$  // and  $W$  as defined in (4.7) and (4.8) are independently / distributed with

$$W \sim \sigma^2 X^2(kn-km), \quad B \sim W_m(k-1, F + \sigma^2 U)$$

and the optimum value of  $c$  which minimizes (4.10) is

$$c = (k-m-2) / (kn-km+2)$$

and, with this choice of  $c$ , the MDE for  $\beta_i^{(e)}$  is

$$Q^{(e)} = \sigma^2 U - \frac{\sigma^4 (n-m) (k-m-2)}{k(n-m)+2} U (F + \sigma^2 U)^{-1} U \quad (4.11)$$

provided  $k \geq m + 2$ .

**Note 1.** We have the inequality

$$Q^{(l)} > Q^{(e)} > Q^{(b)} \quad (4.12)$$

**Note 2.** It may be easily shown that the expectation / of (4.10) for fixed values of  $\beta_1, \dots, \beta_k$  is smaller than /  $\sigma^2 U$ . Thus the empirical Bayes estimators (4.9) are uniformly better than the least squares estimators in the sense of having a smaller mean dispersion error. Thus we have the Stein effect in the simultaneous estimation of vector parameters  $\beta_1, \dots, \beta_k$ . The actual expectation of (4.10) for fixed  $\beta_1, \dots, \beta_k$  is expressed by C.G. Khatri / in the form

$$\sigma^2 U - \frac{\sigma^4 (n-m) (k-m-2)}{k(n-m)+2} E(U B^{-1} U). \quad (4.13)$$

Note 3. Results of a similar nature were obtained / by Efrom (1973) and Efron and Morris (1972).

#### 4.2.- SIMULTANEOUS PREDICTION

Let us consider the  $i$ -th model  $y_i = X\beta_i + \varepsilon_i$ , and a future observation  $y_i$  with the structure

$$y_i = x'\beta_i + n_i$$

$$D(\varepsilon_i) = \sigma^2 \begin{pmatrix} V & A \\ A' & B \end{pmatrix}, \quad i = 1, \dots, k, \quad (4.14)$$

where  $B$  is a scalar,  $A$  is a vector and  $\beta_1, \dots, \beta_k$  are fixed vector parameters. If  $g_i(y_1, \dots, y_k)$  is a predictor of  $y_i$ , then under the assumption of multivariate normality (or simply linearity of regression of  $y_i$  on  $y_i$ )

$$\begin{aligned} \sum_{i=1}^k (y_i - g_i)^2 &= E \sum_{i=1}^k [y_i - x'\beta_i - A'V^{-1}(y_i - X\beta_i)]^2 \\ &\quad + E \sum_{i=1}^k [p'\beta_i - h_i(y_1, \dots, y_k)]^2 \end{aligned} \quad (4.15)$$

where  $h_i = g_i - A'V^{-1}y_i$  and  $p' = x' - A'V^{-1}x$ . In order to minimize the compound loss (4.15) in simultaneous // prediction we need only minimize the second term on the right hand side of (4.15), which is the problem of simultaneous estimation of  $p'\beta_i$ ,  $i = 1, \dots, k$ . Using Proposition 4.2 and result (4.13), we have the following.

*Proposition 4.3* Let  $\beta_i^{(e)}$  and  $\beta_i^{(l)}$  be as defined in (4.9). Then

$$\begin{aligned} E \sum_{i=1}^k (y_i - [x'\beta_i^{(e)} + A'V^{-1}(y_i - X\beta_i^{(e)})])^2 \\ \leq E \sum_{i=1}^k (y_i - [x'\beta_i^{(l)} + A'V^{-1}(y_i - X\beta_i^{(l)})])^2 \end{aligned} \quad (4.16)$$

Note The result (4.16) shows that

$$x'\beta_i^{(e)} + A'V^{-1}(y_i - \beta_i^{(e)}) \quad (4.17)$$

is a better predictor of  $y_i$  than the least squares predictor

$$x'\beta_i^{(l)} + A'V^{-1}(y_i - X\beta_i^{(l)}), \quad (4.18)$$

under the compound loss function (4.15).

#### 4.3.- NONHOMOGENEOUS RIDGE ESTIMATOR

Suppose that we have just one model

$$Y + X\beta + \varepsilon, \quad D(\varepsilon) = \sigma^2 V,$$

where  $Y$  is an  $n$ -vector and  $\beta$  is an  $m$ -vector. Assuming that  $\beta$  is a random variable with

$$E(\beta) = \gamma = g(1, \dots, 1)' \text{ and } D(\beta) = \sigma^2 k^{-1} I$$

we can, using (4.3), write down the Bayes estimator of  $\beta$

$$\begin{aligned} \beta^{(B)} &= \beta^{(L)} - U(U + k^{-1}I)^{-1}(\beta^{(L)} - \gamma) \\ &= \gamma + (X'V^{-1}X + kI)^{-1}X'V^{-1}(Y - X\gamma). \end{aligned} \quad (4.19)$$

With  $Y = 0$ , the estimator (4.19) is the ridge estimator of Hoerl and Kennard (1970). Since  $\gamma$  and  $k$  are unknown, an empirical version is recommended in Rao (1977).

$$\beta^{(e)} = \underline{1}' \hat{\beta} + (X'V^{-1}X + \hat{k}I)^{-1}X'V^{-1}(Y - X\underline{1}'\hat{\beta})$$

where

$$\begin{aligned} \hat{\beta} &= \underline{1}'U^{-1}\beta^{(L)} \div \underline{1}'U^{-1}\underline{1}, \quad \underline{1}' = (1, \dots, 1) \\ \hat{k} &= a \div b \end{aligned}$$

where  $a$  and  $b$  are determined from the equations

$$\begin{aligned} (n-m+2)a &= (Y - X\beta^{(L)})' V^{-1} (Y - X\beta^{(L)}) \\ (m-3)[\frac{b}{m-1} (\text{tr } U^{-1} - \frac{\underline{1}'U^{-2}\underline{1}}{\underline{1}'U^{-1}\underline{1}}) + a] \\ &= (\beta^{(L)} - \underline{1}'\hat{\beta})' U^{-1} (\beta^{(L)} - \underline{1}'\hat{\beta}). \end{aligned}$$

The estimator  $\beta^{(e)}$  may not be better than  $\beta^{(L)}$  in terms of compound quadratic loss function, but it may be a // good competitor to the usual ridge estimator.

#### 5.- ROBUSTNESS OF INFERENCE PROCEDURES

Let  $Y \sim N(X\beta, \sigma^2 V)$ . Then it is well known that the likelihood ratio test (LRT) for testing the hypothesis //  $A\beta = 0$  is

$$\lambda(X, V) = \frac{Y' (I - P_{X, V-1}) Y}{Y' (I - P_{X_0, V-1}) Y} \quad (5.1)$$

It is also the uniformly most powerful invariant (UMPI) / test under a suitable group of transformations. The following proposition due to Khatri (1980), and Mathew and Bhimasankaram (1982) gives the conditions under which /

$$\lambda(X, V) = \lambda(X, I).$$

**Proposition 5.1** The likelihood ratio tests for the / hypothesis  $A\beta = 0$  under the models  $Y \sim N(X\beta, \sigma^2 V)$  and /  $Y \sim N(X\beta, \sigma^2 I)$  are the same iff any one of the following equivalent conditions hold .

(i)  $V$  is of the form

$$V = I + X\Lambda_1 X' + (s-1)ZZ' + X_0\Lambda_3 Z' + Z\Lambda_3' X_0' \quad (5.2)$$

where  $\Lambda_1$  and  $\Lambda_3$  are arbitrary matrices,  $s$  is an arbitrary positive real number subject to  $V$  being p.d. and  $Z'X\Lambda_1 X'Z = (s-1)I_k$  and  $X_0 = X(I - A'A)$  as defined/ in Proposition 5.1.

$$(ii) (I - P_{X_0}) V (I - P_{X_0}) = a(I - P_{X_0}) \text{ for some } a > 0.$$

$$(iii) V^{-1}(I - P_{X_0}, V^{-1}) = a(I - P_{X_0}) \text{ for some } a > 0$$

$$(iv) \begin{bmatrix} I - P_X \\ \vdots \\ L - P_X \end{bmatrix} (V - aI) (I - P_X: P_X L') = 0 \text{ for some } a > 0, / \text{ with } A = LX$$

Finally, we have two other propositions proved by / Mathew and Bhimasankaram (1982).

**Proposition 5.2** The LRT's for the hypothesis  $A\beta = 0$  under the models  $Y \sim N(X\beta, \sigma^2 I)$  and  $Y \sim N(U\beta, \sigma^2 V)$  are the same iff

$$R(X) = R(U) \text{ and } R(X_0) = R(U_0). \quad (5.3)$$

**Proposition 5.3** The LRT's for the hypothesis  $A\beta = 0$  under the models  $Y \sim N(X\beta, \sigma^2 I)$  and  $Y \sim N(U\beta, \sigma^2 V)$  are // same iff (5.2) and (5.3) hold simultaneously.

The following result by Ghosh and Sinha (1980)(see also Sinha and Mukhopadhyay (1980) establishes robustness of the LRT for departure from normality.

**Proposition 5.4** Let  $Y = X\beta + \sigma \varepsilon$  with  $\varepsilon$  distributed according to the density

$$f(\varepsilon) = \int_0^\infty e^{-\tau\varepsilon^2/2} (\tau/2\pi)^{n/2} dL(\tau).$$

Then for testing the hypothesis  $A\beta = 0$ , the F based on  $\lambda(X, I)$  is both LRT and UMPI.

Recently, Sinha and Drygas (1982) generalized this / result to the following:

**Proposition 5.5** Let  $Y = X\beta + \sigma \varepsilon$  with  $\varepsilon$  distributed according to a density function  $g(\varepsilon|\varepsilon)$  where  $g$  is a convex non-increasing function. Then the F-test based on  $\lambda(X, I)$  is both LRT and UMPI.

This result is similar to a robustness property of / Hotelling's  $T^2$ -test proved by Karia (1981) and is based/ on a representation theorem due to Wijsman (1967).

Under a slightly more general distribution of the / errors, Sinha and Drygas (1982) established the fo- // llowing property of the BLUE of  $A\beta$ .

**Proposition 5.6** Let  $Y = X\beta + \sigma \varepsilon$  with  $\varepsilon$  having a spherically symmetric distribution. Then for any real number  $c$  and any non null diagonal matrix  $C$  of appropriate order

$$P \{ (GY - A\beta)' C (GY - A\beta) \leq c^2 \} \geq P \{ (LY - A\beta)' C (LY - A\beta) \leq c^2 \}$$

where  $GY$  is the BLUE of estimable  $A\beta$  and  $LY$  is any unbiased estimator of  $A\beta$ .

## 6.- CANONICAL TRANSFORMATION

Consider the model  $(Y, X\beta, \varepsilon)$ ,  $D(\varepsilon) = \sigma^2 V$  where  $X$  and  $V$  may be deficient in rank. Let  $V$  be  $n \times n$  matrix of rank  $v$ . Then there exists a transformation  $(W:N)$  where  $W$  is  $n \times v$  and  $N$  is  $n \times (n-v)$  matrices such that  $N'W = 0$ ,  $W'VW = I_v$ . Under such a transformation the model  $(Y, X\beta, \sigma^2 V)$  becomes.

$$N'Y = Y_3 = N'X\beta + \varepsilon_3, D(\varepsilon_3) = 0 \quad (6.1)$$

$$W'Y = Y_5 = W'X\beta + \varepsilon_5, D(\varepsilon_5) = \sigma^2 I \quad (6.2)$$

of which one part (6.1) is singular and the other (6.2)/ is regular (non-singular). We make a further transformation.

$$Y_3 = N'X\beta + \varepsilon_3, D(\varepsilon_3) = 0 \quad (6.3)$$

$$Y_5 - \Lambda Y_3 = Y_4 = (W' - \Lambda N')X\beta + \varepsilon_4, D(\varepsilon_4) = \sigma^2 I \quad (6.4)$$

where  $\Lambda$  is such that  $(W' - \Lambda N') X X' N = 0$ . The model (6.3), (6.4) can be written as

$$Y_3 = X_3\beta + \varepsilon_3, D(\varepsilon_3) = 0 \quad (6.5)$$

$$Y_4 = X_4\beta + \varepsilon_4, D(\varepsilon_4) = \sigma^2 I \quad (6.6)$$

where  $R(X'_3) \cap R(X'_4) = 0$ , which provides two distinct uncorrelated linear models with different parameters, one of which is singular. Note that if  $R(X) \subset R(V)$ , then //  $X_3 = 0$  and  $Y_3 = 0$  with probability 1, so that we have only the regular.

Let  $\rho(X_3) = x_1$  and  $\rho(X_4) = x_2$ , then further transformations can be made on each model (in the usual way, Rao, 1973a, p. 189) to obtain the canonical form of  $(Y, X\beta, \sigma^2 V)$  in four distinct submodels.

$$Y_{-1} = \varepsilon_{-1}, \quad D(\varepsilon_{-1}) = 0, \quad Y_{-1} \text{ is } (n-v-x_1)-\text{vector} \quad (6.7)$$

$$Y_1 = X_1\beta + \varepsilon_1, \quad D(\varepsilon_1) = 0, \quad X_1 \text{ is } x_1 \times m \text{ matrix} \quad (6.8)$$

$$Y_2 = X_2\beta + \varepsilon_2, \quad D(\varepsilon_2) = \sigma^2 I, \quad X_2 \text{ is } x_2 \times m \text{ matrix} \quad (6.9)$$

$$Y_0 = \varepsilon_0, \quad D(\varepsilon_0) = \sigma^2 I, \quad Y_0 \text{ is } (v-x_2)-\text{vector} \quad (6.10)$$

Note that

$$R(X'_1) \cup R(X'_2) = R(X') \text{ and } R(X'_1) \cap R(X'_2) = 0 \quad (6.11)$$

so that  $\theta_1 = X_1\beta$  and  $\theta_2 = X_2\beta$  are distinct sets of distinct parameters which are equivalent to  $X\beta$ .

Once the observations are known, we can make a further conditional transformation on the sub model (6.8) / to obtain

$$(Y'_1)' Y_1 = 0 = Y_{11} = (Y'_1)' X_1\beta = X_{11}\beta \text{ with probability 1} \quad (6.12)$$

$$Y'_1 Y_1 = Y_{12} = Y'_1 X_1\beta = X_{12}\beta \text{ with probability 1} \quad (6.13)$$

where  $Y'_1$  is a matrix which is the orthogonal complement/ of  $Y_1$ ,

The canonical form (6.7)-(6.10) together with (6.12) and (6.13) sets out the information in the observations/ in a suitable framework for discussing problems of inference.

For instance the following results easily follow.

#### Proposition 6.1

(i)  $y_1$  is the BLUE of  $\theta_1 = X_1\beta$  with a null dispersion/ matrix.

(ii)  $y_2$  is the BLUE of  $\theta_2 = X_2\beta$  with a null dispersion matrix  $\sigma^2 I$ .

(iii)  $y'_0 y_0 / (v-x_2)$  is an unbiased estimator of  $\sigma^2$ .

(iv) The entire class of BLUE's of linear parametric // functions is

$$\{m'_1 Y_1 + m'_2 Y_2\} \quad (6.14)$$

(v) Since  $\theta_1$  is estimable without any error, the problem reduces to that of estimating  $\theta_2 = X_2\beta$ . For this purpose, we may take the view that only  $y_2$  is relevant (sufficient), since the other variables are uncorrelated with  $y$  and have expectations independent of  $\theta_2$  (see Barnard, 1963 and Rao, 1965, 1973, p. 308, example 4). // Then we may estimate  $\theta_2$  by  $\delta(Y_2)$  a function of  $y_2$  only. Under a condition similar to Fisher consistency, viz., /  $\delta(Y_2)$  reduces to  $\theta_2$  if the error component  $\varepsilon_2 = 0$  (i.e., / when no error is made), we find that  $\delta(\theta_2) = \theta_2 V \theta_2 \in R^{\delta-t}$  which implies that  $\delta(Y_2) = y_2$ , the BLUE of  $\theta_2$ .

The canonical form (6.7)-(6.10) and (6.12)-(6.13) also enables us to discuss concepts such as linear sufficiency in a simple way. We consider the following definitions arising out of the result (v) of Proposition 6.1.

**Definition 6.1** (Extended form of Barnard's definition) A statistic  $L'Y$  is linearly sufficient (l.s.) for /  $K'\theta$  (identifiable set of parametric functions) iff the / expectation of any linear function of  $Y$  uncorrelated // with  $L'Y$  does not have an expectation of the form  $\lambda'k\theta$  where  $K'\lambda \neq 0$ .

**Definition 6.2** A statistic  $M'Y$  is linearly minimal / sufficient (l.m.s.) for  $K'\theta$  iff  $M'Y$  is according to Definition 6.1 and for every linearly sufficient statistic /  $L'Y$ , there exists a transformation  $C_L$  such that  $M'Y = C_L L'Y$

The following proposition is easily established.

**Proposition 6.2** The BLUE of  $K'\theta$  is linearly minimal / sufficient for  $K'\theta$  in the sense of Definition 6.2.

An alternative definition of sufficiency is as follows.

**Definition 6.3** A statistic  $L'Y$  is linearly sufficient for  $X\beta$  iff every linear admissible estimator (in the sense of Section 4) of every identifiable linear parametric function of  $\beta$  is a linear function of  $L'Y$ .

**Definition 6.4** A statistic  $M'Y$  is linearly minimal / sufficient for  $X\beta$  iff  $M'Y$  is linearly sufficient for  $X\beta$  according to Definition 6.3 and for every linearly sufficient statistic  $L'Y$ , there exists a transformation  $C_L$  such that  $M'Y = C_L L'Y$ .

The following proposition is easily established.

**Proposition 6.3** The statistic  $(Y_1, Y_2)$ , which is the same as the BLUE of  $X\beta$ , is linearly minimal sufficient for  $X\beta$  in the sense of Definition 6.4, where  $Y_1, Y_2$  are the linear functions of  $Y$  as defined in the canonical representation (6.7)-(6.10).

In recent papers, Drygas (1983) and Muller (1982) used the following definition of linear sufficiency due to Bakasalary and Kala (1978).

**Definition 6.5** A statistic  $L'Y$  is linearly sufficient for  $X\beta$  iff the BLUE of  $X\beta$  is of the form  $D_L L'Y$ .

**Definition 6.6** A statistic  $M'Y$  is linearly minimal / sufficient for  $X\beta$  iff  $M'Y$  is linearly sufficient and  $M'Y = C_L L'Y$  for any linearly sufficient statistic  $L'Y$ .

The following proposition is easily established.

**Proposition 6.4** The statistic  $(Y_1, Y_2)$  is linearly minimal sufficient for  $X\beta$  in the sense of Definition 6.6

Necessary and sufficient conditions for  $L'Y$  to be linearly sufficient in the sense of any of the Definitions 6.1 and 6.5 are discussed in a paper by Rao and Sinha which also provides a clarification of the results obtained by Drygas (1983a, 1983b), Muller (1982) and Seeley (1978).

## 7.- MODEL SELECTION FOR PREDICTION

Usually model selection is based on some criterion / of goodness of fit of a model to observed data, such as Akaike's information criterion and Mallows  $C_p$ . A model / that fits the data adequately may not be good for prediction beyond the range of the observed values of predictors (extrapolation) as in the prediction of future growth on the basis of past observations. For instance / consider a growth model

$$y_t = \beta_0 \psi_0(t) + \dots + \beta_k \psi_k(t) + \epsilon_t, \quad t = 1, \dots, n \quad (7.1)$$

where  $\psi_0, \dots, \psi_k$  are orthogonal polynomials,  $\epsilon_1, \dots, \epsilon_n$  are uncorrelated errors and  $V(\epsilon_i) = \sigma^2$ . An important problem is the choice of the degree of the polynomial in (7.1) for use in predicting a future observation, say at  $t = n+1$ . If we use the maximum possible degree  $k$ , then the best predictor is

$$\hat{y}_{n+1} = b_0 \psi_0(n+1) + \dots + b_k \psi_k(n+1) \quad (7.2)$$

where  $b_i$  are least squares estimators such that  $V(b_i) = \sigma^2$ ,  $\text{cov}(b_i, b_j) = 0$ . The mean square error of prediction is

$$\sigma^2 [1 + \psi_0^2(n+1) + \dots + \psi_k^2(n+1)] \quad (7.3)$$

If we use only  $(p+1)$  terms in the model (7.1), then the mean square error of prediction is

$$\sigma^2 [1 + \sum_0^p \psi_i^2(n+1)] + [\sum_{p+1}^k \beta_i \psi_i(n+1)]^2 \quad (7.4)$$

which can be estimated by

$$\hat{\sigma}^2 [1 + \sum_0^p \psi_i^2(n+1)] + [\sum_{p+1}^k b_i \psi_i(n+1)]^2 - \hat{\sigma}^2 \sum_0^k \psi_i^2(n+1) \quad (7.5)$$

$$= \hat{\sigma}^2 [1 + \sum_0^p \psi_i^2(n+1)] + [\sum_{p+1}^k b_i \psi_i(n+1)]^2 - \hat{\sigma}^2 \sum_0^k \psi_i^2(n+1)$$

where  $\hat{\sigma}^2$  is the least squares estimator of  $\sigma^2$  based on the full model. From (7.5), we find that the appropriate choice of  $p$  is that value for which

$$2 \sum_{i=0}^p \psi_i^2(n+1) + [\sum_{i=p+1}^k b_i \psi_i(n+1)]^2 \quad (7.6)$$

is a minimum. It may be seen by comparing (7.3) and // (7.4), that even when  $b_{p+1}, \dots, b_k$  are not all zero, the choice of a  $p$ -th degree polynomial (i.e., a wrong model) provides better prediction if

$$\left( \sum_{i=p+1}^k b_i \psi_i(n+1) \right)^2 \leq \sigma^2 \sum_{i=p+1}^k \psi_i^2(n+1) \quad (7.7)$$

Of course, the best wrong model corresponds to the value of  $p$  for which

$$\sigma^2 [\sum_{i=0}^p \psi_i^2(n+1)] + [\sum_{i=p+1}^k b_i \psi_i(n+1)]^2 \quad (7.8)$$

is a minimum, and the criterion (7.5) is an empirical // version of (7.8).

A criterion of the type (7.6) for choice of variables in the general linear model  $y = X\beta + \epsilon$  for predicting at a chosen value (or chosen values) of the predictor variables can be written down in a similar // manner. For acutal expressions and a more detailed study, the reader is referred to Rao (1983).

#### 8.- DIRECT OR INVERSE REGRESSION?

Suppose that  $\{t, m_1, m_2\}$  represent the true  $\{\tau\}$  and // repeated observed measurements  $m_1$  and  $m_2$  of blood pressure (BP) on an individual. Given  $m_1$  and  $m_2$ , and a plausible model

$$\begin{aligned} m_1 &= t + \epsilon_1 \\ m_2 &= t + \epsilon_2 \end{aligned} \quad (8.1)$$

where the errors  $\epsilon_1, \epsilon_2$  are such that  $E[\epsilon_i] = 0, V[\epsilon_i] = \sigma^2 \epsilon, \text{cov}(\epsilon_1, \epsilon_2) = 0$ , how do we specify the rule for estimating  $t$ , the unobservable true BP, in routine practice (say in a clinic)?

We may consider (81) as a linear model involving an unknown parameter  $t$ , and estimate it by the least squares estimator,  $\bar{m} = (m_1 + m_2)/2$ , which is a well known rule.

On the other hand, we may consider  $\{t, m_1, m_2\}$  as three measurements on an individual, one of which is unobservable, and try to predict  $t$  by the regression of  $t$  on  $m_1$  / and  $m_2$ , computed from the distribution of  $\{t, m_1, m_2\}$  in the population of clinic visiting individuals. If  $\tau$  and  $\sigma_t^2$  are the mean and variance of  $t$  among the individuals, then the regression of  $t$  on  $m_1$  and  $m_2$  is

$$\hat{t} = \tau + \frac{2\sigma_t^2}{2\sigma^2 + \sigma_\epsilon^2} (\bar{m} - \tau). \quad (8.2)$$

The unknown parameters  $\tau, \sigma_t^2$  and  $\sigma_\epsilon^2$  can be estimated/ from past data (see Rao, 1973a, p. 337), and can be updated when fresh data become available. Substituting the / estimates for the unknowns in (8.2), we may specify the rule of estimation of  $t$  as

$$\hat{t} = \hat{\tau} + \frac{2\hat{\sigma}_t^2}{2\hat{\sigma}_t^2 + \hat{\sigma}_\epsilon^2} (\bar{m} - \hat{\tau}) \quad (8.3)$$

The expression (8.3) may also be recognized as the / empirical Bayes estimator (EBE). Ignoring errors in //  $\hat{\tau}, \hat{\sigma}_t^2$  and  $\hat{\sigma}_\epsilon^2$ , we find

$$\begin{aligned} E(\hat{t} - t)^2 &= \frac{\sigma_\epsilon^2}{2} \frac{2\sigma_t^2}{2\sigma_t^2 + \sigma_\epsilon^2} \\ &< \frac{\sigma_\epsilon^2}{2} = E(\bar{m} - t)^2 \end{aligned} \quad (8.4)$$

so that the overall mean square error (averaged over // all future predictions) is smaller for  $\hat{t}$  than for  $\bar{m}$ . // Should we, on this account, recommend  $\hat{t}$  as a better rule than  $\bar{m}$ ?

It is seen that  $\hat{t}$  consistently underestimates  $t$  // when  $t - \tau$  is positively large and overestimates when  $t - \tau$  is negatively large. Actually large and small values of  $t$  are poorly estimated while intermediate values of // close to (the general average) are well estimated. Perhaps, in a case like determining BP for diagnostic purposes, underestimating large values and overestimating/ small values are not desirable. However,  $\hat{t}$  may be better in other situations where the quadratic loss function is meaningful.

in other situations where the quadratic loss function is meaningful.

## REFERENCES

- 1.- BAKSALARY, J.K. and KALA, R (1981). Linear transformations preserving the best linear unbiased estimator in a general Gauss-Markov model. *Ann. Stat.* 9, / 913-916.
- 2.- BARNARD, G.A. (1963). The logic of least squares. *J. Roy. Stat. Soc. B* 25, 124-127.
- 3.- DRYGAS, H. (1983a). On the unified theory of least / squares. Reprint No. 4/83, *Mathematische Schriften*,/ Kassel.
- 4.- DRYGAS, H. (1983b). Sufficiency and completeness in/ the general Gauss-Markov model. *Sankhya* 45A, 88-98.
- 5.- EFRON, B (1973). Combining possibly related estimation problems (with discussion). *J. Roy Statist. // Soc.*, 8, 35, 379-421.
- 6.- EFRON B. and MORRIS, C. (1972). Empirical Bayes estimator on vector observations - an extension of // Stein's method. *Biometrika*, 59, 335-347.
- 7.- FAIRFIELD SMITH, H. (1936). A discriminant function/ for plant selection. *Ann. Eugenics*, (London) 7, 240-260.
- 8.- GHOSH, M. and SINHA, BIMAL KUMAR (1980). On the ro-/ bustness of least squares procedures in regression / models. *J. Mult. Analysis*, 10, 332-342.
- 9.- GOLDMAN, A. J. and ZELEN, M. (1964). Weak generali-/ zed inverse and minimum variance unbiased estimation. *J. Research Nat. Bureau of Standards* 68 B, 151-172.
- 10.- HENDERSON,C.R. (1950). Estimation of genetic parame-ters (Abstract). *Ann. Math. Statist.* 21, 309-310.
- 11.- HOERL, A.E. and KENNARD, R.M. (1970). Ridge regre-/ ssion: Biased estimation for monorthogonal problems. *Technometrics*, 12, 55-67.
- 12.- KARIYA, T. (1981). A robustness property of Hote- // lling's  $T^2$ -Test. *Ann. Statist.* 9, 211-214.
- 13.- KEMPTHORNE, O. (1976). Best linear unbiased estima- tion with arbitrary covariance matrix. Chapter 14 / (pp. 203-225) in *Essays in Probability and Statis- / tics*, Sinko Tsusho Co. Ltd., Tokyo.
- 14.- KHATRI, C.G. (1980). Study of F-tests under depen-/ model. *Sankhyaya*, Series A, 43, 107-110.
- 15.- KŁONECHI, ". (1978). Admissible linear estimation / in linear models. Tech. Report, Polish Academy of / Sciences.
- 16.- LA MOTTE, L.R. (1980). Admissibility in linear esti- mation. *Ann. Statisti.* 10, 245-255.

## SOME THOUGHTS ON REGRESSION AND PREDICTION

## 1. INTRODUCTION

Regression is the most widely discussed technique in statistical literature. It is mentioned in all text books on statistics and treated in a rigorous manner in some (for instance, Fisher, 1925; Malinvaud, 1964; Rao, 1965; Searle, 1971; and Theil, 1971). There are full length treatises on the subject (such as Draper and Smith, 1966; Plackett, 1960; and Williams, 1959). Fairly good reviews of regression methods are available in a discussion paper by Cox, Ehrenberg and Nelder (1968) and in papers by Draper and Smith (1969), Williams (1969) and Wold and Lyttkens (1969).

However, there seem to be some aspects of the regression technique which are not properly understood or which do not seem to have been adequately discussed in statistical literature. As a consequence, there have been some misuses of the technique. A classical example is the use of the regression method in what is known as the *disaggregation* problem (section 2), which has led to meaningless results. The phenomenon of *multicollinearity* although widely discussed still seems to need some clarification. The existing definitions are vague and the consequences of multicollinearity have not been fully examined (section 3). There is also the problem of inverse regression (section 4) on which there is considerable controversy (see comments by Lindley on papers by Cox, Ehrenberg and Nelder, 1968). The use of irrelevant variables in prediction continues in spite of the caution conveyed by the excellent examples of spurious regression by Yule (1926) long ago and by Box (1966) recently.

There is also some controversy about the use of two regression lines and the Ehrenberg paradox which is as follows. Let  $h = \alpha_1 + \beta_1 w$  be the regression of height on weight and  $w = \alpha_2 + \beta_2 h$ , that of weight on height. Consider an individual with the combination  $(h_0, w_0)$  of height and weight, which falls within the region  $\{h > \alpha_1 + \beta_1 w, w > \alpha_2 + \beta_2 h\}$  in the  $(h, w)$  plane. A doctor, using the regression of weight on height might declare the observed individual as *too heavy for his height*. But judged by the regression of height on weight Faced with such a paradox, Ehrenberg (1963, 1968) made a case for establishing a single law-like relationship between height and weight is independent of sex, race, etc. Whether there exists such a law-like relationship or not, there will be many (Ehrenberg) paradoxical individuals in any population, but the question whether one/

characteristic is in proper proportion to the other may arise only in a pathological case. Which regression to use in such a case, perhaps, depends on the judgement of the doctor based on an examination of a particular individual?

The object of the present paper is to examine some (but not all) aspects of the regression problem, which have not been adequately covered in earlier reviews.

*Notation.* In the regression problem it is usual to call the variable ( $y$ ) under prediction as dependent and those used for prediction ( $x_1, x_2, \dots, x_m$ ) as independent variables. Psychologists use the term criterion variable (c.v.) for  $y$  and predictor variables (p.v.'s) for  $x_1, x_2, \dots, x_m$ . Economists call  $y$  as the dependent variable (d.v.) and  $x_1, x_2, \dots, x_m$  as explanatory variables (e.v.'s-sounds better when pronounced as eves). We shall adopt the terminology of the economists for discussion of the present paper.

We denote  $n$  uncorrelated observations on  $y$  by the vector  $Y$  and the  $n \times m$  matrix of e.v.'s centred at the observed averages by  $X$ . Then the linear regression model is of the form

$$Y = \alpha u + X\beta + e \quad (1.1)$$

where  $u' = (1, \dots, 1)$ ,  $\alpha$  is a constant and  $\beta$  is the vector of regression parameters. The matrix  $(n-1)^{-1}X'X$  is called the dispersion or D-matrix of the e.v.'s. Dividing the  $(i, j)$ -th element of the D-matrix by the square root of the product of the  $i$ -th and  $j$ -th diagonal elements, we obtain the correlation or C-matrix.

The least squares estimator of  $\hat{\alpha}$  is  $\hat{\alpha} = n^{-1}Y'u$  and the normal equation for estimating  $\beta$  is

$$(X'X)\beta = X'Y \quad (1.2)$$

giving

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (1.3)$$

as the estimator of  $\beta$ , where  $(X'X)^{-1}$  is a g-inverse of  $X'X$  (Rao, 1965; Rao and Mitra 1971). An unbiased estimator of  $\sigma^2$ , the residual variance about the regression, is

$$(n-r) [Y'Y - \hat{a}Y'u - \hat{\beta}'X'Y] \quad (1.4)$$

where  $r$  is the rank of  $X$ . When  $r < m$ , not all linear functions of  $\beta$  are estimable. The estimator of  $p'\beta$ , where  $p$  belongs to the linear manifold generated by the columns of  $X'$  (which is the condition for estimability) is  $p'\hat{\beta}$  and the variance of the estimator is

$$\sigma^2 p'(X'X) - p. \quad (1.5)$$

The predicted value of  $y$  at the values of e.v.'s expressed as deviations from observed averages in past data,  $p' = (\bar{X}_1 - X_1, \dots, \bar{X}_m - X_m)$ , is

$$\hat{a} + p'\hat{\beta} \quad (1.6)$$

with the variance of prediction

$$\sigma^2 (1 + \frac{1}{n} + p'(X'X) - p). \quad (1.7)$$

In our discussion we shall be concerned only with the term  $p'(X'X) - p$  in (1.7), which depends on  $X'X$ . See Rao (1965, 1973) for further detail on least squares theory.

## 2. DISAGGREGATION

Sometimes the regression technique is used to allocate a total to different heads. For instance, consider a survey where from each selected household we obtain total expenditure ( $t$ ) on rice and the family composition in terms of male ( $m$ ) and female ( $f$ ) adults, and male ( $c_1$ ) and female ( $c_2$ ) children. Knowing values of  $(t, m, f, c_1, c_2)$  for  $n$  sample households, can we estimate per capita expenditure on rice for each of the four categories of members in a household in a population under study? If  $\beta_1, \beta_2, \beta_3$  and  $\beta_4$  represent the parameters of average expenditure due to a male adult, female adult, male child and female child respectively, then it is suggested that they may be estimated by the partial regression coefficients of  $t$  on  $m, f, c_1$  and  $c_2$  respectively, i.e., by fitting a regression equation of the form

$$t = \beta_1 m + \beta_2 f + \beta_3 c_1 + \beta_4 c_2 \quad (2.1)$$

to the  $n$  sets of observations on  $(t, m, f, c_1, c_2)$ . An

attempt to fit a regression equation of the type (2.1) at the Indian Statistical Institute resulted in a negative estimate for  $\beta_2$  (the average expenditure for a female adult). Elsewhere, in a similar study of disaggregation of total expenditure of a farm on bulls, cows and calves, a negative estimate was obtained for maintaining a calf. Unfortunately the results of the latter study were published, justifying the negative estimate as resulting from a large standard error of the estimated regression coefficients.

Similar studies are reported from time to time without a critical examination of the computed coefficients. It appears that such anomalous estimates as obtained in the household and farm surveys can arise due to violation of an important assumption in the regression model, viz., the same regression parameters are applicable for all combinations of e.v.'s. For instance, consider households with family compositions  $(m=1, f=0, c_1=0, c_2=0)$  and  $(m=1, f=1, c_1=2, c_2=1)$ . The usual linear model uses the regression functions (expectations)

$$\beta_1$$

$$\beta_1 + \beta_2 + 2\beta_3 + \beta_4$$

for the two cases with the same parameters  $\beta_1, \beta_2, \beta_3$  and  $\beta_4$ . This may not be true; the average expenditure on rice for a male adult (the regression parameter  $\beta_1$ ) may not be the same for single and multiple-member households. The following example due to S.J. Poti shows how completely meaningless estimates can be obtained in extreme cases (which may be described as Poti effect) where the fundamental assumption of the regression model mentioned above is not satisfied.

Consider households with only one member (unmarried) or with a couple (husband and wife) and let  $i$  represent the number of miles per month run by the family automobile. From a survey of households providing information on the variable  $t$  and the family composition, it is desired to estimate the relative utilization (in terms of average number of miles driven) of automobiles by men and women in the population, assuming that an automobile is used by one person at a time. If we represent the corresponding parameters by  $\beta_1$  and  $\beta_2$  then the regression equation is of type

$$t = \beta_1 m + \beta_2 f$$

where the vector  $(m, f)$  can take only two possible values/

(1,0) and (1,1). If in the sample, there are  $n_1$  households of the type (1,0) with an average mileage  $t_1$  and  $n_2$  households of the type (1,1) with an average mileage  $t_2$  then the least squares estimates of  $\beta_1$  and  $\beta_2$  are easily seen to be

$$\beta_1 = t_1 \quad \text{and} \quad \beta_2 = t_2 - t_1$$

Suppose the society from which samples are drawn is such that a couple cannot afford to use the automobile to the same extent as an unmarried male. In such a case  $t_1 > t_2$  giving a negative estimate for  $\beta_2$ , the average number of miles driven by women. On the other hand if the society consists of two classes of males, one so poor that they cannot afford to marry or use the automobile freely and another, rich enough (Maharaja) to have a wife and also use the automobile freely, then  $t_1 < t_2$  in which case  $\beta_1$  would be very small compared to  $\beta_2$  showing that women (wives) use the car most of the time.

In both the extreme cases the estimates are completely meaningless as the regression parameters are dependent on the family composition. It appears that in many situations, Poti effect exists to some extent or other, and disaggregation through the regression technique may not yield reasonable results.

For instance, in the farm survey data, it is quite possible that a farmer owning only two bullocks for farming purposes might feed them well, whereas a farmer who maintains in addition to two bullocks some cows and calves may not be spending a much larger amount for feeding all the animals. In such a case one would expect underestimates or even negative estimates for maintenance of a cow and a calf by fitting a regression equation to data obtained from a sample of all farmers/ some of whom own only two bullocks and others have some cows and calves in addition.

To sum up, the use of regression methods for disaggregation generally leads to meaningless or misleading results because of the existence of Poti effect. There does not seem to be any satisfactory method for the problem except through separate accounting for different categories within a unit.

### 3. MULTICOLLINEARITY

There is considerable literature on the subject of /

multicollinearity, a review of which may be found in a paper by Farrar and Glauber (1967). Multicollinearity is roughly described as high interdependence of the c.v.'s in a regression model, which may effect the precision of estimators and cause some difficulty in the interpretation of individual regression coefficients, and in the use of the estimated regression function for prediction purposes. Interdependence is sought to be measured and judged by the smallness of the determinant of the D or C-matrix of the e.v.'s. Farrar and Glauber (1967) discuss at length methods for detecting the existence, measuring the extent, and pinpointing the location and causes of multicollinearity within a set of independent variables. Their analysis is indeed illuminating, but much of their discussion is misguided partly due to an unsatisfactory definition of multicollinearity, which is not directly related to the precision of prediction in the relevant domain of e.v.'s and partly due to, as pointed out by Krishna Kumar (1973), invalid excursions into tests of significance based on normality assumptions. In the present paper we shall examine some aspects of multicollinearity, which have not been adequately examined before.

Consider  $e_0$  a d.v. and  $e_1, e_2$  two c.v.'s with the D and C-matrices

D-matrix			C-matrix		
1	5	2	1.0	.5	.4
5	100	.	.5	1.0	.
2	.	.25	.4	.	1.0 (3.1)

Since  $e_1$  and  $e_2$  are uncorrelated, one might consider this situation as ideal for an application of the regression technique for prediction. However, instead of  $e_1, e_2$ , one could consider an algebraically equivalent and perhaps equally meaningful e.v.'s,  $x_1 = e_1 + e_2$  and  $x_2 = e_1$ . The D- and C-matrices of  $e_0, x_1, x_2$  are

D-matrix			C-matrix		
1	7	5	1.000	.626	.500
7	125	100	.626	1.000	.894
5	100	100	.500	.894	1.000

(3.2)

Judging from the high value of the correlation, .894, between  $x_1$  and  $x_2$  one might infer some degree of multicollinearity among the e.v.'s. Since  $(x_1, x_2)$  is equivalent to  $(e_1, e_2)$  the situation is paradoxical unless multicollinearity is defined as high interdependence for/

a particular choice of the e.v.'s and not as a property of the system of e.v.'s. (Then we have the additional problem of deciding on the relevant set of e.v.'s for examining multicollinearity).

The situation is not uncommon in econometric work / where some e.v.'s are part of others. It might appear/ that multicollinearity in the sense of smallness of the determinant could be introduced or avoided by redefi- / ning economic variables in an equally meaningful way. / (Thus, if  $e_1$  and  $e_2$  stand for production and net im- // ports, then  $e_1 + e_2$  denote production and/ consumption which are equally meaningful and equivalent economic magnitudes). It is immaterial whether  $e_1$ ,  $e_2$  or  $e_1 + e_2$  are chosen as e.v.'s. But one choice may // exhibit multicollinearity and the other may not. As a/ matter of fact, any vector variable  $X$  with any degree / of dependence (correlations) between its components can be transformed into an equivalent vector variable  $U$  // with uncorrelated components. Since from the point of/ view of the regression technique  $X$  and  $U$  carry the same information, the distinction between high dependence of the components of one and absence of association in the other is not apparent. This shows the need for a more/ careful formulation of multicollinearity in a given // problem, perhaps considering the contemplated use of a/ fitted regression equation. Before attempting to do // this we shall use the D-matrices in (3.1) and (3.2) in/ illustrating some aspects of the regression coefficients which are worth noting.

It is often found that when a number of e.v.'s are/ used, some of the regression coefficients come out to / be negative, although the corresponding variables are / individually highly positively correlated with the d.v. The situation is fairly common when the correlations // between the e.v.'s are high as pointed out by Shourie / (1972) in his critical essay on the regression func- // tions given in the UNCTAD study of (1968). For instance the following regression equations are reported by Faa- land and Dahl to account for the value of agricultural/ output ( $P_A$ ) in terms of GDP and per capita income (GDP/ N):

$$P_A = 34.322 + 0.402(\text{GDP}) \quad R^2 = 0.94 \quad (3.3)$$

(0.036)

$$= -8.001 + 5.698(\text{GDP}/N), \quad R^2 = 0.96 \quad (3.4)$$

(0.397)

$$= -131.716 - 1.226(\text{GDP}) + \quad R^2 = 0.99 \quad (3.5)$$

(0.239)

$$+ 22.816(\text{GDP}/N), \quad (3.345)$$

The coefficient of correlation between (GDP) and (GDP/N) was 0.9985. The regression coefficients in each case // are large compared to standard errors (given in brackets) indicating a reasonable degree of precision in their es- timation. The multiple correlation ( $R$ ) is high in each/ case indicating a high degree of precision for predic- / tion. However, the authors point out that on a priori / grounds the negative partial correlation of  $P_A$  with GDP/ in the equation (3.5) appears improbable.

There may be strong economic reasons for the authors' reactions to the negative coefficient. But a multiple / regression equation such as (3.5) cannot be rejected // solely on the ground that a certain coefficient is nega- tive although the concerned variable has a high positive correlation with the d.v. Of course, one may express // some concern in such a situation. The implausible result of wrong specification of the linear model or of a multi- tude of other reasons including errors in recording. (In many examples, wrong recording of just one entry may // change the values of regression coefficients drastically) However, a negative coefficient may arise in a natural // way because of particular choices of c.v.'s in given sys- tem.

Consider, for instance, the hypothetical example of /  $e_0$ ,  $e_1$ ,  $e_2$  with the D-matrix as in (3.1). The va-/// riabiles  $e_1$  and  $e_2$  may be some economic magnitudes, in /// which case the variables  $x_1 = e_1$  and  $x_2 = e_1 + e_2$  may // also be meaningful economic magnitudes. The regression/ of  $e_0$  on different sets of e.v.'s are as follows (apart/ from a constant):

$$e_0 = .05e_1 + 0.8e_2 \quad (3.6)$$

$$= .05e_1$$

$$= .08e_2$$

$$= -.03x_1 + .08x_2 \quad (3.7)$$

$$= .05x_1$$

$$= .028x_2$$

The equations (3.6) and (3.7) are the same and the nega- / tive coefficient for  $x_1$  in the multiple regression equa- / tion (3.7) arises in a natural way although the correla- / tion between  $e_0$  and  $x_1$  is 0.5. At first sight, this // might appear contrary to expectation, but a closer look / at the choice of the e.v.'s might resolve the situation./ In the above example the regression functions based on //  $x_1$ ,  $x_2$  has the same validity as that based on  $e_1$ ,  $e_2$ ,/ and the use of  $x_1$ ,  $x_2$  or  $e_1$ ,  $e_2$  as instruments yield // the same results.

It must be pointed out that the occurrence of negative coefficients (contrary to expectation) need not be due to multicollinearity or high correlations between e.v.'s. Consider the following C-matrix of three variables  $y, x_1, x_2$ :

$$\begin{array}{ccc} 1.0 & 0.7 & 0.3 \\ 0.7 & 1.0 & 0.5 \\ 0.3 & 0.5 & 1.0 \end{array} \quad (3.8)$$

where the e.v.'s  $x_1$  and  $x_2$  exhibit no sign of multicollinearity. The regression equations of  $y$  on different combinations of the e.v.'s are (apart from a constant)

$$\begin{aligned} y &= 0.7x_1 \\ &= 0.3x_2 \\ &= 0.733x_1 - 0.067x_2 \end{aligned} \quad (3.9)$$

The negative coefficient for  $x_2$  does not detract its value as an explanatory variable. No doubt, one has to be cautious in using  $x_1, x_2$  as instruments, not only in such a situation but in general even when the coefficients appear plausible. The nature of the economic variables and the domain of validity of the prediction function have to be examined, perhaps, using one's knowledge of the underlying economic mechanism which controls  $y$  given specified values of  $x_1, x_2$ . If prediction is needed for values of  $x_1, x_2$  outside a certain domain, there may be other factors, unobserved in the past, which may play an important role.

*What is multicollinearity?* In the strict sense it implies existence of linear relationships among the e.v.'s in the past observed data. Or in algebraic terms,  $|X'X| = 0$  where  $X$  is the matrix of observed e.v.'s expressed as deviations from the means. When this happens the rank of  $X$  is less than the number of e.v.'s and the entire regression function cannot be estimated as not all the individual regression coefficients (i.e., separate influences of the e.v.'s) are identifiable. However, certain combinations of the regression coefficients are estimable implying that prediction is possible only in certain directions of the e.v.'s. The method of estimation when  $X$  is deficient in rank is fully discussed in Rao (1965, 1973).

However, in practice, due to errors of observation in the e.v.'s, identification of linear relationships, if any, may not be possible. Let us write the matrix of

observed e.v.'s as  $X+W$ , where  $X$  denotes the matrix of true unknown values and  $W$  of the errors. Then it might happen that  $| (X + W)' (X + W) | \neq 0$  although  $|X'X| = 0$  giving no clue as to whether it is a case of e.v.'s without errors and with no linear relationships (i.e.,  $|X'X| \neq 0, W = 0$ ), or of e.v.'s with linear relationships but observed with error (i.e.,  $|X'X| = 0$  and  $| (X+W)' (X+W) | \neq 0$ ). We shall examine both the cases and study their consequences on the estimation of parameters.

Case 1:  $|X'X| \neq 0$ .

There is no multicollinearity in the strict sense. According to theory, all the regression coefficients are unbiasedly estimable and prediction is possible in all directions of the e.v.'s. Let us consider a particular direction  $k$  of the e.v.'s, where  $k$  is vector such that  $k'k = 1$ . If  $\hat{\beta}$  denotes the least squares estimator of  $\beta$  then the predicted value of the d.v. in direction  $k$  is (apart from a constant) proportional to  $k'\hat{\beta}$  with variance

$$\sigma^2 k'(X'Y)^{-1}k \quad (3.10)$$

where  $\sigma^2$  is the residual variance. The important component of the expression (3.10) is  $(X'X)^{-1}$ , and the precision of prediction in any given direction depends on the elements of the inverse matrix. However, for any given  $X'X$ , the precision of prediction depends on the direction  $k$ . Let us denote by  $K$  the set of directions in which prediction is contemplated. Then the minimum efficiency of prediction (or maximum variance for vectors in  $K$ ) is

$$r(X, K) = \max_{k \in K} k'(X'X)^{-1}k \quad (3.11)$$

which we may consider as the loss associated with a given system  $X$  of e.v.'s and specified  $K$ .

If  $K$  is the set of all directions in the Euclidean space  $E^m$  of  $m$  of  $m$  dimensions,

$$r(X, E^m) = 1/\lambda_m \quad (3.12)$$

where  $\lambda_m$  is the least eigen value of  $X'X$ , and the direction in which prediction is least efficient is the eigen vector associated with  $\lambda_m$ . Generally, when  $X'X$  is close to zero,  $\lambda_m$  is small and consequently the loss as in (3.12) is high- a situation which may be referred to as

the existence of some degree of multicollinearity. But/in practice, one may not be interested in prediction in/all directions and, therefore, the magnitude (3.12) may/not be relevant. Let us represent all the eigen values/of  $X'X$  by

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$$

and the corresponding eigen vectors by

$$k_1, \dots, k_m$$

If our interest is confined to directions lying in the / linear manifold generated by  $k_1, \dots, k_q$  then the value/ of (3.11) is  $\lambda_q^{-1}$  which may be small although  $\lambda_m^{-1}$  is / large. Thus, the nearness of  $|X'X|$  (which is equal to / the product  $\lambda_1 \lambda_2 \dots \lambda_m$ ) to zero due to  $\lambda_m$  being // small is of no consequences for our purpose. Indeed one can give examples  $X_1$  and  $X_2$  of  $X$  where

$$\lambda_q(X_1'X_1) > \lambda_q(X_2'X_2) \quad X_2$$

while  $\lambda_m(X_1'X_1) < \lambda_m(X_2'X_2)$  and/or  $|X_1'X_1| < |X_2'X_2|$ , / in which case  $X_1$  is a better choice of the e.v.'s, al- / though on the criterion of multicollinearity based on / the least eigen value,  $X_2$  may be considered better.

Thus, one may define (3.11) as the degree of multi-/ collinearity in  $X$  taking into account the set of direc-/ tions of interest. However, it must be understood that/ if data exhibit some degree of multicollinearity it is / reflected in the standard errors being large (there is / no method of avoiding it except by the use of suplemen-/ tary information). In such a case the width of the con-/ fidence interval will be large and consequently the use-/ fulness of the predicted value is reduced. Thus we have a built-in safeguard against any rash interpretation of/ predicted values.

Let us consider two e.v.'s  $e_1$  and  $e_2$  with the // C-matrix

$$\begin{array}{cc} 1.00 & 0.99 \\ 0.99 & 1.00 \end{array} \quad (3.13)$$

The determinant of the correlation matrix is 0.0199 which may be considered small indicating high multicollinearity of the e.v.'s. The eigen values of (3.13) are 1.99 and / 0.01 so that the variance of the predicted values is proportional to 0.503 in the direction (1:1) for standardised

e.v.'s and 100 in the direction (1:-1). Thus if we are/ interested in the direction (1:1), the high variance for the other direction is of no relevance. Indeed if the / e.v.'s in the past exhibited no correlation, indicating/ perfect non-collinearity, the variance of prediction in/ the direction (1:1) would be proportional to 1.00 which/ is higher than that for the multicollinearity case (3.13)

Case 2:  $|X'X| = 0$  and  $|U'U| \neq 0$ .  $U = X + W$ .

There is indeed multicollinearity in the strict sense but is obscured by the unknown error matrix  $W$ , and case/ 2 is indistinguishable from case 1 unless some additio-/ nal information is provided. Using the method of least/ squares the regression parameters would be estimated by

$$\hat{\beta} = (U'U)^{-1}U'Y \quad (3.14)$$

with the associated D-matrix

$$\sigma^2(U'U)^{-1} \quad (3.15)$$

The predicted value in any direction  $k$  is  $k'\hat{\beta}$  with the / attributed variance  $\sigma^2 k'(U'U)^{-1}k$ .

Let  $U'U - X'X = G$  be positive definite\* and the rank / of  $X$  be  $r$ . Then there exists a matrix  $P$  such that

$$X'X + PAP', \quad G = PP' \quad (3.16)$$

where  $\Lambda$  is a diagonal matrix with diagonal elements //  $\lambda_1 \geq \dots \geq \lambda_r \neq 0, 0, \dots, 0$ . If  $P_i$  is the  $i$ -th co-/ lumn of  $P$ , then it is easy to see that  $P_1 \dots P_r$  span  $X'X$ / (or  $X'$ ) and no combination of  $P_{r+1}, \dots, P_m$  belongs to / the linear manifold of  $X'X$ .

In terms of the  $P$  and  $\Lambda$  matrices

$$(U'U)^{-1} = (P^{-1})'(\Lambda + I)^{-1}P^{-1}. \quad (3.17)$$

The predicted value in the direction  $P_i$  is  $P_i'\hat{\beta}$ , where /  $\hat{\beta}$  is as given in (3.14). Then,

$$E(P_i'\hat{\beta}) = P_i'(P^{-1})'(\Lambda + I)^{-1}P^{-1}(X'X + W'X)$$

---

\* If  $X$  and  $W$  are uncorrelated, then  $E(U'U - X'X)$  is non negative defini-te, but  $U'U - X'X$  may not be for all values of  $W$ .

$$= \frac{\lambda_i}{1+\lambda_i} P_i' \beta + \frac{1}{1+\lambda_i} Q_i' W' X \beta, \\ i=1, \dots, r \quad (3.18)$$

$$= Q_i' W' X \beta, \quad i > r. \quad (3.19)$$

where  $Q_i$  is the  $i$ -th column of  $(P^{-1})'$ . For  $i = 1, \dots, r$ .

$$E(P_i' \hat{\beta}) - P_i' \beta = \frac{1}{1+\lambda_i} P_i' \beta + Q_i' W' X \beta \quad (3.20)$$

which is of the order  $(1 + \lambda_i)^{-1}$  ignoring the error term in (3.20). If the elements of  $W'W$  are small in magnitude compared to those of  $X'X$ ,  $\lambda_i$  are likely to be large and the bias in (3.20) may be small, provided the direction of e.v.'s is in the linear manifold of  $X'X$ . But for  $i > r$  when  $P_i$  does not belong to the linear manifold of  $X'X$ .

$$E(P_i' \hat{\beta}) = Q_i' W' X \beta \quad (3.21)$$

which seems to have no relation to  $P_i' \beta$  the parameter under estimation, and whose magnitude depends only on the errors in  $X$ . Thus any claim for  $P_i' \hat{\beta}$  to be an estimator of  $P_i' \beta$  for  $i > r$  is devoid of any meaning.

Let us examine the attributed variance of  $P_i' \hat{\beta}$  which is apart from the multiplier  $\sigma^2$

$$P_i' (U'U)^{-1} P_i = \frac{1}{1+\lambda_i}, \quad i=1, \dots, r \quad (3.22)$$

$$= 1, \quad i > r. \quad (3.23)$$

Thus if  $\lambda_i$  are large, the functions  $P_i' \beta$  for  $i = 1, \dots, r$ , are estimated with reasonable precision. The variance associated with the proposed estimators of  $P_i' \beta$  for  $i > r$  are of a relatively large magnitude.

To sum up, if the e.v.'s had been observed without error, we might have discovered that linear functions of the type  $P_i' \beta$  for  $i > r$  are not estimable. Non-recognition of errors when they exist might lead us to claim that  $P_i' \beta$  for  $i > r$  is also estimable and its estimate and variance are, respectively,

$$P_i' \hat{\beta} \quad \text{and} \quad \sigma^2 P_i' (U'U)^{-1} P_i \quad (3.24)$$

If the rank of  $X$  is  $r$ , the estimator  $P_i' \hat{\beta}$  for  $i > r$  will be in the nature of pure noise subject to a large variance and bears no relation to the parameter  $P_i' \beta$  under estimation.

In such a situation, wrong claims will also be made about functions  $k' \beta$  where  $k$  is a combination of  $P_i'$ 's for some values of  $i \leq r$  and some greater than  $r$ . The estimator  $k' \beta$  may have large bias as well as large variance.

Thus large variance for a predictor may be due to nonestimability as in case 2 or poor estimability as in case 1 of certain functions, and, often, it is difficult to distinguish between the two cases. The situation is not serious in practice provided due caution is exercised in using estimates subject to large standard errors (by computing confidence intervals).

An investigation on different lines when  $|X'X| = 0$  and there are errors in  $X$  has been carried out by Rao and Potluri (1973). The results obtained by him are of interest from a practical point of view.

#### 4. VARIABLES WITH A FACTOR STRUCTURE

Let us consider  $y$  (d.v.) and  $x_1, \dots, x_m$  (e.v.'s) with a factor structure, i.e., there exist hypothetical variables  $f_1, \dots, f_k$  such that

$$y = a_{01} f_1 + \dots + a_{0k} f_k + e_0 \quad (4.1)$$

$$x_i = a_{i1} f_1 + \dots + a_{ik} f_k + e_i, \quad i=1, \dots, m, \quad (4.2)$$

where  $e_0, e_1, \dots, e_m$  are uncorrelated among themselves and also uncorrelated with the factors  $f_1, \dots, f_k$ . How do we use the information of the factor structure (4.1, 4.2) in predicting  $y$  using  $x_1, \dots, x_m$ ?

To understand the problem let us consider the case of a single factor variable and the structure

$$y = a_f f + e_0, \quad x_i = f + e_i, \quad i=1, \dots, m \quad (4.3)$$

where the coefficients of  $f$  in  $x_i$  can be taken as unity without loss of generality. Further, let

$$V(f) = \sigma_f^2, \quad V(e_i) = \sigma_i^2. \quad (4.4)$$

Then it is easily seen that the regression of  $y$  on  $f$  is

$$y^* = a_0 f \quad (4.5)$$

and that of  $y$  on  $x_1, \dots, x_m$  is

$$y = a_0 \sum \frac{\sigma_i^2}{1 + \sum \alpha_i^2} x_i, \quad \alpha_i^2 = \sigma_f^2 / \sigma_i^2. \quad (4.6)$$

A noticeable feature of the regression equation is that as  $m$  increases, the coefficient of any fixed  $x$  decreases and the sum of the coefficients is

$$a_0 \frac{\sum \alpha_i^2}{1 + \sum \alpha_i^2} \quad (4.7)$$

which is near to  $a_0$  as  $m$  increases. The square of the multiple correlation of  $y$  on  $x_1, \dots, x_m$  is

$$\alpha_0^2 = \left(1 + \frac{1}{\sum \alpha_i^2}\right) \left(a_0^2 + \frac{1}{a_0^2}\right) \quad (4.8)$$

which has the upper bound

$$\alpha_0^2 \leq \left(a_0^2 + \frac{1}{a_0^2}\right) \quad (4.9)$$

This implies that as  $m \rightarrow \infty$ , the rate of reduction in residual variance of  $y$  given  $x_1, \dots, x_m$  will be small.

Suppose we fit the regression of  $y$  on  $x_1, \dots, x_m$  on the basis of  $n$  sets of observations on  $(y, x_1, \dots, x_m)$  in the usual way. Then the magnitudes of the partial regression coefficients decrease as  $m$  increases while the standard errors of these coefficients might increase from a certain stage of  $m$  value. The result is that if significance of these coefficients is tested at 5 per cent or 1 per cent level, a large number may be declared insignificant.

When there is more than one common factor, a similar situation arises. The number of regression coefficients increase, but their magnitudes are subject to linear constraints. Since such constraints are not recognised in fitting the regression, the precision of individual estimators decreases as  $m$  increases. Further, there is a natural limit to the multiple correlation as  $m$  increases, and as a consequence tests on individual regression coefficients become less and less powerful unless the

sample size is simultaneously increased.

The classical example of Fisher (1938) in the prediction of high altitude acclimatization using 7 sea level / characteristics may be explained by the phenomenon described above. In the formula with 7 characteristics, no coefficient was significantly different from zero at the 5 % level of significance. Commenting on this phenomenon Fisher observed:

This does not mean that the formula is worthless, but that all the individual coefficients might be varied // largely, and provided the other coefficients were suitably adjusted, the predictive value would not appreciably be impaired. The formula in seven variables is, therefore, to a large extent arbitrary and one or more of the variables used must certainly be redundant.

Fisher proceeded to eliminate some variables one at a time omitting each time the variable whose coefficient // has the least F (or t) ratio, till the coefficients in the regression equation with reduced number of variables reached significance at the 5% level. Fisher also provided another explanation for eliminating some of the variables arguing that the variance of the error of prediction by the regression formula should not be below that of independent errors in the d.v. The use of more than four characters seemed to indicate spurious precision.

It is not clear what Fisher had in mind in stating // that the regression equation is arbitrary indicating that some variables are redundant. Perhaps he was referring to multicollinearity which is discussed in section 3 of the lecture. However, the situation as in Fisher's example can arise when the variables have a factor structure/ as demonstrated in the simple case of a single factor. // In such a case we may proceed as follows.

Let us again consider the structure as in (4.3) and / denote  $\bar{x} = (x_1 + \dots + x_m)/m$ . The regression of  $y$  on  $\bar{x}$  is

$$y = \frac{a_0}{1 + m^{-2} \sum \alpha_i^2} \bar{x}. \quad (4.10)$$

If the coefficient of  $x$  is estimated in the usual way / from  $n$  sets of observations on  $(y, x)$ , then no problem of/ the type encountered in the case of multiple regression / coefficients arise. Error of prediction decreases as //  $m \rightarrow \infty$ , and no variable need be considered redundant. We note that  $x$  is an estimate (not necessarily the best)

of the factor variable  $f$ , so that the regression of  $y$  on  $\bar{x}$  is a close approximation to that of  $y$  on  $f$ . In practice one does not know the factor structure or the number of factors. In such a case the possibility of extracting factors using the estimated dispersion matrix of  $(x_1, \dots, x_m)$  may be considered. Let  $\hat{f}_1, \hat{f}_2, \dots, \hat{f}_k$  be estimated factors as linear combinations of  $x_1, \dots, x_m$  using any standard method of factor analysis. If the factor structure as in (4.1, 4.2) holds, then the usual regression of  $y$  on  $\hat{f}_1, \dots, \hat{f}_k$  may provide a better prediction formula than the usual regression of  $y$  on  $x_1, \dots, x_m$ .

The technique described above is, probably, well known to factor analysis, but has not been popular in statistical literature. However, attempts have been made by economists to use the dominant principal components of  $x_1, \dots, x_m$  in the prediction formula, which is not the same thing as using the factors unless the variances  $\sigma_i^2 = V(\epsilon_i)$  in (4.2) are nearly equal (see for instance Rao, 1955). Even under factor structure, the least important principal components may have high correlations with the d.v., and neglecting them may impair precision of prediction. It must be understood that the suggested method is applicable only when the factor structure of the type (4.1, 4.2) holds both for dependent as well as for explanatory variables.

## 5. ESTIMATION OF TRUE MEASUREMENT- AN ANOMALY

Suppose  $x$  is a true (unobservable) measurement of a characteristic of an individual and  $y$  is the observed measurement subject to error, i.e.,  $y = x + e$  where  $e$  is an error independent of  $x$ . The problem is one of estimating (or predicting)  $x$  given  $y$ .

Let us consider the bivariate distribution of  $(x, y)$ . It may be seen that

$$E(y) = E(x) = \alpha \quad (5.1)$$

$$V(y) = \sigma_x^2 + \sigma_e^2, \text{ cov}(x, y) = \sigma_x^2 \quad (5.2)$$

where  $\sigma_x^2$  and  $\sigma_e^2$  are the variances of the marginal distributions of  $x$  and  $e$  respectively. By hypothesis, the regression of  $y$  on  $x$  is linear, which may not imply that the regression of  $x$  on  $y$  is linear. The exact conditions under which this happens is given in Rao (1974). If we assume that the regression of  $x$  on  $y$  is also linear, then the regression equation is

$$\begin{aligned} \hat{x} &= \alpha + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} (y - \alpha) \\ &= \frac{\sigma_e^2}{\sigma_x^2 + \sigma_e^2} \alpha + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} y. \end{aligned} \quad (5.3)$$

$y$  is an unbiased (or inverse regression) estimator of  $x$  and  $\hat{x}$  is the (direct) regression estimator of  $x$ , which are equal only when  $\sigma_e^2$ , the error variance is zero. The mean square errors (m.s.e.) of these two estimators are

$$E(y - x)^2 = \sigma_e^2 \quad (5.4)$$

$$E(\hat{x} - x)^2 = \frac{\sigma_x^2 \sigma_e^2}{\sigma_x^2 + \sigma_e^2} \leq \sigma_e^2 \quad (5.5)$$

so that judged by the criterion of m.s.e. (or quadratic loss), the direct regression estimator  $\hat{x}$  is better than the inverse regression estimator  $y$ .

In practice  $\alpha$ ,  $\sigma_x$  and  $\sigma_e$  may not be known. However, if past data provide multiple determinations of the specified characteristic on at least some observed individuals, then the parameters  $\alpha$ ,  $\sigma_x$  and  $\sigma_e$  may be estimated as shown in Rao (1965). Using estimates  $\hat{\alpha}$ ,  $\hat{\sigma}_x$ ,  $\hat{\sigma}_e$ , the regression estimator can be written

$$\hat{x} = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} \hat{\alpha} + \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} y \quad (5.6)$$

which will still be better than  $y$  provided the past data are sufficiently numerous and stable estimates of  $\hat{\alpha}$ ,  $\hat{\sigma}_x$  and  $\hat{\sigma}_e$  are available.

Let us consider a possible application of the estimation procedure described above. Suppose an individual goes to a hospital and the doctor wants an examination of his blood cholesterol (b.c.) from the clinical laboratory which reports it as  $y$  units. It is known that the determination of b.c. is subject to error whose variance is  $\sigma_e^2$  say. From past data on individuals referred to the b.c. test one may have estimates of  $\hat{\alpha}$  the general mean,  $\hat{\sigma}_x^2$  the variance of true b.c. between individuals and  $\hat{\sigma}_e^2$  the variance of error in an individual determination. Should the doctor, as a routine, use the estimator  $y$  or alternative regression estimator with possibly smaller m.s.e.,

$$\hat{x} = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} \hat{\alpha} + \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} y \quad (5.7)$$

for diagnostic purposes? It may be seen that  $\hat{x}$  is a / biased estimator of  $x$  and  $|\hat{x} - \bar{\alpha}| < |y - \bar{\alpha}|$ , i.e., /  $\hat{x}$  deviates somewhat less from the general mean than  $y$  / thus indicating a less pathological situation than that revealed by  $y$ .

In order to answer this question, let us compute / the mean square errors in  $y$  and  $\hat{x}$  for given  $x$ , assuming that  $\alpha, \sigma_x, \sigma_e$  are known, which are

$$E[(y-x)^2/x] = \sigma_e^2 \quad (5.8)$$

$$E[(\hat{x}-x)^2/x] = \frac{\sigma_e^2 \sigma_x^4}{(\sigma_x^2 + \sigma_e^2)^2} + \frac{\sigma_e^4}{(\sigma_x^2 + \sigma_e^2)^2} (x-\bar{x})^2 \quad (5.9)$$

For given  $x$  (given individual) the m.s.e. of  $\hat{x}$  is lar-/ ger than that of  $y$  when

$$(x-\bar{x})^2 > 2\sigma_x^2 + \sigma_e^2 \quad (5.10)$$

although the over all (individuals) m.s.e. of  $\hat{x}$  is /

smaller than that of  $y$ . Thus, for large deviations of / the true value  $x$  from the general mean,  $y$  is a better / estimator than  $\hat{x}$  and for small deviations the reverse // holds. Since  $\hat{x}$  underestimates (biased downwards) and // underestimation in the case of large deviations may not/ be desirable from the point of view of medical diagnosis and treatment,  $y$  seems to be a safer choice than  $\hat{x}$  as an estimator of the unknown  $x$  in an individual case.

Further, suppose an individual is put under treat- / ment because of a large deviation of his b.c. from the / general average and the b.c. level is tested every week. Let  $y_1, y_2, \dots$  be the observed values and  $\hat{x}_1, \hat{x}_2, \dots$  / the corresponding regression estimates. Since the re- / gression estimates are biased towards the general ave- / rage, the rate of recovery as indicated by the  $\hat{x}$  series/ will be (biased) lower than that for the  $y$  series which/ provides an unbiased estimate of trend. Thus, in the // situation under consideration,  $y$  seems to serve a better purpose than  $\hat{x}$ . However, there may be other situations/ where individual considerations are not important and // loss function is quadratic, which favour the choice of  $\hat{x}$  and  $y$ .

## BIBLIOGRAFIA

- 1.- BOX, G.E.P. (1966). "Use and Abuse of regression". / *Technometrics*, 8, 625-630).
- 2.- COX, D.R., EHRENBURG, A.S.C. and NELDER, J.A. // / (1968). "Regression methods (with discussion)". *J. Roy Stat. Soc.*, 131, 265-326.
- 3.- DRAPER, N.R. and SMITH, H. (1966). "Applied Regression Analysis". *Wiley*, New York.
- 4.- DRAPER, N.R. and SMITH, H. (1969). "Methods for selecting variables from a given set of variables for regression analysis". *Proc. 37th session, ISI*, // / XLIII, Book 1, 7-16.
- 5.- EHRENBURG, A.S.C. (1963). "Bivariate regression analysis useless". *Applied Statistics*, 12, 161-179.
- 6.- FARRAR, D.E. and GALUBER, R.R. (1967). "Multicollinearity in regression analysis: The problem revisited". *The Review of Economic Statistics*, XLIX, 91- / -107.
- 7.- FISHER, R.A. (1938). "Individual variations in ability to acclimatize to high altitude" *Proc. Roy. // Soc. London, Series B*, 126, 25-29.
- 8.- KRISHNA KUMAR, T. (1973). "Note on multicollinearity in regression analysis". *Department of Economics*, Florida State University, Tallahassee.
- 9.- MALINVAUD, E. (1970). "Statistical Methods of Econometrics". *North Holland Publishing Co*, Amsterdam- / -London.
- 10.- PLACKETT, R.L. (1960). "Regression Analysis". *Clarendon Press*, Oxford.
- 11.- RAO, C.R. (1955) "Estimation and Tests of significance in factor analysis". *Psychometrika*, 20, 93- / -111.
- 12.- RAO, C.R. (1965). "Linear Statistical Inference and its Applications". (Second Edition, 1973), *Wiley*, // / New York.
- 13.- RAO, C.R. (1974). "Characterization of prior distributions and solution to a compound decision problem" (in press).
- 14.- RAO, C.R. and MITRA, S.K. (1971). "Generalized Inverse of Matrices and its Applications". *Wiley*, New // / York.
- 15.- RAO, POTLURI (1973). "Some notes on errors-in-variables model". *Am. Statistician*, 27, 217-218.
- 16.- SEARLE, S.R. (1971). "Linear Models". *Wiley*, New // / York.
- 17.- SHOURIE ARUN (1972). "The use of macro-economic regression models of developing countries for forecasts and policy prescription: Some reflections on / current practice". *Oxford Economic Papers*, 24, 1-35.
- 18.- THEIL, H. (1971). "Principles of Econometrics". // / North Holland, Amsterdam.
- 19.- UNCTAD (1968). "Trade Prospects and Capital Needs of Developing Countries". *United Nations*, New York.
- 20.- WILLIAMS, E.J. (1959). "Regression Analysis". *Wiley*, New York.
- 21.- WILLIAMS; E.J. (1969). "Regression methods in calibration problems". *Proc. 37 th session, ISI XLIII*, Book 1, 29-51.
- 22.- YULE, G. UDNY (1926). "Why do we sometimes get non-sense correlations between time series? *J. Roy. Stat. Soc.*, LXXXIX, 1-64.

## ESTADISTICOS, ESTADISTICAS Y POLITICA PUBLICA

### I. ESTADISTICAS

*No puede haber un buen plan para el progreso de la economía sin unos datos adecuados y no puede haber unos datos adecuados sin un buen plan para obtenerlos. P. C. Mahala nobis*

Cualquier discusión de las estadísticas oficiales comienza con un recuento de sus defectos en términos de / lagunas, economía, oportunidad, exactitud, presentación, fases prioritarias, personal, coordinación interdepartamental, etc., y termina con sugerencias y esperanzas de progreso. Se nombran comisiones y se constituyen comités periódicamente para profundizar en los problemas de las estadísticas oficiales y se hacen recomendaciones para / su mejora.

Hay naturalmente algunas limitaciones inherentes en el desarrollo de sistemas estadísticos eficientes. Algunos de los requerimientos como la oportunidad y el coste entran en conflicto con la demanda de exactitud e información más detallada. Los recursos, la pericia y el personal necesario en cualquier momento para mantener un / sistema estadístico coherente y para planificar la cre-

ciente demanda de información son a menudo desatendidos por los Gobiernos. Naturalmente de cualquier forma si---guen existiendo defectos a pesar de los continuos esfuerzos que se hacen para remediarlos. Sin embargo, actualmen---te, las estadísticas oficiales de cualquier país están en mejores condiciones de lo que solían estarlo anteriormen---te.

Esto se debe principalmente a la creciente demanda / de Estadísticas en orden a formular planes nacionales pa---ra el desarrollo socio-económico que cualquier país desarrollado trata de realizar para mejorar la calidad de vida de sus gentes.

La planificación implica una valoración del presente comparado con el conocimiento y la experiencia del pasado y el establecimiento de metas para el futuro. Deben / examinarse las tendencias económicas anteriores y deben determinarse los insermos actuales necesarios para lograr objetivos futuros específicos.

Todos estos procesos requieren el uso de la estadística

tica en sus dos aspectos, datos y metodología. De hecho el desarrollo de las estadísticas es un requisito previo para la formulación de los primeros planes socio-económicos. La demanda de estadísticas puede incrementarse en el futuro cuando la planificación hecha a nivel nacional se extienda ahora a niveles regionales y subregionales.

Todo esto requiere la formulación de un extenso plan en base a un sistema nacional de estadística coordinado y la creación de un departamento especial en la Oficina Nacional de Estadística con un grupo de expertos encargados de las siguientes responsabilidades:

- Revisión y mejoras periódicas en los métodos / existentes de recogida y recopilación de datos / en áreas tradicionales como agricultura e industria. Como dice Claus Moser<sup>(1)</sup>, Hay una tendencia inherente en las series económicas a deteriorarse a menos que se haga un buen trabajo de mantenimiento.
- Evaluación de las necesidades de datos para resolver los nuevos problemas que continuamente / surgen cuando se toman decisiones políticas.
- Planificar la recogida de datos de censos especiales y encuestas muestrales en áreas donde no es posible conseguir información a través de los canales normales administrativos.
- Establecer las prioridades en la recopilación y publicación de datos.
- Organizar<sup>(2)</sup> periódicamente seminarios y reuniones con la participación de estadísticos del Gobierno y especialistas de fuera para discutir / mejoras de datos y para evaluar las necesidades futuras de datos.
- Evitar duplicidades y coordinar las actividades estadísticas de los diversos organismos gubernamentales.

El último aspecto, la coordinación, es extremadamente importante si un país tiene un sistema estadístico / descentralizado. Pero la coordinación es también una labor difícil y puede haber obstrucciones administrativas y técnicas de difícil resolución<sup>(3)</sup>.

## 2. ESTADÍSTICOS

... la ciencia estadística ha sido el aspecto peculiar / del progreso humano que ha dado al siglo XX su carácter especial ... es al estadístico a quien la presente era / transforma para la más esencial de todas sus actividades

más importantes. R. A. Fisher.

Cuales son las funciones de una oficina de estadística y las responsabilidades de los estadísticos del Gobierno y del director de una oficina de estadística?. Consiste su trabajo principalmente en recopilar estadísticas para posibles usos de otros (por ejemplo, análisis e interpretación estadística)? Deben también dedicarse al análisis de datos y ayudas a los políticos en la toma de decisiones? Hay opiniones diversas en cuanto a estos temas. Voy a referirme a un artículo de Subramaniam<sup>(4)</sup> quien cita los archivos del periódico británico en la India para / resaltar dos diferentes opiniones.

Comentando sobre el papel del Director de Estadística la Comisión Industrial de India de 1918 decía:

El Director de Estadística debe ser solamente un oficial recopilador. Su relación con las estadísticas debe ser meramente aritmética, es decir, no debe hacer comentarios / sobre ellas.

La preparación de previsiones era un asunto altamente técnico. La práctica existente por la cual el Director de Estadística era responsable de aprobar las previsiones de / los encargados provinciales, fue derogada. Dicha tarea searía más apropiada para ser realizada por un responsable / técnico con experiencia general de la agricultura hindú.

Por otra parte la Conferencia Estadística Imperial / que tuvo lugar en Londres en 1920 apuntaba:

El trabajo aritmético de un estadístico oficial constituye solamente el trabajo preparatorio y menos importante / de esta oficina. Si se ignorase este principio, los servicios más importantes que en base al trabajo estadístico, / se pueden rendir a la comunidad, serían sacrificados.

Los anteriores puntos de vista contrarios, mantenidos desde hace más de 60 años, todavía persisten a pesar de / la aparición de la estadística como una disciplina separada y el reconocimiento de la estadística como una herramienta indispensable en las investigaciones científicas, especialmente en las ciencias sociales y en la toma de decisiones.

La distinción entre abastecedores y usuarios de la estadística es desafortunada. Si es responsabilidad del estadístico suministrar datos útiles que sean actuales, precisos y sin lagunas, él debe tener una idea apropiada de la utilización posterior de los datos y también estar involucrado personalmente en análisis de datos y en la reso-

lución de problemas. La utilidad de los datos no se puede entender en términos absolutos. No existen datos completamente exactos (5), (6) ó sin lagunas. Su valor es relevante sólo en términos de su utilidad para resolver problemas determinados.

No se sugiere que las oficinas de estadística gubernamentales se transforman en departamentos de investigación para producir artículos académicos, pero una oficina de estadística debe tener la necesaria experiencia / para realizar las siguientes funciones:

(i) Encargarse de la investigación de los métodos de recogida de datos (tipos de cuestionarios a utilizar, contratación de agencias para la recogida de datos, métodos de depuración e imputación, información secundaria útil para detectar errores e ilustrar resultados).

(ii) Hacer un mínimo análisis estadístico que ayude a juzgar la adecuación de los datos dados para tomar ciertas decisiones políticas y a evaluar los errores que se cometen al hacer previsiones. Este tipo de investigación debe orientarse a la mejora de datos, a la compilación de / síntesis de resultados más significativos tales como números índice, estadísticas input / output, análisis de series cronológicas, predicciones a corto plazo útiles al Gobierno para tomar decisiones administrativas diariamente.

Las funciones de una oficina estadística deberían de ir más allá de la recogida de datos, ampliándose a su análisis que es necesario para mantener un sistema estadístico apropiado, para hacer posible una mayor comunicación sobre los datos que necesitan los que toman decisiones y para informarles de la capacidad actual de la estadística.

### 3. POLITICA PUBLICA

No hay que negar que muchas decisiones políticas tomadas por el Gobierno que nos afectan a cada uno de nosotros dependen crucialmente de su base estadística, aún más se puede argumentar que ninguna otra profesión tiene actualmente tanta responsabilidad en los asuntos públicos como nosotros (los estadísticos). Claus Moser.

Cual es la mejor forma en que los estadísticos gubernamentales pueden contribuir a los esfuerzos nacionales para el beneficio de la humanidad? Como pueden ellos ayudar al gobierno especialmente en los países en vías de /

desarrollo, a encontrar nuevas fórmulas que acaben con la pobreza y aseguren un deseable crecimiento económico?

La eficacia de un gobierno depende de la naturaleza / de las decisiones que toma. La calidad de las decisiones a su vez depende de la disponibilidad de datos actuales y precisos y también de lo que es igualmente tan importante un correcto análisis estadístico de los datos. La recopilación de datos y el análisis estadístico requieren habilidades para cuyo ejercicio el estadístico está especialmente entrenado. Por lo tanto, un estadístico tiene la / responsabilidad y un importante papel que jugar en los problemas nacionales. Argumentar que un estadístico no / tiene cabida en la mesa de decisiones políticas y que él es solamente un observador de hechos puros, es ignorar el cuerpo entero de la rama especializada del conocimiento / conocida como metodología estadística por la cual se llega a decisiones óptimas a partir de restricciones dadas y para lograr las metas deseadas.

Muchos países tienen adecuadas facilidades para preparar a los estadísticos a diversos niveles, y también sistemas estadísticos nacionales bien desarrollados en relación a datos estadísticos rutinariamente recogidos en la Administración, así como en relación a datos específicos recogidos a través de censos y encuestas muestrales. Sin embargo a pesar de tanta experiencia y competencia estadística, de tantos datos estadísticos y disponibles, debo confesar con un sentimiento de insatisfacción que la ciencia y la profesión estadística no están cumpliendo su cometido de lograr un mejor entendimiento cuantitativo de / los problemas socio-económicos y de sugerir y proponer soluciones factibles a los mismos. La razón de este serio / fallo proviene de la falta de relación entre los diferentes grupos de expertos que parece que trabajan independientemente unos de otros. Tenemos estadísticos profesionales que recogen datos sin estar enterados de sus usos, estadísticos académicos que desarrollan teoría estadística para el análisis de datos sin ninguna relevancia respecto a los problemas reales y científicos sociales que / formulan decisiones políticas basadas en datos con los cuales no están familiarizados, usando análisis estadísticos que pueden ser no eficientes.

Como podemos unir a todos estos expertos, coordinar / sus actividades y usar sus conocimientos especializados / para lograr los mejores resultados posibles? Estoy convencido de que esto sólo se puede llevar a cabo mediante un proceso gradual. Sugiero que este esfuerzo debe hacerse / en varias direcciones:

- Los departamentos estadísticos en las Universi-

dades se han mantenido o bien integrados en los departamentos de matemáticas, o donde son independientes, han tendido a mantenerse separados de las áreas aplicadas en general y de las ciencias sociales en particular. En los departamentos de Ciencias Sociales el cuerpo de profesores y los estudiantes no sólo carecen de una / adecuada experiencia matemática sino que de hecho son generalmente débiles en razonamiento matemático y cuantitativo. Por tanto, hay un gran vacío entre las Ciencias Estadísticas y las / Ciencias Sociales en las Universidades. Se deben realizar serios esfuerzos en reorganizar / cursos de estadística y ciencias sociales para llenar ese vacío y formar estadísticos profesionales con predisposición hacia las aplicaciones y profesionales de ciencias sociales con una / tendencia cuantitativa y competentes en análisis cuantitativos.

- Se debe reclutar a la vez estadísticos y de / ciencias sociales en las oficinas de Estadística y en las divisiones políticas del gobierno / para proporcionar una apropiada conjunción de / personal con diferentes técnicas que actúen conjuntamente y produzcan los mejores resultados / posibles.
- Debe tenerse en cuenta la posibilidad, de invitar a expertos de Universidades e Institutos de Investigación para trabajar en las oficinas de estadística gubernamentales y la de enviar estadísticos del Gobierno a enseñar en las Universidades aspectos prácticos de la estadística. Tal cambio de expertos debe dar unos efectos beneficiosos al mejorar la eficacia y utilidad de las oficinas de estadística del Gobierno y también en la creación de estadísticos mejor preparados para trabajar en las oficinas estadísticas después de completar su educación universitaria.

El hacer política ya no es un juego de imprevisibles posibilidades de acierto o un negocio a la aventura. Ahora está dentro del campo de las técnicas de la ciencia el tomar decisiones óptimas sobre la base de la evidencia / disponible y los resultados pueden ser continuamente controlados por realimentación y control.

#### 4. ALGUNOS PROBLEMAS NUEVOS

... los riesgos que implica la política económica y social son enormes y nosotros (los estadísticos) tenemos una buena parte de la responsabilidad. Harold Wilson.

Hay grandes áreas de la estadística social que están descuidadas porque su importancia en el contexto de la / planificación no ha sido bien entendida. En un reciente / estudio <sup>(7)</sup>, se hizo una comparación de los perfiles socio-económicos de seis países durante el período 1950- / 1965. Se encontró que la presencia o la ausencia de rigideces en la estructura social, y las divisiones basadas / en las diferencias étnicas influyen en el crecimiento económico. La consideración de factores sociales ha introducido una nueva dimensión en la planificación económica y da una mayor responsabilidad a los estadísticos para desarrollar un sistema integrado económico y social de Estadística, y para dirigir estudios sobre la influencia de / los factores sociales en el desarrollo económico.

Mientras que tenemos estadísticas económicas adecuadas para calcular magnitudes como el PNB, no tenemos aún datos básicos sobre aspectos sociales tales como ingresos y gastos de familias, tamaño de la familia, status educacional y sanitario de los individuos, que son necesarios para construir indicadores sociales. Creo que aproximadamente dos de cada cinco estadísticos profesionales en U.K están ahora trabajando sobre estadísticas sociales. Este porcentaje era muy pequeño hace cinco años, lo cual indica la gran prioridad concedida a las estadísticas sociales, en el U.K. (Reino Unido) en los últimos años. El uso de indicadores sociales como funciones objetivas además / del PNB es muy importante para la planificación de los / países en vías de desarrollo.

Durante mucho tiempo hubo la creencia general de que cualquier tipo de educación confiere habilidades productivas al educado, el único tema a debatir era si el flujo / de ganancias adicionales que acumula el educado a causa de las habilidades adquiridas puede compararse al del no educado en proporción al costo de la educación. En resumen, desde el punto de vista ortodoxo la adquisición de / educación es como cualquier otra inversión, sólo que pertenece al ser humano. Tenemos aquí por consiguiente la / teoría del capital humano.

Recientemente, a causa de las realidades observadas / ha surgido un punto de vista totalmente nuevo y contradictorio del capital humano. Es la denominada Teoría de la / Selección de la educación, expuesta entre otros por / Kenneth J. Arrow Premio Nobel de Economía. Una versión extrema de esta teoría sostiene que la educación no confiere habilidades. La contribución productiva del individuo a la sociedad depende sólo de su habilidad innata y no de la educación. Por otra parte los que poseen una mayor habilidad tienen una mayor posibilidad de éxito de obtener grado universitario. En este caso el grado universitario

sirve únicamente como filtro ó criba para separar aquellos con habilidades innatas potencialmente mayores, del resto. Aún así cuando desde el punto de vista de la sociedad el grado universitario no confiere ventajas productivas, y cualquier utilización de los recursos en proporcionar una educación universitaria es un gasto inútil el obtener un grado universitario es evidentemente una / ganancia positiva para los individuos y ofrece a los empresarios información además de los que tienen mayor habilidad innata obteniendo así sueldos más altos. La ventaja para los empresarios es también evidente, pueden / contratar empleados con mayores capacidades innatas eligiendo aquellos con grado universitario. Esto no es sorprendente, por ello hay una demanda creciente de educación, que los gobiernos están forzados a satisfacer. La rápida expansión de la educación, particularmente la superior, es un indicador en esta dirección. Naturalmente este es un punto de vista extremo, algunos tipos de educación denominadas educación técnica y vocacional evidentemente confieren habilidades. Pero la mayor parte de la expansión en nuestra educación superior consiste en proveer de un grado universitario en ciencias o en Letras y aquí parece bastante probable que la teoría de la selección sea relevante.

Es de gran importancia conocer si la educación confiere habilidades o no y qué tipo de educación confiere habilidades útiles. Aunque en el pasado se han hecho intentos para estimar estadísticamente la contribución de la educación en las diferencias observadas de los salarios de los educados (universitarios) comparados con los de los no educados, la mayor parte de estos cálculos no son satisfactorios porque se confunden los efectos de las habilidades innatas con las que proporciona la educación. Cualquiera que pueda desarrollar elementos estadísticos y técnicos para aislar los efectos de la educación, proporcionará una base política educativa racional para un buen número de países en vías de desarrollo y de países pobres

Otro problema que interesa actualmente es la situación de personas desnutridas en países en vías de desarrollo. Los únicos datos disponibles de cada país son estimaciones indirectas del consumo de alimentación per cápita (con un coeficiente de error desconocido) y unas distribuciones empíricas de los ingresos (gastos) con toda su falta de fiabilidad. Aún disponiendo de estimaciones fiables de estas cantidades, no es posible obtener estimaciones / válidas de las personas desnutridas en cada país sin hacer algunas suposiciones sobre ingestión de calorías, / elasticidad de renta, modelos de distribución y variabilidad desconocida de las necesidades caloríficas individuales. Las estimaciones recientes del número de personas /

desnutridas hechas por la FAO y del Banco Mundial difieren considerablemente porque están basadas en diferentes conjuntos de suposiciones. Es difícil justificar cualquiera de las dos estimaciones a la vista de la extrema falta de adecuación base de los datos y a las hipótesis razonables hechas por estas organizaciones. Creo que la FAO proyecta una encuesta mundial para recoger datos a / fin de determinar la extensión de la desnutrición en los países en vías de desarrollo. Este es un problema interesante y extremadamente importante donde los estadísticos y los profesionales de ciencias sociales pueden colaborar haciendo una contribución útil.

## 5. CONCLUSIONES

Para concluir digamos lo siguiente:

Las responsabilidades de los estadísticos no terminan sólo con proveer datos actuales y precisos. Deben / participar en el análisis e interpretación de los datos. El análisis de datos también permite la retroalimentación para su mejora y para las futuras necesidades de / los mismos.

Debe haber una mayor colaboración entre los estadísticos y los profesionales de ciencias sociales que puede conseguirse mediante intercambios convenientes de los / cursos que se dan en las Universidades, y en la política de reclutamiento de los departamentos gubernamentales para conseguir la necesaria interacción entre estadísticos y profesionales de ciencias sociales.

Con el énfasis dado a los indicadores sociales y los índices de desarrollo se han abierto nuevos caminos en / las estadísticas oficiales haciendo recaer responsabilidades más pesadas sobre los estadísticos gubernamentales

El conocimiento estadístico es un recurso nacional y como tal debe ser explotado.

Las buenas estadísticas no pueden sustituir a los / buenos estadísticos y de forma similar los buenos estadísticos no sustituyen a las buenas estadísticas. Necesitamos de ambos. Pero esto no es suficiente, también es / necesario que los estadísticos trabajen junto a los políticos en la toma de decisiones, si es que el país quiere beneficiarse al máximo del conocimiento estadístico disponible.

En un discurso dirigido a los delegados de la International Statistical Conference que tuvo lugar en 1951 / el entonces Primer Ministro de la India Jawaharlal Nehru

destacó la importancia de las estadísticas para comprender los problemas del país y buscar soluciones. Sus palabras fueron:

*Yo creo que en la India, y posiblemente en otras partes del mundo también, se está haciendo un intento de responder preguntas antes de haberlas hecho. / Desde luego este es un hecho extraordinario, que to-*

*dos respondamos a las preguntas sin conocer cuales / son. En otras palabras todo el mundo está buscando / algún remedio sin conocer cual es la enfermedad.*

*La estadística es una entrada al conocimiento y al / progreso, un país depende vitalmente de la eficiencia de su sistema estadístico.*

- (1) Claus Moeser (1976). El papel de la oficina central de Estadística, como asistencia a los usuarios públicos. *The American Statistician*, 30, 59-66.
- (2) En la India, la Econometric Society mantienen seminarios periódicos sobre las bases de datos de la economía de la India y publicaciones con los trabajos presentados sobre las necesidades y perfeccionamiento de los datos junto con recomendaciones para el Gobierno.
- (3) Puede hacerse referencia al artículo "Informe y Comentarios sobre / el desarrollo del Sistema Estadístico Federal", *The American Statistician*, 35, 183-209, fechado en Noviembre 1981.
- (4) Subramanian S. (1969), Sabiduría Estadística - vieja ó nueva *J. Roy Statist. Soc. A*, 132, 229-234.
- (5) Weirus, físico alemán del siglo XVI, tiempo en que la mayoría de / los países europeos estaban afligidos por el miedo a los demonios y las brujas, calculó que exactamente 7.405.926 demonios habitaban la tierra. La mayoría de la gente pensó que la cifra 7.405.926, era / bastante exacta porque para ellos eran en realidad y Weirus era un hombre instruido.
- (6) La historia trata sobre un hombre que cuando se le preguntó acerca / de la edad de un cierto río replicó que tenía 3.000.004 años. Cuando se le preguntó como podía dar una cifra tan exacta, su respuesta fue que hace cuatro años se le informó que ese río en particular tenía tres millones de años.
- (7) Informe de la U. N. Research Institute for Social Development, Ginebra, titulado "Levels of Living and Economic Growth".

**UN EJEMPLO NATURAL DE UNA DISTRIBUCION  
BINOMINAL PONDERADA**

**1.- UN RESULTADO EMPIRICO**

Si se pide a cada estudiante de sexo masculino de // una clase que escriba el número de hermanos y de hermanas que tiene, incluyéndose él mismo, surge la siguiente cuestión. Sea  $k$  el número de estudiantes de sexo masculino interrogados, y  $B$  y  $S$  el número total de hermanos y hermanas que tienen. ¿Cuál sería aproximadamente el valor de  $B/(B + S)$ , la razón de hermanos al número total de hijos? La tendencia general es decir, que  $B/(B + S)$  se aproxima a 1/2 o es ligeramente superior. La respuesta es, por supuesto, errónea. Sorprendentemente, cuando el número de muchachos en clase no es muy pequeño, se puede hacer una predicción exacta de las magnitudes relativas de  $B$  y de  $S$  y de la razón  $B/(B + S)$ . Esto debe anunciarse como un teorema empírico.

**Teorema Empírico:** Sean  $k$  los muchachos observados en una clase o en cualquier agrupación que tienen un total de hermanos (incluidos ellos mismos) y un total de  $S$  // hermanas. Entonces pueden hacerse las predicciones siguientes:

- (i)  $B$  es mucho mayor que  $S$ .
- (ii)  $B - k$  es aproximadamente igual a  $S$ .
- (iii)  $(B - k)/(B + S - k)$  se approxima a  $\frac{1}{2}$ .
- (iv)  $B/(B + S)$  es mayor que  $\frac{1}{2}$ ,  $[B/(B + S)] - [K/2(B + S)]$  se approxima a  $\frac{1}{2}$ .

Antes de dirigir el experimento ha escrito el teorema en un folio y se lo ha dado a uno de los estudiantes en un sobre cerrado, que se abrirá después de que sean recogidos los datos para su verificación. Los papeles / de  $B$  y de  $S$  están invertidos en el Teorema Empírico si los datos se obtienen en una clase o agrupación de muchachas.

La Tabla 1 ofrece una recopilación de datos de estudiantes del sexo masculino comprendida entre 19 y 24 años y que están estudiando cursos de Estadística en // Universidades de diferentes lugares y países. La coincidencia con los valores predichos parece ser buena como lo indican las figuras de las tres últimas columnas. Los valores de las dos últimas columnas se aproximan  $\frac{1}{2}$ . Además,  $A$  y  $B - k$  están en razón 1:1.

En la Sección 2 se da un esbozo del modelo en el que están basadas las predicciones.

**2.- CONSIDERACIONES TEORICAS-MUESTREO PROPORCIONAL AL TAMAÑO**

Empecemos con una distribución binomial  $b(n, \pi)$  con índice  $n$  y probabilidad de acierto  $\pi$ . La probabilidad de  $r$  aciertos es

$$cp_r \binom{n}{r} \pi^r (1 - \pi)^{n-r}, r = 0, 1, \dots, n \quad (2.1)$$

Construyamos otro modelo donde la probabilidad del evento  $r$ -ésimo se cambia a

$$cp_r \binom{n}{r} \pi^r (1 - \pi)^{n-r}, r = 0, \dots, n \quad (2.2)$$

donde  $p_r$  son los números no negativos asignados y  $c$  es / la constante normalizada. Obsérvese que la distribución / (2.2) es aplicable si hacemos a partir de (2.1) lo que / los estadísticos encargados de las encuestas llaman muestreo proporcional al tamaño de la muestra, es decir, damos probabilidades desiguales a los eventos con diferentes valores de  $r$ . En un trabajo anterior (Rao, 1965) la distribución (2.2) se definía como una distribución binomial ponderada. En un estudio reciente de Patil y Ord. (1975) vienen dados los desarrollos subsecuentes sobre / dichas distribuciones, así como en las referencias dadas en su trabajo.

El caso  $p_r = r$  es de mucho interés. En este caso, (2.2) se reduce a

$$\cdot \binom{n-1}{r-1} \pi^{r-1} (1 - \pi)^{n-r}, r = 1, \dots, n \quad (2.3)$$

que es de nuevo binomial con índice  $n$  y la observación  $r$  reducida en una unidad. Si el modelo apropiado para nuestro muestreo es (2.3), entonces  $E(r-1) = (n-1)\pi$  dando

$$E(r) = n\pi + (1 - \pi) \quad (2.4)$$

y no  $n\pi$  como en el caso (2.1).

Interpretaremos los datos recogidos a los muchachos en función de la nueva distribución (2.3) con elección especial de  $p_r$  en (2.2). La distribución del número de hijos varones en familias con un número dado de hijos, di-

TABLA 1.

Verificación de las predicciones basadas en los datos recopilados sobre estudiantes de sexo masculino de edades entre 19 y 24 años.

Fuente y fecha de la encuesta	k	B	S	B - K	$\frac{S}{B + S}$	$\frac{2B - k}{2(B + S)}$	$\frac{B - k}{B - k - S}$
Teheran (Iran)	Abril 1975	21	65	40	.619	.519	.524
Ispahan (Iran)	Abril 1975	11	45	32	.584	.513	.515
Tokio (Japón)	Junio 1975	50	90	34	.726	.524	.540
Delhi (India)	Febr. 1975	29	92	66	.582	.490	.488
Calcuta (India)	Agos. 1963	104	414	312	.570	.498	.498
Waltair (India)	Junio 1969	39	123	88	.583	.491	.488
Ahmedabad (India)	Agos. 1975	29	84	49	.632	.523	.529
Bangalore (India)	Agos. 1975	55	180	127	.586	.496	.497

gamos  $n$ , es  $b(n, 1/2)$ , es decir, binomial con  $\pi = 1/2$  / (al menos aproximadamente). Supongamos que elegimos una familia de acuerdo con algún mecanismo, el cual no proporciona probabilidades iguales a las familias, sino // probabilidades proporcionales a los números de sus hijos varones. Entonces el modelo apropiado para nuestras observaciones es (2.3) y no (2.1).

Así, si  $b$  y  $s$  son los números de hermanos y hermanas en una familia como nos informó un muchacho, resulta:

$$E(b - 1/b + s) = \frac{b + s - 1}{2} \quad (2.5)$$

donde  $E(x/y)$  designa la esperanza condicional de  $x$ , dado  $y$ . Si  $(b_i, s_i)$ ,  $i = 1, \dots, k$  son observaciones sobre muchachos, entonces, con  $B = \sum b_i$ ,  $S = \sum s_i$ ,

$$E(B - k/B + S) = \frac{B + S - k}{2} \quad (2.6)$$

lo que implica

$$E \left[ \frac{B - k}{B + S - k} \right] = \frac{1}{2} \quad (2.7)$$

y

$$E \left[ \frac{B}{B + S} - \frac{k}{2(B + S)} \right] = \frac{1}{2} \quad (2.7)$$

Los resultados en (2.7) forman la base de los enunciados (i) - (iv) del Teorema Empírico, utilizando los teoremas del límite habituales. Podemos establecer el segundo resultado en (2.7) de la forma

$$\frac{B}{B + S} \text{ es aproximadamente } \left( \frac{1}{2} + \frac{1}{2f} \right) \quad (2.8)$$

donde  $f = [B + S]/k$ , el tamaño de la familia media. De este modo,  $B/(B + S)$  sobreestima la verdadera probabilidad/ en  $(1/f)$ .

### 3.- DISCUSIÓN ACERCA DEL MECANISMO DE SELECCIÓN

Demostramos el Teorema Empírico sobre la suposición / de que la probabilidad de selección de una familia es proporcional al número de hijos varones. Esto es cierto si / nuestra estructura de muestreo es la relación de todos // los hijos varones y se selecciona una familia cuando se / observa a un hijo varón perteneciente a ella. Pero nuestra estructura de muestreo es un poco diferente, siendo / los hijos varones de un grupo de determinada edad.

Designemos a un hijo varón de este grupo de edad por  $M$  y supongamos que aquí solamente puede haber uno con // edad  $M$  en una familia. Entonces, debido a la naturaleza/ de nuestra estructura de muestreo, la probabilidad de que la familia seleccionada tenga  $r$  hijos varones de  $n$  es

$$P(r/M, n) = cP(M/r, n) \binom{n}{r} \pi^r (1 - \pi)^{n-r} \quad (3.1)$$

donde  $c = 1/P(M/n)$  y  $P(M/r, n) = P_r$ , es decir, es la probabilidad de que una familia con  $r$  hijos varones entre  $n$  tenga una edad  $M$  ( $r = 1, 2, \dots, n$ ). Supongamos que seleccionamos al azar un hijo varón del grupo de todos los hijos varones, pertenecientes a familias que tienen  $r$  hijos varones entre  $n$ . La probabilidad de que el hijo seleccionado sea de edad  $M$  es  $\pi^{-1} p_r$ . Si es razonable creer que  $\pi^{-1} p_r = p_1$ , entonces  $p_r = \pi p_1$ , en cuyo caso, es aplicable el modelo (2.3) (haciendo  $p_0 = 0$ ). En los datos de la Tabla 1 parece aplicarse dicho mecanismo. He recogido datos similares procedentes de estudiantes graduados y profesores

T A B L A 2.

Resumen de datos recopilados sobre estudiantes graduados y profesores universitarios de sexo masculino (25-60 años)

Fuente y fecha de la encuesta	k	B	S	B - k	$\frac{B}{B - S}$	$\frac{2B - k}{2(B + S)}$	$\frac{B - k}{B - k - S}$
Varsovia (Polonia)	Agosto 1975	18	41	21	.660	.525	.523
Poznan (Polonia)	Septi. 1975	24	50	17	.746	.567	.605
State College (USA)	Agosto 1975	30	87	39	.690	.571	.546
Bilbao	Septi. 1983	24	95	56	.629	.549	.559

T A B L A 3.

Resumen de datos recopilados sobre estudiantes graduados y profesores universitarios de sexo femenino (20-40 años)

Fuente y fecha de la encuesta	k	S	B	S - k	$\frac{S}{B + S}$	$\frac{2S - k}{2(B + S)}$	$\frac{S - k}{B + S - k}$
Poznan (Polonia)	Septi. 1975	20	42	15	.737	.562	.594
State College (U.S.A.)	Agosto 1975	12	34	16	.680	.560	.579

res de Departamentos de Estadística, del grupo de edad // comprendido entre los 25 y los 60 años, en otras tres // universidades, Pennsylvania State University, en el State College de Estados Unidos, en Agricultural University Poznan (polonia) y en la Universidad del País Vasco / en Bilbao. Además, los datos también fueron obtenidos de las delegaciones asistentes a una conferencia sobre la / enseñanza de la estadística en Varsovia (Polonia). Los / datos y valores predichos de acuerdo con el Teorema Empírico están recogidos en la Tabla 2 para hombres y en la Tabla 3 para mujeres. Se ve que los valores en las dos / últimas columnas de la Tabla 2 y de la Tabla 3 son algo mayores de  $\frac{1}{2}$ , indicando la posibilidad de que las probabilidades del muestreo para las familias son mayores que los cocientes, en el caso de estudiantes graduados y profesores universitarios. Esto puede esperarse cuando los / estudiantes graduados y profesores universitarios están/ supeditados a selección bietápica, una cuando ingresan / en la facultad y otra cuando son seleccionados para /// puestos de investigación o de enseñanza. En la primera etapa, las probabilidades de seleccionar a un muchacho de una familia con  $n$  muchachos de  $n$  puede ser proporcional/ a  $n$ . Esto da lugar a la distribución binomial ponderada/

discutida en la sección 2. Si los investigadores y profesores son reclutados entre aquellos que han completado/ estudiados medios y se han licenciado, parece que debe haber una selección mayor a favor de las familias con un / número mayor de varones. Será de interés determinar que/ probabilidades de selección operan en las dos etapas.

## REFERENCES

- 1.- FISHER, R.A. (1934): The effect of the methods of // ascertainment upon the estimation of frequencies. // Ann. Eugen. London, 6, 13-25.
- 2.- PATIL, G.P. and ORD. J.K. (1975): On size-biased // sampling and related form invariant weighted distributions. Sankhya. Serie A, 38.
- 3.- RAO, C. RADHAKRISHNA (1965): On discrete distributions arising out of methods of ascertainment. Clas- sical and Contagious Distributions. G.P. Patil(ed.) S.P.S. Calcutta, pp. 320-332. Also reprinted in San- khya A. 27, 311-324.

## INFERENCIA DEL MODELO LINEAL CON LOS EFECTOS ESTABLECIDOS - RESULTADOS RECIENTES Y DIVERSAS RESPUESTAS.

### 1.- INTRODUCCION

Se considera el modelo general de Gauss-Markoff

$$y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 V) \quad (1.1)$$

donde las matrices  $X$  y  $V$  pueden ser deficientes de rango, se analizan los resultados recientes sobre inferencia del modelo (1.1) cuando  $\beta$  y  $\sigma^2$  son desconocidas y se presentan algunos problemas de interés en aplicaciones prácticas. La Sección 2 trata de las aproximaciones unificadas a la estimación lineal insesgada óptima aplicable al caso general, cuando  $X$  y  $V$  son posiblemente deficientes en rango. La Sección 3 presenta criterios para la estimación como alternativas a la insesgacidad y a la varianza mínima. La Sección 4 trata de estimación y predicción simultáneas. La Sección 5 está dedicada a la robustez de algunos procedimientos de inferencia. La Sección 6 presenta los conceptos de suficiencia y complejidad en los modelos lineales sin hacer hipótesis sobre la distribución. La Sección 7 estudia la selección del modelo con propósitos de predicción. La Sección 8 examina un aspecto no deseable de los estimadores empíricos de Bayes. Los temas no incluidos explícitamente son la admisibilidad de los estimadores lineales [Rao, (1976); Klonecki, (1979); La Motte, (1980)], la robustez de estimadores para errores en la especificación / de las matrices  $X$  y  $V$  [Rao, (1967a); Zyskind, (1967)] , y la estimación  $X\beta$  cuando  $V$  tiene una estructura específica con parámetros desconocidos [Rao, y Kleffe, (1980)].

En esta conferencia se utilizan las notaciones siguientes:

- (1)  $\rho(A)$  designa el rango de una matriz y  $R(A)$ , el espacio vectorial generado por las columnas de  $A$ .
- (ii)  $A^-$  es una g-inversa de  $A$  (definida por  $AA^-A = A$  con las propiedades expuestas en Rao, 1962).
- (iii)  $Z$  es una matriz de rango completo tal que  $X'Z = 0$  donde  $X$  es la matriz del diseño en el modelo lineal (1.1). También se escribe  $Z = X^T$ . Obsérvese que  $R(X) \cap R(VZ) = 0$  que implica

$$\rho(V:X) = \rho(X:VZ), \quad \rho(VZ) = \rho(V:X) - \rho(X) = \rho(Z'VZ) \quad (1.2)$$

- (iv)  $P_{A/B}$  se define como el operador de proyección generalizado sobre  $R(A)$  a lo largo de un subconjunto / disjunto  $R(B)$  aunque  $R(A) \cup R(B)$  puede no ser el espacio completo si

$$P_{A/B}y = y_A, \quad y_A \in R(A:B)$$

Donde  $y_A$  es el componente único de  $y \in R(A:B)$  en la descomposición  $y = y_A + y_B$ ,  $y_A \in R(A)$  e  $y_B \in R(B)$  / [ver Rao (1974)] y Rao y Yanaia (1979) P<sub>A/R</sub> en // (1.3) sólo necesita satisfacer las condiciones // no necesariamente únicas.

$$PA = A, \quad PB = 0 \quad (1.4)$$

Los siguientes hechos acerca del modelo (1.1) se usan en la exposición de diversos resultados.

(1)  $Y \in R(V:X)$  con probabilidad 1.

(2)  $\beta_1 \neq \beta_2 \Leftrightarrow X\beta_1 \neq X\beta_2$  si y sólo si  $X$  es de rango completo, lo cual implica que  $\beta$  no es identificable si  $X$  es deficiente en rango. Una función parámetrica lineal  $p'\beta$  se dice que es identificable si  $p'\beta_1 \neq p'\beta_2 \Rightarrow X\beta_1 \neq X\beta_2$  para lo cual una condición necesaria y suficiente es que  $p' \in R(X')$ . Esta condición es idéntica a la de estimabilidad insesgada de  $p'\beta$  por medio de una función lineal de las observaciones. Desde luego, no tiene sentido para estimar funciones parámetricas no identificables.

### 2.-APROXIMACION UNIFICADA A LA ESTIMACION LINEAL INSESGADA OPTIMA

#### 2.1.- PROYECCION GENERALIZADA

Sea  $(y, X\beta, \sigma^2 V)$  el modelo lineal en que  $X$  y  $V$  / pueden ser deficientes en rango. A partir de (iii), Sección 1,  $R(X) \cap R(VZ) = 0$  de modo que existe un operador de proyección  $P$  sobre  $R(X)$ , a lo largo de  $R(VZ)$ , que satisface las ecuaciones

$$PX = X, \quad PVZ = 0 \quad (2.1)$$

Obsérvese que, resolviendo (2.1)

$$P = X(X'V^{-1}X)^{-1}X'V^{-1}, \quad (\text{cuando } V \text{ es no singular}) \quad (2.2)$$

$$= X(X'X)^{-1}X, \quad [\text{cuando } V = I, \text{ ver Rao (1967b)}] \quad (2.3)$$

$$= X(X'TX)^{-1}X'T, \quad (\text{en general}) \quad (2.4)$$

siendo  $T = (V+XUX')^{-1}$  para cualquier g-inversa, con  $U/$

tal que  $\rho(V+XUX') = \rho(V:X)$ . Se tiene la proposición/general siguiente que responde a una cuestión propuesta/por Kempthorne (1976), acerca de operadores de proyección cuando  $V$  es singular.

*Proposición 2.1* sea  $P$  un operador de proyección como se define en (2.1), y sea  $L$  una matriz tal que  $LX = X$ . Entonces

$E(PY - X\beta) (PY - X\beta)' \leq E(LY - X\beta) (LY - X\beta)'$  es decir,  $PY$  es el estimador insesgado de mínima dispersión del parámetro vectorial identificable  $\theta = X\beta$ .

*Demostración* Hagamos  $M = L - P$  de modo que  $MX = 0$  // o  $M = AZ'$  para algún  $A$ . Entonces

$$\begin{aligned} E[L(Y - X\beta) (Y - X\beta)'] &= (M + P) V (M + P)' \\ &= MVM' + PVP' \geq PVP' \\ &= E[P(Y - X\beta) (Y - X\beta)'P] \end{aligned}$$

puesto que  $MVP' = AZ'VP' = 0$  por (2.1). La proposición está demostrada y facilita los resultados posteriores.

*Nota 1*  $P_y$  es único con probabilidad 1 para todo  $P$  que satisface (2.1), puesto que  $y \in R[X:VZ]$  con probabilidad 1.

*Nota 2* Si  $p'\beta$  es una función paramétrica identificable,  $p'\beta = \lambda'X\beta$  y  $\lambda'PY$  es el estimador insesgado de mínima varianza de  $p'\beta$ . Esto es consecuencia inmediata de  $PY$  que es el estimador insesgado de mínima dispersión.

*Nota 3* Eligiendo  $P$  como en (2.4)  
 $\lambda'PY = \lambda'X(X'TX)^{-1}X'Ty = p'\hat{\beta}$   
 siendo  $\hat{\beta} = (X'TX)^{-1}X'Ty$ .

*Nota 4*  $\hat{\beta}$  como ha sido definida en la Nota 3 es un valor estacionario de la forma cuadrática  
 $(Y - X\beta)'T(Y - X\beta) \quad (2.5)$   
 y por consiguiente, se considera a  $\hat{\beta}$  como un estimador mínimo cuadrático con respecto a  $T$ . Desde luego, si  $V$  es no singular,  $T = V^{-1}$  es una elección obvia, pero  $T$  podría ser también  $(V + XUX')^{-1}$  siendo  $U$  una matriz definida no negativa.

*Nota 5* Un estimador insesgado de  $\sigma^2$  es  
 $\sigma^2 = f^{-1}(Y - X\hat{\beta})'T(Y - X\hat{\beta})$   
 $= f^{-1}y'T(I-P)y, f = \rho(V:X) - \rho(X).$

*Nota 6* Con  $\hat{\beta} = (X'TX)^{-1}X'Ty$ ,  
 $V(p'\hat{\beta}) = \sigma^2 p'[(X'TX)^{-1}U]p$   
 $Cov(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'[(X'TX)^{-1}U]q$

*Nota 7* En Rao (1973) se ha demostrado que la  $T$  de la forma cuadrática (2.4) debe ser de la forma  $(V + XUX')$  si los valores estacionarios proporcionan la estimación lineal insesgada óptima.

La proposición 2.1 y las notas anteriores suministran la geometría de la estimación a través del concepto de proyección generalizada y el álgebra de la estimación a través de una técnica del tipo de mínimos cuadrados, sin hacer ninguna hipótesis acerca de  $\rho(X)$  y  $\rho(V)$ .

Puede mencionarse que el caso  $\rho(X) < n$  y  $\rho(V) = n$  fue estudiado en primer lugar por R.C. Bose, quien demostró que en tal caso algunas funciones paramétricas no son estimables insesgadamente.

## 2.2.- MINIMOS CUADRADOS CON RESTRICCIONES

El caso  $\rho(V) < n$  fue estudiado en primer lugar por Goldman y Zelen (1964) quienes redujeron el problema a uno de la teoría de los mínimos cuadrados con restricciones en los parámetros. Mitra y Rao (1968) hicieron una formulación ligeramente más sencilla.

*Proposición 2.2* Sea  $N$  una matriz de rango  $s = n - \rho(V)$  tal que  $N'V = 0$  lo que implica que  $N'Y = N'X\beta$  con probabilidad 1, y  $V^-$  una g-inversa de  $V$ . Además sea  $\hat{\beta}$  tal que

$$\min(Y - X\beta)'V^-(Y - X\beta) = (Y - X\hat{\beta})'V^-(Y - X\hat{\beta}) = R_0^2. \quad (2.6)$$

$$N'X\beta = N'Y$$

Entonces:

- (i)  $\hat{\beta}$  es el estimador lineal insesgado óptimo de  $p'\beta, p \in R(X')$
- (ii)  $R_0^2$  es un estimador insesgado de  $\sigma^2$ , siendo  $f = \rho(V:X) - \rho(X)$
- (iii)  $R_0^2 \sim \sigma^2 X^2(f)$  si  $Y \sim N_n(X\beta, \sigma^2 V)$
- (iv) Si  $Y \sim N_n(X\beta, \sigma^2 V)$  y  $K'\beta = w$  es una determinada hipótesis consistente sobre  $\beta$ ,  $R_1^2 - R_0^2 \sim \sigma^2 X^2(h)$  independientemente de  $R_0^2 \sim \sigma^2 X^2(f)$  siendo  $h = \rho[D(K'\beta)]$

$$R_1^2 = \min(Y - X\beta)'V^-(Y - X\beta). \quad (2.7)$$

$$N'X\beta = N'Y$$

$$K'\beta = w$$

*Nota 1* La proposición 2.2 es esencialmente la teoría de mínimos cuadrados con restricciones en los parámetros. También  $K'\beta = w$  es una hipótesis consistente si y solo si  $[D(K'\beta)] [D(K'\beta)]' [K'\beta - w] = K'\beta - w$ .

*Nota 2* Se ha intentado demostrar que se pueden obtener los mismos valores  $R_0^2$  y  $R_1^2$  suprimiendo las restric-

ciones  $N'XB = N'Y$  en las expresiones (2.6) y (2.7) y eligiendo una g-inversa especial de  $V$ . En Rao (1978) se demostró que ésto no es posible.

### 2.3.- ENFOQUE DE LA MATRIZ INVERSA PARTICIONADA

Otro enfoque unificado llamado de la matriz inversa particionada fue desarrollado por Rao (1971) para el caso general en que  $V$  y  $X$  pueden ser deficientes en rango.

*Proposición 2.3* Sea

$$\begin{bmatrix} V & X \\ X' & 0 \end{bmatrix}^{-1} = \begin{bmatrix} C_1 & C_2' \\ C_3 & -C_4 \end{bmatrix}$$

para cualquier g-inversa. Entonces, se verifica

- (i) [uso de  $C_2$  y  $C_3$ ]. El estimador lineal insesgado / óptimo de cualquier función identificable  $p'\beta$  es  $p'\hat{\beta}$  donde  $\hat{\beta} = C_2 Y \oplus C_3 Y$  (que puede no ser lo mismo)

La condición de estimabilidad de  $p'\beta$  es

$$P'C_3X = p' \quad \text{y} \quad P'C_2X = p'.$$

- (ii) [uso de  $C_4$ ] la matriz de dispersión de  $\hat{\beta}$  es  $\sigma^2 C_4$  en el sentido de que

$$V(p'\hat{\beta}) = \sigma^2 p'C_4 P, \text{cov}(p'\hat{\beta}, q'\hat{\beta}) = \sigma^2 p'C_4 Q = \sigma^2 q'C_4 P$$

- (iii) [uso de  $C_1$ ] Un estimador insesgado de  $\sigma^2$  es

$$f^{-1}Y'C_1Y \text{ siendo } f = R(V:X) - R(X) \quad (2.8)$$

*Proposición 2.4* Sea  $S'\beta$  el vector de la estimación/lineal insesgada óptima de  $k$  funciones parámetricas estimables  $S'\beta$ ,  $R_\theta^2 = Y'C_1Y$  y  $\delta$  como se ha definido en ///(2.6). Si  $Y \sim N_n(X\beta, \sigma^2 V)$ ,

- (i)  $S'\beta$  y  $R_\theta^2$  se distribuyen independientemente con

$$S'\hat{\beta} \sim N_k(S'\beta, \sigma^2 G), \quad R_\theta^2 \sim \sigma^2 X^2(\delta)$$

siendo  $G = S'C_4 S$ .

- (ii) Sea  $S'\beta = w$  una hipótesis nula. La hipótesis es consistente si y solo si

$$G G^- u = u, \quad u = p'\hat{\beta} - w$$

en cuyo caso el estadígrafo de la dócima

$$F = \frac{u'G^-u}{h} \div \frac{R_\theta^2}{f}, \quad h = P(G)$$

tiene una distribución F central con  $h$  y  $f$  grados de libertad

Puede observarse que un enfoque general a una estimación insesgada de varianza mínima cuando  $X$  y  $V$  son posiblemente deficientes en rango ha sido dado también por Zyskind (1967), a través de ciertas transformaciones lineales del modelo. En la Sección 6 desarrollamos una transformación más detallada de este tipo y estudiamos la estimación de los parámetros a través del concepto de suficiencia lineal.

### 3.- CRITERIOS ALTERNATIVOS DE ESTIMACIÓN

Se han planteado preguntas de si son apropiadas las condiciones de insesgaciedad y varianza mínima en la estimación lineal. Podemos considerar criterios alternativos que pueden tener un atractivo intuitivo. Con este fin consideraremos que  $X$  y  $V$  son de rango completo (el caso general se estudia en la Sección 6). Sea  $P\Delta Q$  la descomposición del valor singular de  $V^{-1/2}X$ , donde  $P$  es una matriz semiortogonal  $n \times m$ . Si  $R$  de orden  $n \times (n-m)$  es el complemento ortogonal de  $P$ , el modelo  $(Y, XB, \sigma^2 V)$  puede transformarse en

$$Y_1 = P'V^{-1/2}Y = \Delta Q'b + P'V^{-1/2}\epsilon \quad (3.1)$$

$$Y_2 = R'V^{-1/2}Y = R'V^{-1/2}\epsilon \quad (3.2)$$

Escribiendo  $\theta = \Delta Q'b$ , el modelo  $(Y, XB, \sigma^2 V)$  es equivalente a los dos modelos incorrelacionados

$$(Y_1, \theta, \sigma^2 I_m) \quad y \quad (Y_2, 0, \sigma^2 I_{n-m}). \quad (3.3)$$

Bajo las hipótesis sobre el modelo lineal  $(Y, XB, \sigma^2 V)$ ,  $Y_2$  no tiene información sobre  $\theta$  ni sobre su esperanza ni sobre su correlación con  $Y_1$ , y es, por consiguiente, redundante para la estimación de  $\theta$ . Podemos formular el criterio siguiente para la estimación de una función paramétrica identificable  $p'\beta$ . Obsérvese que  $p'\beta$  puede escribirse/equivalentemente como una función lineal de  $\theta$ , por ejemplo,  $q'\theta$ .

- (i) El estimador de  $q'\theta (= p'\beta)$  es una función sólo de  $Y_1$ , es decir,  $f(Y_1)$ .

- (ii) El estimador  $f(Y_1)$  es de método consistente (semejante al concepto de consistencia de Fisher), es decir, si  $Y_1$  no tiene error,  $f(Y_1) = q'\theta$ , o sea, el estimador se reduce al verdadero valor.

La condición (ii) implica que  $f(\theta) = q'\theta$ ,  $\forall \theta \in R^m$ , es decir,  $f(Y_1) = q'Y_1$ , que es el estimador mínimo cuadrático

de  $q'\theta = p'\beta$ . De esta forma, llegamos a los estimadores/mínimos cuadráticos sin recurrir a la linealidad, insensibilidad o a la varianza mínima del estimador. [una solución de lo anterior se da en el ejemplo 4, pag. 309 de Rao, (1973)].

#### 4.- PREDICCIÓN Y ESTIMACION SIMULTÁNEAS

##### 4.1.- ESTIMACION SIMULTÁNEA

Consideramos  $k$  modelos lineales

$$y_i = X\beta_i + \varepsilon_i, i = 1, \dots, k \quad (4.1)$$

donde  $y_i$  es un  $n$ -vector,  $X$  una matriz  $n \times m$  dada y  $\beta_i, \varepsilon_i$  son variables aleatorias independientes e identicamente/distribuidas tales que

$$E \begin{pmatrix} \beta_i \\ \varepsilon_i \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \quad D \begin{pmatrix} \beta_i \\ \varepsilon_i \end{pmatrix} = \begin{pmatrix} F & 0 \\ 0 & \sigma^2 V \end{pmatrix} \quad (4.2)$$

Donde  $\beta, F$  y  $\sigma^2$  son desconocidas. Este es un problema de predicción simultánea (estimación) de  $p'\beta$ ,  $i = 1, \dots, k$ , con fines de comparación. Un ejemplo primitivo de dicho problema de estimación (que puede llamarse empírico-Bayes)/es la construcción del índice de selección en genética/sugerido por R.A. Fisher y desarrollado por Fairfield // Smith (1936), Henderson (1950), Panse (1946) y Rao (1952, 1953).

**Proposición 4.1** Sea  $\beta_i^{(l)}$  el estimador mínimo cuadrático de  $\beta_i$  del modelo  $i$ -ésimo y sea  $u = (X'V^{-1}X)^{-1}$ . Si  $\sigma^2, \beta$  y  $F$  son conocidas, el predictor lineal óptimo de  $p'\beta$  es  $p'\beta_i^{(b)}$  (es decir, para el cual el error mínimo / de predicción es mínimo) siendo

$$\beta_i^{(b)} = \beta_i^{(l)} - \sigma^2 U(F + \sigma^2 U)^{-1}(\beta_i^{(l)} - \beta) \quad (4.3)$$

con error mínimo cuadrático de predicción  $p'Q^{(b)}p$ , // siendo:

$$Q^{(b)} = \sigma^2 U - \sigma^4 U(F + \sigma^2 U)^{-1}U \quad (4.4)$$

**Nota 1.**  $\beta_i^{(b)}$  es el estimador de Bayes de  $\beta_i$  bajo / las hipótesis (4.2).

**Nota 2.** El error medio de la dispersión para  $\beta_i^{(b)}$ , /  $i = 1, \dots, k$ , es

$$K^{-1} \sum_1^K E(\beta_i^{(b)} - \beta_i)(\beta_i^{(b)} - \beta_i)' = \sigma^2 U - \sigma^4 U(F + \sigma^2 U)^{-1}U \quad (4.5)$$

que es más pequeño que  $Q^{(l)} = \sigma^2 U$ , el error medio de la / dispersión para  $\beta_i^{(l)}$ ,  $i = 1, \dots, k$ . Cuando  $\sigma^2, \beta$  y  $F$  son desconocidas podemos estimarlas mediante

$$k\beta_* = \sum_1^k \beta_i^{(l)}, \quad (4.6)$$

$$k(n-m)\sigma_*^2 = \sum_1^k (Y_i'V^{-1}Y_i - Y_i'V^{-1}X\beta_i^{(l)}) = W, \quad (4.7)$$

$$(k-1)(F_* + \sigma_*^2 U) = \sum_1^k (\beta_i^{(l)} - \beta_*)(\beta_i^{(l)} - \beta_*)' = B. \quad (4.8)$$

Sustituyendo un múltiplo constante de estos estimadores/por  $\sigma^2, \beta$  y  $F$  en (4.3) obtenemos un estimador empírico / Bayes de  $\beta_i$ .

$$\beta_i^{(e)} = \beta_i^{(l)} - c W U B^{-1} (\beta_i^{(l)} - \beta_*), \quad i = 1, \dots, k. \quad (4.9)$$

Determinamos  $c$  minimizando

$$E \sum_1^k (\beta_i^{(e)} - \beta_i)(\beta_i^{(e)} - \beta_i)' \quad (4.10)$$

bajo la hipótesis de que  $\varepsilon_i$  y  $\beta_i$  tienen distribuciones/ normales multivariantes independientes. La proposición / siguiente se demuestra en Rao (1975)

**Proposición 4.2.** Si  $\beta_i$  y  $\varepsilon_i$  tienen distribuciones normales multivariantes con parámetros como en (4.2),  $B$  y  $W$  como se definen en (4.7) y (4.8) están independientemente distribuidas con

$$W \sim \sigma^2 X^2 (kn - km), \quad B \sim W_m^{(k-1)}(F + \sigma^2 U)$$

y el valor óptimo de  $c$  que minimiza (4.10) es

$$c = (k-m-2)/(kn - km + 2)$$

y con esta elección de  $c$ , el error medio de la dispersión de  $\beta_i^{(e)}$  es

$$Q^{(e)} = \sigma^2 U - \frac{\sigma^4 (n-m)(k-m-2)}{k(n-m)+2} U (F + \sigma^2 U)^{-1} U \quad (4.11)$$

supuesto que  $k = m + 2$

**Nota 1.** Tenemos la desigualdad

$$Q^{(l)} > Q^{(e)} > Q^{(b)} \quad (4.12)$$

**Nota 2.** Puede demostrarse fácilmente que la esperanza de (4.10) para valores fijos de  $\beta_1, \dots, \beta_k$  es más pequeña que  $\sigma^2 U$ . Por consiguiente, los estimadores empírico-Bayes (4.9) son uniformemente mejores que los estima-

dores mfnimo-cuadráticos en el sentido de que tienen un error medio de dispersión más pequeño. Así tenemos el efecto Stein en la estimación simultánea de los parámetros vectoriales  $\beta_1, \dots, \beta_k$ . C.G. Katri expresa la esperanza real de (4.10), para  $\beta_1, \dots, \beta_k$  fijos, en la forma:

$$\sigma^2 U = \frac{\sigma^2 (n-m) (k-m-2)}{k(n-m)+2} E(UU^{-1}) \quad (4.13)$$

*Nota 3.* Efron (1973) y Efron y Morris (1972) obtuvieron resultados de naturaleza semejante.

#### 4.2.- PREDICCIÓN SIMULTÁNEA

Consideremos el modelo  $i$ -ésimo  $y_i = XB_i + \varepsilon_i$ , y una observación futura  $y_i$  con la estructura

$$y_i = x'\beta_i + n_i$$

$$D\begin{pmatrix} \varepsilon_i \\ n_i \end{pmatrix} = \sigma^2 \begin{pmatrix} V & A \\ A' & B \end{pmatrix}, \quad i = 1, \dots, k, \quad (4.14)$$

donde  $B$  es un escalar,  $A$  es un vector y  $\beta_1, \dots, \beta_k$  son parámetros vectoriales fijos. Si  $g_i(y_1, \dots, y_k)$  es un predictor de  $y_i$ , entonces bajo la hipótesis de normalidad multivariante (o simple linealidad de la regresión de  $y_i$  sobre  $y_j$ )

$$\begin{aligned} \sum_{i=1}^k (y_i - g_i)^2 &= E \sum_{i=1}^k [y_i - x'\beta_i - A'V^{-1}(y_i - XB_i)]^2 \\ &+ E \sum_{i=1}^k [p'\beta_i - h(y_1, \dots, y_k)]^2 \end{aligned} \quad (4.15)$$

donde  $h_i = g_i - A'V^{-1}y_i$  y  $p' = x' - A'V^{-1}X$ . Para minimizar la pérdida combinada (4.15) en la predicción/simultánea sólo necesitamos minimizar el segundo término del segundo miembro de (4.15), que es el problema de estimación simultánea de  $p'\beta_i$ ,  $i = 1, \dots, k$ . Usando / la Proposición 4.2 y el resultado (4.13) obtenemos la / siguiente:

*Proposición 4.3.* Si  $\beta_i^{(e)}$  y  $\beta_i^{(\ell)}$  están definidas como en (4.9)

$$E \sum_{i=1}^k (y_i - [x'\beta_i^{(e)} + A'V^{-1}(y_i - XB_i^{(e)})])^2$$

$$\leq E \sum_{i=1}^k (y_i - [x'\beta_i^{(\ell)} + A'V^{-1}(y_i - XB_i^{(\ell)})])^2 \quad (4.16)$$

*Nota.* El resultado (4.16) demuestra que

$$x'\beta_i^{(e)} + A'V^{-1}(y_i - XB_i^{(e)}) \quad (4.17)$$

es un predictor mejor de  $y_i$  que el predictor mfnimo-cuadrático

$$x'\beta_i^{(\ell)} + A'V^{-1}(y_i - XB_i^{(\ell)}) \quad (4.18)$$

bajo la función de pérdida combinada (4.15).

#### 4.3.- ESTIMADOR ACANALADO NO HOMOGENEO

Supongamos que tenemos ahora el modelo

$$Y = XB + \varepsilon, \quad D(\varepsilon) = \sigma^2 V,$$

siendo  $Y$  un  $n$ -vector y  $B$  es un  $m$ -vector. Suponiendo que  $B$  es una variable aleatoria con

$$E(B) = \gamma = g(1, \dots, 1)' \quad D(B) = \sigma^2 k^{-1} I$$

podemos, utilizando (4.3), escribir el estimador de Bayes de  $B$  como

$$\begin{aligned} \beta^{(b)} &= \beta^{(\ell)} - U(U + k^{-1}I)^{-1} (\beta^{(\ell)} - \gamma) \\ &= \gamma + (X'V^{-1}X + kI)^{-1} X'V^{-1}(Y - X\gamma). \end{aligned} \quad (4.19)$$

Con  $\gamma = 0$ , el estimador (4.19) es el estimador acanalado / de Hoerl y Kennard (1970). Puesto que  $\gamma$  y  $k$  son desconocidos, en Rao (1977) se recomienda una versión empírica

$$\hat{\beta}^{(e)} = \underline{1}' \hat{g} + (X'V^{-1}X + \hat{k}I)^{-1} X'V^{-1}(Y - \underline{1}\underline{1}' \hat{g})$$

siendo

$$\begin{aligned} \hat{g} &= \underline{1}' U^{-1} \beta^{(\ell)} : \underline{1}' U^{-1} \underline{1}, \quad \underline{1}' = (1, \dots, 1)' \\ \hat{k} &= a : b \end{aligned}$$

donde  $a$  y  $b$  se determinan a partir de las ecuaciones

$$(n-m+2)a = (Y - X\beta^{(\ell)})' V^{-1} (Y - X\beta^{(\ell)})$$

$$(m-3) \left[ \frac{b}{m-1} (\text{tr } U^{-1} - \frac{1'U^{-2}1}{1'U^{-1}1}) + a \right] = (\beta^{(l)} - \hat{1}'\hat{g})'U^{-1}(\beta^{(l)} - \hat{1}\hat{g}).$$

El estimador  $\beta^{(e)}$  puede no ser mejor que  $\beta^{(l)}$  en función de la función cuadrática de pérdida combinada, pero puede ser un buen competidor para el estimador acanalado / usual.

## 5.- ROBUSTEZ DE LOS PROCEDIMIENTOS DE INFERENCIA

Sea  $Y \sim N(X\beta, \sigma^2 V)$ . Como es bien conocido, la dócima de la razón de verosimilitudes para docimar la hipótesis  $A\beta = 0$  es

$$\lambda(X, V) = \frac{Y' (I - P_{X_0} V^{-1}) Y}{Y' (I - P_{X_0} V^{-1}) Y} \quad (5.1)$$

Es también la dócima uniformemente más potente invariantemente bajo un grupo de transformaciones apropiado. La siguiente proposición debida a Khatri (1980) y Mathew y // Bhimansankaram (1982) de las condiciones bajo las cuales  $\lambda(X, V) = \lambda(X, I)$ .

**Proposición 5.1** Las dócimas para la razón de verosimilitudes para la hipótesis  $A\beta = 0$  bajo los modelos //  $Y \sim N(X\beta, \sigma^2 V)$  e  $Y \sim N(X\beta, \sigma^2 I)$  son las mismas si y solo si se verifican cualquiera de las siguientes condiciones equivalentes.

(i)  $V$  es de la forma

$$V = I + X\Lambda_1 X' + (s-1)ZZ' + X_0\Lambda_3 Z' + Z\Lambda_3' X_0' \quad (5.2)$$

siendo  $\Lambda_1$  y  $\Lambda_3$  matrices arbitrarias,  $s$  un número real positivo arbitrario sometido a  $V$ , que es dependiente en // probabilidad y  $Z'X\Lambda_1 X'Z = (s-1)I_k$  y  $X_0 = X(I - A^{-1}A)$  como se definió en la Proposición 5.1.

(ii)  $(I - P_{X_0})V(I - P_{X_0}) = a(I - P_{X_0})$  para algún  $a > 0$

(iii)  $V^{-1}(I - P_{X_0} V^{-1}) = a(I - P_{X_0})$  para algún  $a > 0$

(iv)  $\begin{bmatrix} I - P_{X_0} \\ L \\ P_X \end{bmatrix} (V - aI) (I - P_{X_0} V^{-1}) = 0$  para algún  $a > 0$ , con  $A = LX$

Finalmente, tenemos otras dos proposiciones demostradas por Mathew y Bhimasankaran (1982).

**Proposición 5.2** Las dócimas de la razón de verosimili-

tudes para la hipótesis  $A\beta = 0$  bajo los modelos //  $Y \sim N(X\beta, \sigma^2 I)$  e  $Y \sim N(U\beta, \sigma^2 I)$  son idénticas si y solo si

$$R(X) = R(U) \text{ y } R(X_0) = R(U_0) \quad (5.3)$$

**Proposición 5.3** Las dócimas de la razón de verosimilitudes para la hipótesis  $A\beta = 0$  bajo los modelos //  $Y \sim N(X\beta, \sigma^2 I)$  e  $Y \sim N(U\beta, \sigma^2 I)$  son idénticas si y solo si se verifican simultáneamente (5.2) y (5.3).

El resultado siguiente, debido a Ghosh y Sinha (1980) (ver también Sinha y Mukhopadhyay (1980)), establece la robustez de las dócimas de la razón de verosimilitudes para desviación de la normalidad.

**Proposición 5.4** Sea  $Y = X\beta + \sigma \epsilon$  con  $\epsilon$  distribuida de acuerdo a la densidad

$$f(\epsilon) = \int_0^\infty e^{-\tau \epsilon \epsilon' / 2} (\tau / 2\pi)^{n/2} dL(\tau).$$

La dócima F basada sobre  $\lambda(X, I)$  para docimar la hipótesis  $A\beta = 0$  es a la vez la dócima de la razón de verosimilitud y la dócima uniformemente más potente invariante.

Recientemente, Sinha y Drygas (1982) generalizaron este resultado de la forma siguiente:

**Proposición 5.5** Sea  $Y = X\beta + \sigma \epsilon$  con  $\epsilon$  distribuida de acuerdo a la función de densidad  $q(\epsilon' \epsilon)$ , siendo  $q$  una función convexa no creciente. La dócima F basada sobre //  $\lambda(X, I)$  es a la vez la dócima de la razón de verosimilitudes y la dócima uniformemente más potente invariante.

Este resultado es semejante a la propiedad de robustez de la dócima  $T^2$  de Hotelling, demostrada por Karia / (1981), y se basa en un teorema de representación debido a Wijsman (1967).

Bajo una distribución de los errores ligeramente más/ general, Sinha y Drygas (1982) establecieron la propiedad siguiente de la estimación lineal insesgada óptima de  $A\beta$ .

**Proposición 5.6** Sea  $Y = X\beta + \sigma \epsilon$  teniendo  $\epsilon$  una distribución esférica simétrica. Para cualquier número real  $c$  y cualquier matriz diagonal no nula  $C$  de orden apropiado.

$$P \{ (GY - A\beta)' C (GY - A\beta) \leq c^2 \} \geq P \{ (LY - A\beta)' C (LY - A\beta) \leq c^2 \}$$

siendo  $GY$  el estimador lineal insesgado óptimo del estimable  $A\beta$  y  $LY$  cualquier estimador insesgado de  $A\beta$ .

## 6.- TRANSFORMACION CANONICA

Consideramos el modelo  $(Y, X\beta, \epsilon)$ ,  $D(\epsilon) = \sigma^2 I$ , donde  $X$  y  $V$  pueden ser deficientes en rango. Sea  $V$  una matriz  $n \times m$  de rango  $V$ . Existe una transformación  $(W:N)$ , siendo  $W$  una matriz  $n \times v$  y  $N$  una matriz  $n \times (n-v)$ , tal que  $N'W = 0$ ,  $W'VW = I_v$ . Bajo tal transformación, el modelo  $(Y, X\beta, \sigma^2 V)$  se transforma en

$$N'Y = Y_3 = N'X\beta + \epsilon_3, \quad D(\epsilon_3) = 0 \quad (6.1)$$

$$W'Y = Y_5 = W'X\beta + \epsilon_5, \quad D(\epsilon_5) = \sigma^2 I \quad (6.2)$$

en el cual la parte (6.1) es singular y la otra (6.2) es regular (no singular). Hacemos una nueva transformación

$$Y_3 = N'X\beta + \epsilon_3, \quad D(\epsilon_3) = 0 \quad (6.3)$$

$$Y_5 - \Lambda Y_3 = Y_4 = (W' - \Lambda N')X\beta + \epsilon_4, \quad D(\epsilon_4) = \sigma^2 I \quad (6.4)$$

siendo  $\Lambda$  tal que  $(W' - \Lambda N')XX'N = 0$ . El modelo (6.3), / (6.4) puede escribirse como

$$Y_3 = X_3\beta + \epsilon_3, \quad D(\epsilon_3) = 0 \quad (6.5)$$

$$Y_4 = X_4\beta + \epsilon_4, \quad D(\epsilon_4) = \sigma^2 I \quad (6.6)$$

siendo  $R(X'_3) \cap R(X'_4) = 0$ , lo que proporciona dos modelos lineales distintos no correlacionados con parámetros diferentes, uno de los cuales es singular. Obsérvese que si  $R(X) \subset R(V)$ , entonces  $X_3 = 0$  e  $Y_3 = 0$  con probabilidad 1, de modo que tenemos sólo la parte regular.

Sea  $\rho(X_3) = x_1$  y  $\rho(X_4) = x_2$ , entonces se pueden hacer nuevas transformaciones sobre cada modelo (de la forma usual, Rao (1973a), p. 189) para obtener la forma canónica de  $(Y, X\beta, \sigma^2 V)$  en cuatro submodelos distintos

$$Y_{-1} = \epsilon_{-1}, \quad D(\epsilon_{-1}) = 0, \quad Y_{-1} \text{ es un } (n-v-x_1)-\text{vector} \quad (6.7)$$

$$Y_1 = X_1\beta + \epsilon_1, \quad D(\epsilon_1) = 0, \quad X_1 \text{ es una matriz } x_1 \times m \quad (6.8)$$

$$Y_2 = X_2\beta + \epsilon_2, \quad D(\epsilon_2) = \sigma^2 I, \quad X_2 \text{ es una matriz } x_2 \times m \quad (6.9)$$

$$Y_0 = \epsilon_0, \quad D(\epsilon_0) = \sigma^2 I, \quad Y_0 \text{ es } (v-x_2)-\text{vector} \quad (6.10)$$

Obsérvese que

$$R(X'_1) \cup R(X'_2) = R(X') \quad y \quad R(X'_1) \cap R(X'_2) = 0$$

de modo que  $\theta_1 = X_1\beta$  y  $\theta_2 = X_2\beta$  son conjuntos distintos de parámetros diferentes que son equivalentes a  $X\beta$ .

Una vez que se conocen las observaciones, podemos hacer una nueva transformación condicional sobre el submodelo (6.8) para obtener

$$(Y'_1)'Y_1 = 0 = Y_{11} = (Y_1)'X_1\beta = X_{11}\beta \text{ con probabilidad 1} \quad (6.12)$$

$$Y'_1 Y_1 = Y_{12} = Y'_1 X_1\beta = X_{12}\beta \text{ con probabilidad 1} \quad (6.13)$$

siendo  $Y'_1$  la matriz ortogonal complementaria de  $Y_1$ .

La forma canónica (6.7) - (6.10) junto con (6.12) y (6.13) expone la información en las observaciones de una forma adecuada para tratar problemas de inferencia.

Por ejemplo, los resultados siguientes se deducen fácilmente.

*Proposición 6.1.*

(i)  $Y_1$  es el estimador lineal insesgado óptimo de  $\theta_1 = X_1\beta$  con matriz de dispersión nula.

(ii)  $Y_2$  es el estimador lineal insesgado óptimo de  $\theta_2 = X_2\beta$  con matriz de dispersión  $\sigma^2 I$ .

(iii)  $Y'_0 Y_0 + (v-x_2)$  es un estimador insesgado de  $\sigma^2$

(iv) La clase completa de estimadores lineales insesgados óptimos de las funciones lineales paramétricas es

$$\left\{ m'_1 Y_1 + m'_2 Y_2 \right\} \quad (6.14)$$

(v) Puesto que  $\theta_1$  es estimable sin ningún error, el problema se reduce a estimar  $\theta_2 = X_2\beta$ . Con este fin, podemos considerar que solamente es apropiado (suficiente)  $Y_2$ , puesto que las otras variables están incorrelacionadas con  $Y_2$  y tienen esperanzas / independientes de  $\theta_2$  (ver Barnard 1963) y Rao // (1965), (1973), p. 308, ejemplo 4). Luego, podemos estimar  $\theta_2$  por medio de  $f(Y_2)$ , función sólo de  $Y_2$ . Bajo una condición semejante a la consistencia de Fisher, esto es,  $f(Y_2)$  se reduce a  $\theta_2$  si el componente de error  $\epsilon_2 = 0$  (es decir, cuando no se cometen errores), hallamos que  $f(\theta_2) = \theta_2, \forall \theta_2 \in R^{\delta-t}$  lo cual implica que  $f(Y_2) = Y_2$ , el estimador lineal insesgado óptimo de  $\theta_2$ .

La forma canónica (6.7)-(6.10) y (6.12) nos permite también discutir de una manera sencilla conceptos tales/ como la suficiencia lineal. Consideraremos las definiciones siguientes que se obtienen del resultado (v) de la / Proposición 6.1

*Definición 6.1* (Forma extendida de la definición de Barnard) Un estadígrafo  $L'Y$  es suficiente linealmente para  $K'\theta$  (conjunto identificable de funciones paramétricas)

si y sólo si la esperanza de cualquier función lineal de  $y$ , incorrelacionada con  $L'y$ , no posee una esperanza de la forma  $\lambda'K\theta$ , siendo  $K'\lambda \neq 0$ .

**Definición 6.2** Un estadígrafo  $M'y$  es suficiente linealmente mfnimo para  $K'\theta$  si y solo si  $M'y$  es suficiente linealmente, según la Definición 6.1, y para todo estadígrafo suficiente linealmente existe una transformación  $C_L$  tal que  $M'y = C_L L'y$ .

Se puede establecer fácilmente la proposición siguiente.

**Proposición 6.2** El estimador lineal insesgado óptimo de  $K'\theta$  es suficiente linealmente mfnimo para  $K'\theta$  en el sentido de la Definición 6.2.

Una definición alternativa de suficiencia es la siguiente.

**Definición 6.3** Un estadígrafo  $L'y$  es suficiente linealmente para  $X\beta$  si y solo si todo estimador lineal admissible (en el sentido de la Sección 4) de toda función lineal paramétrica identificable de  $\beta$  es función lineal de  $L'y$ .

**Definición 6.4** Un estadígrafo  $M'y$  es suficiente linealmente mfnimo para  $X\beta$  si y sólo si  $M'y$  es suficiente linealmente para  $X\beta$ , según la Definición 6.3, y para todo estadígrafo suficiente linealmente  $L'y$  existe una transformación  $C_L$  tal que  $M'y = C_L L'y$ .

La proposición siguiente se establece fácilmente.

**Proposición 6.3** El estadígrafo  $(y_1, y_2)$ , que es el mismo que el estimador lineal insesgado óptimo de  $X\beta$ , es suficiente linealmente mfnimo para  $X\beta$ , según la Definición 6.4, siendo  $y_1, y_2$  las funciones lineales de  $y$ , como se definieron en la representación canónica (6.7) - (6.10).

En trabajos recientes, Drygas (1983) y Muller (1982) utilizaron la definición siguiente de suficiencia lineal debida a Bakasalary y Kala (1978).

**Definición 6.5** Un estadígrafo  $L'y$  es suficiente linealmente para  $X\beta$  si y sólo si el estimador lineal insesgado óptimo de  $X\beta$  es de la forma  $D_L L'y$ .

**Definición 6.6** Un estadígrafo  $M'y$  es suficiente linealmente mfnimo para  $X\beta$  si y sólo si  $M'y$  es suficiente linealmente y  $M'y = C_L L'y$  para cualquier estadígrafo/

suficiente linealmente  $L'y$

**Proposición 6.4** El estadígrafo  $(y_1, y_2)$  es suficiente linealmente mfnimo para  $X\beta$  según la Definición 6.6.

Las condiciones necesarias y suficientes para que  $L'y$  sea suficiente linealmente, según cualquiera de las Definiciones 6.1, 6.3, y 6.5, se discuten en un trabajo de Rao y Sinha que también aclara los resultados obtenidos por Drygas (1983a, 1983b), Muller (1982) y Seely (1978).

## 7.- SELECCION DEL MODELO DE PREDICCION

La selección del modelo se basa, en general, sobre algún criterio de la bondad del ajuste del modelo a los datos observados, tal como el criterio de información de Akaike y el  $C_p$  de Mallows. Un modelo, que ajusta los datos adecuadamente, puede no ser bueno para la predicción más allá del recorrido de los valores observados de los predictores (extrapolación), como en la predicción del crecimiento futuro en base a las observaciones anteriores. Por ejemplo, consideremos el modelo de crecimiento

$$y_t = \beta_0 \psi_0(t) + \dots + \beta_k \psi_k(t) + \epsilon_t, \quad t = 1, \dots, n \quad (7.1)$$

donde  $\psi_0, \dots, \psi_k$  son polinomios ortogonales,  $\epsilon_1, \dots, \epsilon_k$  son errores incorrelacionados y  $V(\epsilon_i) = \sigma^2$ . Un problema importante es la elección del grado del polinomio en (7.1) a usar para predecir una observación futura, por ejemplo, en  $t = n + 1$ . Si usamos el grado máximo posible  $k$ , el predictor óptimo es

$$\hat{y}_{n+1} = b_0 \psi_0(n+1) + \dots + b_k \psi_k(n+1) \quad (7.2)$$

siendo  $b_i$  los estimadores mínimos cuadráticos tales que  $V(b_i) = \sigma^2$ ,  $\text{cov}(b_i, b_j) = 0$ . El error mfnimo cuadrático de predicción es

$$\sigma^2 \left[ 1 + \psi_0^2(n+1) + \dots + \psi_k^2(n+1) \right] \quad (7.3)$$

Si usamos solo  $(p + 1)$  términos en el modelo (7.1), el error mfnimo cuadrático de predicción es

$$\sigma^2 \left[ 1 + \sum_0^p \psi_i^2(n+1) \right] + \left[ \sum_{p+1}^k b_i \psi_i(n+1) \right]^2 \quad (7.4)$$

que puede estimarse

$$\begin{aligned} \hat{\sigma}^2 & \left[ 1 + \sum_0^p \psi_i^2(n+1) \right] + \left[ \sum_{p+1}^k b_i \psi_i(n+1) \right]^2 - \hat{\sigma}^2 \sum_0^p \psi_i^2(n+1) = \\ & \hat{\sigma}^2 \left[ 1 + 2 \sum_0^p \psi_i^2(n+1) \right] + \left[ \sum_{p+1}^k b_i \psi_i(n+1) \right]^2 - \hat{\sigma}^2 \sum_0^p \psi_i^2(n+1) \end{aligned} \quad (7.5)$$

siendo  $\hat{\sigma}^2$  el estimador mfnimo cuadrático de  $\sigma^2$  basado en el modelo completo. A partir de (7.5) hallamos que la elección apropiada de  $p$  es el valor para el cual

$$2 \sum_{i=0}^p \psi_i^2(n+1) + \left[ \sum_{i=p+1}^k b_i \psi_i(n+1) \right] \hat{\sigma}^2 = 0 \quad (7.6)$$

es mfnimo. Puede verse, comparando (7.3) y (7.4), que aún no todos los  $b_{p+1}, \dots, b_k$  son nulos, la elección de un polinomio de grado  $p$ -ésimo (es decir, un modelo erróneo) proporciona la mejor predicción si

$$\left( \sum_{i=p+1}^k b_i \psi_i(n+1) \right)^2 \leq \sigma^2 \sum_{i=p+1}^k \psi_i^2(n+1) \quad (7.7)$$

Desde luego, el mejor modelo erróneo corresponde al valor de  $p$  para el cual

$$\sigma^2 \left[ \sum_{i=0}^p \psi_i^2(n+1) \right] + \left[ \sum_{i=p+1}^k b_i \psi_i(n+1) \right]^2 = 0 \quad (7.8)$$

es mfnimo y el criterio (7.5) es una versión empírica de (7.8).

Un criterio del tipo (7.6) para la elección de variables en el modelo lineal general  $y = X\beta + \varepsilon$  para la predicción en un valor elegido (o valores elegidos) de las variables predictoras puede escribirse de manera semejante. Para expresiones más exactas y para un estudio más detallado el lector puede consultar Rao (1983).

#### 8.- ¿REGRESIÓN DIRECTA O INVERSA?

Supongamos que  $(t, m_1, m_2)$  representa las medidas verdaderas  $t$  y las observadas y repetidas  $m_1$  y  $m_2$  de la presión sanguínea de una persona. Dadas  $m_1$  y  $m_2$  y el modelo

$$\begin{aligned} m_1 &= t + \varepsilon_1 \\ m_2 &= t + \varepsilon_2 \end{aligned} \quad (8.1)$$

donde los errores  $\varepsilon_1, \varepsilon_2$  son tales que  $E(\varepsilon_i) = 0$ ,  $V(\varepsilon_i) = \sigma^2 \varepsilon$ ,  $\text{cov}(\varepsilon_1, \varepsilon_2) = 0$ , ¿cómo especificamos la regla para estimar  $t$  la presión sanguínea verdadera no observada, en una práctica rutinaria (por ejemplo, en una clínica)?

Podemos considerar (8.1) como un modelo lineal que involucra un parámetro desconocido  $t$ , y estimarlo mediante el estimador mfnimo cuadrático,  $\hat{m} = (m_1 + m_2)/2$ , que es una regla bien conocida.

Por otra parte, podemos considerar  $(t, m_1, m_2)$  como tres medidas sobre una persona una de las cuales es no observable e intentar predecir  $t$  por medio de la regresión de  $t$  sobre  $m_1$  y  $m_2$ , calculada a partir de la distribución de  $(t, m_1, m_2)$  en la población de las personas que visitan una clínica. Si  $\tau$  y  $\sigma^2$  son la media y la varianza de  $t$  entre las personas, la regresión de  $t$  sobre  $m_1$  y  $m_2$  es

$$\hat{t} = \tau + \frac{2\sigma^2}{2\sigma^2 + \sigma_\varepsilon^2} \frac{t - \bar{m}}{\bar{m} - \tau} \quad (8.2)$$

Los parámetros desconocidos  $\tau, \sigma_\tau^2$  y  $\sigma_\varepsilon^2$  pueden estimarse a partir de los datos anteriores [ver Rao (1973a) pag. 337], y puede ponerse al dfa cuando se consigan nuevos / datos. Sustituyendo las estimaciones para los valores desconocidos en (8.2), podemos especificar la regla de estimación de  $t$  como

$$\hat{t} = \hat{\tau} + \frac{2\hat{\sigma}^2}{2\hat{\sigma}_t^2 + \hat{\sigma}_\varepsilon^2} \frac{t - \bar{m}}{\bar{m} - \hat{\tau}} \quad (8.3)$$

La expresión (8.3) puede también reconocerse como el estimador empírico-Bayes.

Ignorando errores en  $\hat{\tau}, \sigma_\tau^2$  y  $\sigma_\varepsilon^2$  se halla:

$$\begin{aligned} E(\hat{t} - t)^2 &= \frac{\sigma_\varepsilon^2}{2} \frac{2\sigma^2 t}{2\sigma^2 t + \sigma_\varepsilon^2} \\ &< \frac{\sigma_\varepsilon^2}{2} = E(\bar{m} - t)^2 \end{aligned} \quad (8.4)$$

de modo que el error cuadrático medio total (promediado/sobre todas las futuras predicciones) es más pequeño para  $\hat{t}$  que para  $\bar{m}$ . ¿Podríamos recomendar, a partir de estos datos,  $\hat{t}$  como una regla mejor que  $\bar{m}$ ?

Se ve que  $\hat{t}$  subestima consistentemente  $t$  cuando  $t - \tau$  es grande positivamente y sobreestima cuando  $t - \tau$  es grande negativamente. Realmente los valores grandes y pequeños de  $t$  se estiman pobremente, mientras que los valores intermedios de  $t$  próximos a  $\tau$  (el promedio general) se estiman bien. Quizás, en un caso como el de determinar la presión sanguínea con fines de diagnóstico, no sea deseable subestimar valores grandes y sobreestimar valores pequeños. Sin embargo,  $\hat{t}$  puede ser mejor en otras situaciones donde la función cuadrática de pérdida es significativa.

## ALGUNAS CONSIDERACIONES SOBRE PREDICCIÓN Y REGRESIÓN

## 1. INTRODUCCIÓN

La regresión es la técnica más extensamente estudiada en la literatura estadística. Se menciona en todos los textos de Estadística y se trata de una manera rigurosa en algunos, por ejemplo, Fisher (1925), Malinvaud (1964), Rao, (1965), Searle, (1971) y Theil, (1971). Hay tratados completos sobre el tema tales como Draper y Smith (1966), Plackett (1960) y Williams (1959). Resúmenes bastante buenos acerca de los métodos de regresión se pueden ver en el trabajo de Cox, Ehrenberg y Nelder (1968) y en los artículos de Draper y Smith (1969), Williams (1969) y Wold y Lyttkens (1969).

Sin embargo, parece que hay algunos aspectos de la técnica de regresión, que no se han entendido apropiadamente, o que no parecen haber sido expuestos adecuadamente en la literatura estadística. En consecuencia ha habido algunos abusos de la técnica. Un ejemplo clásico es el uso del método de regresión, en lo que se conoce como el problema de la desagregación (Sección 2), que ha conducido a resultados sin sentido. El fenómeno de la multicolinearidad aunque ampliamente estudiado, todavía parece que necesita alguna clarificación. Las definiciones existentes son vagas y las consecuencias de la multicolinearidad no se han examinado completamente (Sección 3). También existe el problema de la regresión inversa (Sección 4) acerca del cual hay una considerable controversia. (Ver comentarios de Lindley en el estudio de Cox, Ehrenberg, Nelder (1968)). El uso de variables no apropiadas en la predicción continua, a pesar de la cautela que aconsejan los excelentes ejemplos de regresión espuria de Yule (1926) y de Box (1966).

También existe alguna controversia acerca del uso de las dos líneas de regresión y de la paradoja de Ehrenberg que es como sigue: Sea  $h = \alpha_1 + \beta_1 w$  la regresión de altura sobre peso y  $w = \alpha_2 + \beta_2 h$  la de peso sobre altura. Consideremos un individuo con la combinación  $(h_0, w_0)$  de altura y peso, que cae dentro de la región  $\{h > \alpha_1 + \beta_1 w, w > \alpha_2 + \beta_2 h\}$  en el plano  $(h, w)$ . Un médico, utilizando la regresión de peso sobre altura puede considerar que el individuo observado es demasiado pesado para su estatura. Pero a juzgar por la regresión de altura sobre peso es, desde luego, demasiado alto para su peso. Encarándose con tal paradoja, Ehrenberg (1963-1968), planteó el caso para establecer una sola relación semejante a una

ley entre altura y peso que es independiente de sexo, raza, etc. Si existe dicha relación o no, habrá muchos individuos paradójicos (Ehrenberg) en cualquier población, pero la pregunta de si una característica está en la proporción adecuada a la otra puede surgir solamente en un caso patológico. La regresión a usar en tal caso, dependerá quizás del juicio del médico basándose en el examen de un determinado individuo?

El objeto de la presente conferencia es examinar algunos aspectos (no todos) del problema de la regresión, que no lo han sido adecuadamente hasta ahora.

**Notación.** En el problema de la regresión es usual llamar dependiente a la variable ( $y$ ) bajo predicción e independiente a las usadas para la predicción ( $x_1, x_2, \dots, x_m$ ). Los psicólogos usan la palabra *variable criterio* para  $y$ , y *variables predictoras* para  $x_1, x_2, \dots, x_m$ . Los economistas nombran a  $y$  *variable dependiente* y a  $x_1, x_2, \dots$  *variables explicativas*. Nosotros adoptaremos aquí la terminología de los economistas.

Designaremos  $n$  observaciones incorrelacionadas sobre  $y$  mediante el vector  $\mathbf{Y}$ , y a la matriz  $n \times m$ , de las variables explicativas centrada sobre las medias observadas, mediante  $\mathbf{X}$ . El modelo de la regresión lineal será de la forma

$$\mathbf{Y} = \alpha \mathbf{u} + \mathbf{X}\beta + \mathbf{e} \quad (1.1)$$

donde  $\mathbf{u}' = (1, \dots, 1)$ ,  $\alpha$  es una constante y  $\beta$  es el vector de los parámetros de la regresión. A la matriz  $(n - 1)^{-1} \mathbf{X}' \mathbf{X}$  se le llama matriz de dispersión o matriz  $\mathbf{D}$  de las variables explicativas. Dividiendo el elemento  $(i, j)$  de la matriz  $\mathbf{D}$  por la raíz cuadrada del producto de los elementos diagonales  $i$ -ésimo y  $j$ -ésimo, obtenemos la matriz de correlación o matriz  $\mathbf{C}$ .

El estimador mínimo cuadrado de  $\hat{\alpha}$  es  $\hat{\alpha} = \mathbf{n}^{-1} \mathbf{Y}' \mathbf{u}$  y la ecuación normal para estimar  $\beta$  es

$$(\mathbf{X}' \mathbf{X}) \beta = \mathbf{X}' \mathbf{Y} \quad (1.2)$$

lo que da

$$\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y} \quad (1.3)$$

como el estimador de  $\beta$ , donde  $(\mathbf{X}' \mathbf{X})^{-1}$  es una g-inversa de  $\mathbf{X}' \mathbf{X}$  (Rao (1965), Rao y Mitra (1971)). Un estimador insesgado de  $\sigma^2$ , la varianza residual de la regresión,

es

$$(n-r) [Y'Y - \hat{a}Y'u - \hat{\beta}'X'Y] \quad (1.4)$$

siendo  $r$  el rango de  $X$ . Cuando  $r < m$ , no todas las funciones lineales de  $\beta$  son estimables. El estimador de  $p'\beta$  donde  $p$  pertenece a la variedad lineal generada por las columnas de  $X'$  (que es la condición de estimabilidad) es  $p'\hat{\beta}$  y la varianza del estimador es

$$\sigma^2 p'(X'X) - p \quad (1.5)$$

El valor predicho de  $y$  en los valores de las variables explicativas, expresadas como desviaciones de las medias observadas en los datos anteriores,  $p' = (\bar{x}_1 - / - x_1, \dots, \bar{x}_m - x_m)$ , es

$$\hat{a} + p'\hat{\beta} \quad (1.6)$$

con la varianza de predicción

$$\sigma^2 (1 + \frac{1}{n} + p'(X'X) - p) \quad (1.7)$$

En nuestro estudio sólo nos preocupará el término  $p'(X'X) - p$  en (1.7) que depende de  $X'X$ . Para una exposición más detallada sobre la teoría de los mínimos cuadrados ver Rao (1965, 1973).

## 2. DESAGREGACION

Algunas veces la técnica de regresión se usa para asignar un total a diferentes encabezamientos. Por ejemplo, consideremos una encuesta donde a partir de cada unidad familiar seleccionada obtenemos el total de consumo de arroz ( $t$ ) y la composición de la familia en términos de hombres ( $m$ ) y mujeres ( $f$ ), niños ( $c_1$ ) y niñas ( $c_2$ ). Conociendo los valores de ( $t, m, f, c_1, c_2$ ) de una muestra de  $n$  unidades familiares, ¿podemos estimar el consumo de arroz per cápita de cada una de las cuatro categorías de los miembros de una unidad familiar de la población bajo estudio? Si  $\beta_1, \beta_2, \beta_3$  y  $\beta_4$  representan los parámetros del consumo medio de un hombre, una mujer, un niño y una niña respectivamente, se piensa que pueden estimarse mediante los coeficientes de regresión parcial de  $t$  sobre  $m, f, c_1, c_2$  respectivamente, es decir, ajustando una ecuación de regresión de la forma

$$t = \beta_1 m + \beta_2 f + \beta_3 c_1 + \beta_4 c_2 \quad (2.1)$$

a los  $n$  conjuntos de observaciones sobre ( $t, m, f, c_1$  y

$c_2$ ). En el Indian Statistical Institute, el intento de ajustar una ecuación de regresión del tipo (2.1), dió como resultado una estimación negativa de  $\beta_2$  (el consumo medio para una mujer). Por otra parte, en un estudio semejante de desagregación del gasto total de una granja en toros, vacas y terneros, se obtuvo una estimación negativa de la manutención de un ternero. Desgraciadamente, los resultados del último estudio fueron publicados justificando la estimación negativa como resultante de un error típico grande de los coeficientes de regresión estimados.

De vez en cuando, se publican estudios semejantes, sin un examen crítico de los coeficientes calculados. Parece que estimaciones anómalas, tales como las obtenidas en las encuestas sobre unidades familiares y granjas, pueden surgir debido a la violación de un hipótesis importante en el modelo de regresión, o sea, los mismos parámetros de regresión son aplicables a todas las combinaciones de las variables explicativas. Por ejemplo, consideremos hogares con composiciones familiares ( $m=1, f=0, c_1=0, c_2=0$ ) y ( $m=1, f=1, c_1=2, c_2=1$ ). El modelo lineal más corriente usa las funciones de regresión (esperanzas).

$$\begin{aligned} &\beta_1 \\ &\beta_1 + \beta_2 + 2\beta_3 + \beta_4 \end{aligned}$$

para los dos casos, con los mismos parámetros  $\beta_1, \beta_2, \beta_3$  y  $\beta_4$ . Esto puede no ser verdad; el consumo medio de arroz para un hombre (parámetro de regresión  $\beta_1$ ) puede no ser el mismo para unidades familiares con uno o más miembros. El ejemplo siguiente debido a S.J. Poti muestra como pueden obtenerse estimaciones completamente sin sentido en casos extremos (que puede describirse como el efecto Poti), donde no se satisface la hipótesis fundamental del modelo de regresión, mencionada anteriormente.

Consideremos unidades familiares con un sólo hombre (soltero) o con una pareja (marido y mujer) y sea  $t$  el número de millas por mes recorridas por el automóvil familiar. A partir de una encuesta de unidades familiares que proporciona información sobre la variable  $t$  y la sobrecomposición de las familias, se desea estimar la utilización relativa (en función del promedio de millas recorridas) de los automóviles por los hombres y las mujeres de la población, suponiendo que el automóvil es utilizado por una sola persona cada vez. Si representamos los parámetros correspondientes por  $\beta_1$  y  $\beta_2$ , la ecuación de regresión es de la forma

$$t = \beta_1 m + \beta_2 f$$

donde el vector ( $m, f$ ) puede tomar sólo dos valores posibles (1,0) y (1,1). Si en la muestra hay  $n_1$  unidades familiares del tipo (1,0), con un millaje medio  $t_1$ , y  $n_2$  unidades de la clase (1,1), con un millaje medio //  $t_2$ , se ve fácilmente que las estimaciones mínima-cuadráticas de  $\beta_1$  y  $\beta_2$  son

$$\beta_1 = t_1 \quad \beta_2 = t_2 - t_1$$

Supongamos que la sociedad de la que se ha extraído la muestra es tal que una pareja no puede usar el automóvil tanto como un hombre soltero. En tal caso,  $t_1 > t_2$ , / que da una estimación negativa para  $\beta_2$ , el promedio de millas recorridas por las mujeres. Por otra parte, si / la sociedad consta de dos clases de hombres, una tan pobre que no pueden afrontar el matrimonio o usar el automóvil libremente, y otra, lo suficientemente rica (Maharaja) como para tener esposa y también usar el automóvil libremente,  $t_1 << t_2$ , en cuyo caso  $\beta_1$  sería muy pequeño comparado con  $\beta_2$ , lo que indica que las mujeres // (esposas) usan el automóvil la mayor parte del tiempo.

En estos casos extremos, las estimaciones carecen de sentido cuando los parámetros de la regresión dependen / de la composición de la familia. Esto ocurre en muchas/ situaciones, el efecto Poti existe con mayor o menor / extensión, y la desagregación a través de la técnica de/ la regresión puede conducir a resultados no razonables.

Por ejemplo, en los datos de la encuesta sobre las / granjas es bastante posible que un granjero, que posea / solamente dos bueyes, con fines de labranza, pueda alimentarlos bien, mientras que un granjero que mantiene, / además de los dos bueyes, algunas vacas y terneros, pude de que no gaste una suma mucho más grande en alimentar a todos los animales. En tal caso, podría esperarse sub/ estimaciones, o incluso estimaciones negativas, para la/ manutención de una vaca y un ternero, al ajustar una e/ cuación de regresión a los datos obtenidos, a partir de/ una muestra de todos los granjeros, algunos de los cuales posee sólo dos bueyes y otros tienen además algunas/ vacas y terneros.

En resumen, el uso de los métodos de regresión, con/ fines de desagregación, generalmente conducen a resultados sin sentido o erróneos, a causa de la existencia del efecto Poti. Parece que no hay ningún método satisfactorio para el problema, excepto a través de una contabilidad separada para las diferentes categorías dentro de/

una unidad.

### 3. MULTICOLINEARIDAD

Se ha escrito mucho sobre el tema de la multicolinealidad, que puede encontrarse resumido en el artículo de / Farrar y Glauber (1967). La multicolinealidad se describe de manera general como una fuerte interdependencia de/ las variables explicativas en un modelo de regresión, que puede afectar a la precisión de los estimadores y causar/ alguna dificultad en la interpretación de los coeficientes individuales de regresión, y en el uso de la función/ estimada de regresión con fines de predicción. La interdependencia se busca para ser medida y juzgada por medio/ de la insignificancia del determinante de la matriz D o C de las variables explicativas. Farrar y Glauber (1967) / estudian con detalle métodos para detectar la existencia, medir la extensión y señalar con precisión la ubicación y las causas de la multicolinealidad dentro de un conjunto/ de variables independientes. Su análisis es verdaderamente esclarecedor, pero mucho de su exposición conduce / en parte a conclusiones erróneas, debido a una definición insatisfactoria de la multicolinealidad, que no está directamente relacionada con la precisión de la predicción/ en el dominio pertinente de las variables explicativas, y en parte, como ha indicado Krishna Kumar (1973), a disposiciones no válidas dentro de las décimas de significación, basadas en hipótesis de normalidad. Aquí, examinaremos algunos aspectos de la multicolinealidad, que no // han sido examinadas adecuadamente con anterioridad.

Consideremos una variable dependiente  $e_0$  y dos variables explicativas  $e_1$  y  $e_2$ , con las matrices D y C

D-matrix			C-matrix		
1	5	2	1.0	.5	.4
5	100	.	.5	1.0	.
2	.	25	.4	.	1.0 (3.1)

Puesto que  $e_1$  y  $e_2$  no están correlacionadas, se podría considerar esta situación como ideal para aplicar la técnica de regresión para la predicción. Sin embargo, en lugar de  $e_1$  y  $e_2$  se podrían considerar dos variables explicativas equivalentes algebraicamente, y quizás igualmente significativas,  $x_1 = e_1 + e_2$  y  $x_2 = e_1$ . Las matrices D y C de  $e_0$ ,  $x_1$  y  $x_2$  son

D-matrix			C-matrix		
1	7	5	1.000	.626	.500
7	125	100	.626	1.000	.894
5	100	100	.500	.894	1.000 (3.2)

A partir del alto valor de la correlación entre  $x_1$  y  $x_2$ , 0.894, podría inferirse algún grado de multicolinearidad entre las variables explicativas. Puesto que  $(x_1, x_2)$  es equivalente a  $(e_1, e_2)$  la situación es paradógica, excepto que se defina la multicolinearidad como una alta interdependencia, para una elección particular de las variables explicativas, y no como una propiedad del sistema de variables explicativas (Tenemos, pues, el problema adicional de decidir sobre el conjunto apropiado de variables explicativas para el examen de la multicolinearidad).

La situación no es rara en Econometría, donde algunas variables explicativas forman parte de otras. Podría parecer que la multicolinearidad, en el sentido de la insignificancia del determinante, podría introducirse o evitarse volviendo a definir las variables económicas de una forma igualmente significativa. (Así, si  $e_1$  y  $e_2$  representan la producción y la importación neta,  $e_1$  y  $c = e_1 + e_2$  representan la producción y el consumo, que son magnitudes económicas igualmente significativas y equivalentes). Es indiferente que se elijan  $e_1$  y  $e_2$  o  $e_1$  y  $c$  como variables explicativas. En realidad, cualquier variable vectorial  $X$  con algún grado de dependencia entre sus componentes (correlaciones) puede transformarse en una variable vectorial equivalente  $U$  con componentes incorrelacionados. Puesto que desde el punto de vista de la técnica de regresión  $X$  y  $U$  contienen la misma información, no se observa la distinción entre alta dependencia entre los componentes de uno y la ausencia de asociación en el otro. Esto muestra la necesidad de una formulación más cuidadosa de la multicolinearidad en un problema dado, quizás considerando el uso proyectado de una ecuación ajustada de regresión. Antes de intentar hacer esto, usaremos las matrices  $D$  en (3.1) y (3.2) para explicar algunos aspectos de los coeficientes de regresión, que merece la pena observar.

A menudo se encuentra que, cuando se utilizan varias variables explicativas, algunos de los coeficientes de regresión resultan negativos, aunque las variables correspondientes están correlacionadas individualmente, de una forma altamente positiva, con la variable dependiente. La situación es bastante corriente cuando las correlaciones entre las variables explicativas son altas, como indicó Shourie (1972) en su estudio crítico sobre las funciones de regresión presentadas en el estudio UNCTAD (1968). Por ejemplo, las siguientes ecuaciones de regresión, establecidas por Faaland y Dahl para estimar el valor del consumo agrícola ( $P_A$ ) en función del PNB y la renta per cápita (PNB/N):

$$P_A = 34.322 + 0.402(GDP), \quad \bar{R}^2 = 0.94 \quad (3.3)$$

$$(0.036) \quad (0.397) \quad \bar{R}^2 = 0.96 \quad (3.4)$$

$$= -8.001 + 5.698(GDP/N), \quad \bar{R}^2 = 0.99 \quad (3.5)$$

$$= -131.716 - 1.226(GDP) + (0.239)$$

$$+ 22.816(GDP/N), \quad (3.345)$$

El coeficiente de correlación entre (PNB) y (PNB/N) era 0.9985. Los coeficientes de regresión en cada caso son grandes comparados con los errores típicos (dados entre paréntesis), que indican un grado razonable de precisión en su estimación. La correlación múltiple ( $R$ ) es alta en cada caso, lo cual indica un alto grado de precisión en la predicción. Sin embargo, los autores indican que a priori la correlación parcial negativa de  $P_A$  con respecto al PNB en la ecuación (3.5) parece imposible. Puede haber fuertes razones económicas para las reacciones de los autores al coeficiente negativo. Pero una ecuación de regresión múltiple tal como (3.5) no puede rechazarse sólamente sobre la base de que un cierto coeficiente sea negativo, aunque la variable considerada tenga una alta correlación positiva con la variable dependiente. Desde luego, puede hacerse algún reparo a tal situación. Este resultado, algo inaceptable, puede ser consecuencia de una especificación errónea del modelo lineal, o de una multitud de otras razones, incluso errores de transcripción. (En muchos ejemplos, errores de transcripción de un sólo valor pueden cambiar drásticamente los valores de los coeficientes de regresión). Sin embargo, un coeficiente negativo puede aparecer de una forma natural a causa de una elección particular de las variables explicativas en un sistema dado.

Consideremos, por ejemplo, el caso hipotético de  $e_0$ ,  $e_1$  y  $e_2$  con la matriz  $D$ , que aparece en (3.1). Las variables  $e_1$  y  $e_2$  pueden ser ciertas magnitudes económicas, en cuyo caso las variables  $x_1 = e_1$  y  $x_2 = e_1 + e_2$  pueden ser también magnitudes económicas significativas. La regresión de  $e_0$  sobre diferentes conjuntos de variables explicativas son como sigue (prescindiendo de una constante):

$$e_0 = .05e_1 + 0.8e_2 \quad (3.6)$$

$$= .05e_1$$

$$= .08e_2$$

$$= -.03x_1 + .08x_2 \quad (3.7)$$

$$= .05x_1$$

$$= .028x_2$$

Las ecuaciones (3.6) y (3.) son las mismas y el coeficiente negativo de  $x_1$  en la ecuación de regresión múltiple (3.7) aparece de forma natural, aunque la correlación entre  $e_0$  y  $x_1$  es 0.05. A primera vista, esto puede parecer contrario a lo esperado, pero un estudio más detallado de la elección de las variables explicativas puede resolver la situación. En el ejemplo anterior, la función de regresión basada sobre  $x_1$  y  $x_2$  tiene la misma validez que la basada en  $e_1$  y  $e_2$  y el uso de  $x_1$ ,  $x_2$  o  $e_1$ ,  $e_2$  como instrumentos conducen a los mismos resultados.

Debe destacarse que la presencia de coeficientes negativos (en contra de lo esperado) no es necesario que se deban a la multicolinealidad o altas correlaciones entre las variables explicativas. Consideremos la siguiente matriz C de las tres variables y,  $x_1$ ,  $x_2$ :

$$\begin{array}{ccc} 1.0 & 0.7 & 0.3 \\ 0.7 & 1.0 & 0.5 \\ 0.3 & 0.5 & 1.0 \end{array} \quad (3.8)$$

donde las variables explicativas  $x_1$  y  $x_2$  no indican signos de multicolinealidad. Las ecuaciones de regresión de y sobre diferentes combinaciones de las variables explicativas son (prescindiendo de una constante):

$$\begin{aligned} y &= 0.7x_1 \\ &= 0.3x_2 \\ &= 0.733x_1 - 0.067x_2 \end{aligned} \quad (3.9)$$

El coeficiente negativo de  $x_2$  no disminuye su valor como una variable explicativa. No hay duda, uno debe ser cuidadoso al usar  $x_1$  y  $x_2$  como instrumentos, no sólo en una situación tal, sino en general, incluso cuando los coeficientes parecen plausibles. La naturaleza de las variables económicas y el dominio de validez de la función de predicción deben examinarse, quizás, usando el propio conocimiento del mecanismo económico subyacente, que controla y, dados valores específicos de  $x_1$  y  $x_2$ . Si se necesita la predicción para valores de  $x_1$  y  $x_2$ , fuera de un cierto dominio, puede haber factores, no observados en el pasado, que pueden jugar un papel importante.

*¿Qué es la multicolinealidad?* En sentido estricto implica la existencia de relaciones lineales entre las variables explicativas en los datos observados con anterioridad. En términos algebraicos,  $|X'X| = 0$  siendo X la matriz de las variables explicativas observadas, expresadas como desviaciones de las medias. Cuando esto ocurre,

el rango de X es menor que el número de variables explicativas y no puede estimarse la función de regresión completa, ya que no todos los coeficientes de regresión individual (es decir, las influencias separadas de las variables explicativas) son identificables. Sin embargo ciertas combinaciones de los coeficientes de regresión son estimables, lo que implica que la predicción es posible sólo en ciertas direcciones de las variables explicativas. El método de estimación cuando X es definiente en rango puede estudiarse en Rao (1965, 1973).

En la práctica, sin embargo, debido a errores de observación en las variables explicativas, puede no ser posible la identificación de las relaciones lineales, si las hay. Escribamos la matriz de las variables explicativas observadas como  $X + W$ , donde X designa la matriz de los valores verdaderos desconocidos y W la matriz de los errores. Podría suceder que  $(X + W)'(X + W) \neq 0$ , aunque  $|X'X| = 0$ , lo que no nos indica si es un caso de variables explicativas sin errores y sin relaciones lineales (es decir,  $|X'X| \neq 0$  y  $W = 0$ ), o de variables explicativas con relaciones lineales, pero observadas con error (es decir,  $|X'X| = 0$  y  $(X + W)'(X + W) \neq 0$ ). Examinaremos ambos casos y estudiaremos sus consecuencias en la estimación de los parámetros.

Caso 1:  $|X'X| \neq 0$

No hay multicolinealidad en sentido estricto. De acuerdo con la teoría, todos los coeficientes de regresión son estimables insegadamente y la predicción es posible en todas las direcciones de las variables explicativas. Consideremos una determinada dirección k de las variables explicativas, siendo k un vector tal que  $k'k = 1$ . Si designamos por  $\hat{\beta}$  al estimador mínimo-cuadrático de  $\beta$ , el valor predicho de la variable dependiente en la dirección k es (prescindiendo de una constante) proporcional a  $k'\hat{\beta}$  con varianza

$$\sigma^2 k'(X'Y)^{-1}k \quad (3.10)$$

siendo  $\sigma^2$  la varianza residual. La componente importante de la expresión (3.10) es  $(X'X)^{-1}$ , y la precisión de la predicción en cualquier dirección dada depende de los elementos de la matriz inversa. Sin embargo, para cualquier  $X'X$  dado, la precisión de la predicción depende de la dirección k. Designemos por K el conjunto de direcciones en las cuales se proyecta la predicción. La eficiencia mínima de predicción (o varianza máxima de los vectores en K) será

$$r(X, K) = \max_{k \in K} k'(X'X)^{-1}k \quad (3.11)$$

que se puede considerar como la pérdida asociada con un sistema dado  $X$  de variables explicativas con un determinado  $K$ .

Si  $K$  es el conjunto de todas las direcciones en el espacio euclíadiano  $E^m$  de  $m$  dimensiones,

$$r(X, E^m) = 1/\lambda_m \quad (3.12)$$

siendo  $\lambda_m$  el menor valor propio de  $X'X$  y el vector propio asociado con  $\lambda_m$ . En general, cuando  $X'X$  es casi cero,  $\lambda_m$  es pequeño, y, en consecuencia, la pérdida, como aparece en (3.12), es grande -situación que puede considerarse como existencia de algún grado de multicolinealidad. En la práctica, sin embargo, se puede no estar interesado en la predicción en todas las direcciones  $y$ , por consiguiente, el valor de (3.12) puede no ser pertinente. Representemos todos los valores propios de  $X'X$  por:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$$

y los correspondientes vectores propios por

$$k_1, \dots, k_m$$

Si nuestro interés se limita a las direcciones situadas en la variedad lineal generada por  $k_1, \dots, k_m$ , el valor de (3.11) es  $\lambda_q^{-1}$ , que puede ser pequeño, aunque  $\lambda_m^{-1}$  sea grande. De esta forma, la proximidad de  $|X'X|$  (que es igual al producto  $\lambda_1 \lambda_2 \dots \lambda_m$ ) a cero, debido a que  $\lambda_m$  es pequeño, no tiene consecuencias para nuestro propósito. Naturalmente, se pueden poner ejemplos  $x_1$  y  $x_2$  de  $X$  tales que

$$\lambda_q(x_1'x_1) > \lambda_q(x_2'x_2) \quad x_2$$

mientras que  $\lambda_m(x_1'x_1) < \lambda_m(x_2'x_2)$  y/o  $|x_1'x_1| < |x_2'x_2|$ , en cuyo caso  $x_1$  es la mejor elección de las variables explicativas, aunque con el criterio de multicolinealidad, basado en el menor valor propio,  $x_2$  puede considerarse mejor.

Por lo tanto, se puede definir (3.11) como el grado de multicolinealidad en  $X$ , teniendo en cuenta el conjunto de direcciones de interés. Sin embargo, debe tenerse presente que, si los datos exhiben algún grado de multicolinealidad, esto se refleja en los errores típicos, que son grandes (no hay forma de evitarlo, excepto mediante el uso de información suplementaria). En tal caso, la amplitud del intervalo de confianza será grande y, en consecuencia, se reduce la utilidad del valor predicho

Consideremos dos variables explicativas  $e_1$  y  $e_2$  con la matriz  $C$ :

$$\begin{array}{cc} 1.00 & 0.99 \\ 0.99 & 1.00 \end{array} \quad (3.13)$$

El determinante de la matriz de correlación es 0.0199, que puede considerarse pequeño, lo que indica una alta multicolinealidad de las variables explicativas. Los valores propios de (3.13) son 1.99 y 0.01, de modo que la varianza del valor predicho es proporcional a 0.503 en la dirección (1:1) para variables explicativas tipificadas y 100 en la dirección (1 : -1). Por consiguiente, si estamos interesados en la dirección (1 : 1), el gran valor de la varianza para la otra dirección no es pertinente. Asimismo, si antes las variables explicativas no mostraron correlación, lo que indica una colinearidad perfecta, la varianza de la predicción en la dirección (1 : 1) sería proporcional a 1.00, que es más grande que para el caso de multicolinealidad (3.13).

$$\text{Caso 2: } |X'X| = 0 \text{ y } |U'U| \neq 0, \quad U = X + W$$

Hay, naturalmente, multicolinealidad en sentido estricto, pero está oculta por la matriz de error desconocida  $W$ , y el caso 2 no se distingue del caso 1, a menos que se disponga de información adicional. Usando el método de los mínimos cuadrados, los parámetros de la regresión se estimarían mediante

$$\hat{\beta} = (U'U)^{-1}U'Y \quad (3.14)$$

con la matriz  $D$  asociada:

$$\sigma^2(U'U)^{-1} \quad (3.15)$$

El valor predicho en cualquier dirección  $k$  es  $k'\hat{\beta}$  con varianza  $\sigma^2 k'(U'U)^{-1}k$ .

Sea  $U'U - X'X = G$  definida positiva\* y  $r$  el rango de  $X$ . Existe una matriz  $P$  tal que

$$X'X + P\Lambda P', \quad G = PP' \quad (3.16)$$

siendo  $\Lambda$  una matriz diagonal con elementos diagonales  $\lambda_1 \geq \dots \geq \lambda_r \neq 0, 0, \dots, 0$ . Si  $P_j$  es la columna  $i$ -ésima de  $P$ , se ve fácilmente que  $P_1, \dots, P_r$  abarca a  $X'X$  (o  $X'$ ) y ninguna combinación de  $P_{r+1}, \dots, P_m$  pertenece a la variedad lineal de  $X'X$ .

En función de las matrices  $P$  y  $\Lambda$ :

$$(U'U)^{-1} = (P^{-1})'(\Lambda + I)^{-1}P^{-1} \quad (3.17)$$

El valor predicho en la dirección  $P_i$  es  $P_i^{\hat{\beta}}$ , es- / tanto  $\hat{\beta}$  dado por (3.14). Luego :

$$E(P_i^{\hat{\beta}}) = P_i(P^{-1})'(\Lambda + I)^{-1}P^{-1}(X'X + W'X)$$

$$= \frac{\lambda_i}{1+\lambda_i} P_i^{\beta} + \frac{1}{1+\lambda_i} Q_i^{\top} W' X \beta, \\ i=1, \dots, r \quad (3.18)$$

$$= Q_i^{\top} W' X \beta, \quad i > r. \quad (3.19)$$

siendo  $Q_i$  la columna  $i$ -ésima de  $(P^{-1})'$ . Para  $i = 1, \dots, r$ :

$$E(P_i^{\hat{\beta}}) - P_i^{\beta} = \frac{1}{1+\lambda_i} P_i^{\beta} + Q_i^{\top} W' X \beta \quad (3.20)$$

que es del orden  $(1 + \lambda_i)^{-1}$ , ignorando el error en // (3.20). Si los elementos de  $W'W$  son pequeños en magnitud comparados con los de  $X'X$ , las  $\lambda_i$  son probablemente // grandes y el sesgo en (3.20) puede ser pequeño, supuesto que la dirección de las variables explicativas está en la variedad lineal de  $X'X$ . Pero para  $i > r$ , cuando  $P$  no // pertenece a la variedad lineal de  $X'X$ :

$$E(P_i^{\hat{\beta}}) = Q_i^{\top} W' X \beta \quad (3.21)$$

que no parece tener relación con  $P_i^{\hat{\beta}}$ , el parámetro/ bajo estimación, y cuya magnitud depende sólo de los e-// rrores en  $X$ . De aquí, cualquier pretensión de que  $P_i^{\hat{\beta}}$  / sea un estimador de  $P_i^{\beta}$  para  $i > r$  no tiene sentido.

Examinemos la varianza atribuida de  $P_i^{\hat{\beta}}$ , prescin- diendo del multiplicador  $\sigma^2$

$$P_i^{\beta}(U'U)^{-1}P_i = \frac{1}{1+\lambda_i}, \quad i=1, \dots, r \quad (3.22)$$

$$= 1, \quad i > r. \quad (3.23)$$

De este modo, si las  $\lambda_i$  son grandes, las funciones  $P_i^{\hat{\beta}}$  para  $i = 1, \dots, r$ , se estiman con precisión razonable. Las varianzas asociadas con los estimadores propuestos de  $P_i^{\hat{\beta}}$  para  $i > r$ , son de magnitud relativamente grande.

\* Si  $X$  y  $W$  están incorrelacionadas,  $E(U'U - X'X)$  es definida no negativa, pero  $U'U - X'X$  puede no serlo para todos los valores de  $W$ .

En resumen, si las variables explicativas se han ob- / servado sin error, podríamos haber descubierto que las // funciones lineales del tipo  $P_i^{\hat{\beta}}$ , para  $i > r$ , no son estimables. No reconocer los errores, cuando existen, // podría conducirnos a afirmar que  $P_i^{\hat{\beta}}$ , para  $i > r$ , es también estimable y que su estimación y varianza son

$$P_i^{\hat{\beta}} \quad \sigma^2 P_i^{\beta} (U'U)^{-1} P_i \quad (3.24)$$

respectivamente. Si el rango de  $X$  es  $r$ , el estimador /  $P_i^{\hat{\beta}}$ , para  $i > r$ , estará en la naturaleza del puro / ruido, sujeto a una varianza grande, y no guardará rela- / ción con el parámetro  $P_i^{\beta}$  bajo estimación.

En tales situaciones, se harán también afirmaciones / erróneas acerca de las funciones  $k^{\beta}$  siendo  $k$  una com- binación de las  $P_i^{\beta}$  para algunos valores de  $i \leq r$  y para/ algunos mayores que  $r$ . El estimador  $k^{\beta}$  puede tener / un gran sesgo, así como una varianza grande.

Por consiguiente, una varianza grande para un predictor puede deberse a la no-estimabilidad, como en el caso/ 2, o a una estimabilidad pobre, como en el caso 1, de // ciertas funciones, y, a menudo, es difícil distinguir entre los dos casos. La situación no sería en la práctica/ supuesto que se tiene la precaución debida al utilizar // estimadores sujetos a errores típicos grandes (calculando los intervalos de confianza).

Rao Potluri (1973) llevó a cabo una investigación so- bre diferentes líneas de acción cuando  $|X'X| = 0$  y cuan- do hay errores en  $X$ . Los resultados que obtuvo son inte- resantes desde un punto de vista práctico.

#### 4. VARIABLES CON ESTRUCTURA FACTORIAL

Sean  $y$ , una variable dependiente, y  $x_1, \dots, x_m$  / variables explicativas, con una estructura factorial, es/ decir, existen variables hipotéticas  $f_1, \dots, f_k$  tales / que

$$y = a_{01}f_1 + \dots + a_{0k}f_k + e_0 \quad (4.1)$$

$$x_i = a_{i1}f_1 + \dots + a_{ik}f_k + e_i, \quad i=1, \dots, m, \quad (4.2)$$

donde  $e_0, e_1, \dots, e_m$  están incorrelacionadas entre sf / y también están incorrelacionadas con los factores  $f_1, \dots, f_m$ . ¿Cómo utilizamos la información de la estructural factorial (4.1) - (4.2) al predecir  $y$  usando  $x_1, \dots, x_m$ ?

Para entender el problema consideremos el caso de una

variable con un sólo factor y la estructura

$$y = a_0 f + e_0, \quad x_i = f + e_i, \quad i=1, \dots, m \quad (4.3)$$

donde los coeficientes de  $f$  en  $x_1$  pueden tomarse como unitarios, sin pérdida de generalidad. Además sea

$$V(f) = \sigma_f^2, \quad V(e_i) = \sigma_i^2. \quad (4.4)$$

Es evidente que la regresión de  $y$  sobre  $f$  es

$$y = a_0 f \quad (4.5)$$

y que la de  $y$  sobre  $x_1, \dots, x_m$  es

$$y = a_0 \frac{\sigma_f^2}{1 + \sum_{i=1}^m \sigma_i^2} x_i, \quad a_0^2 = \sigma_f^2 / \sigma_i^2 \quad (4.6)$$

Un hecho digno de mención de la ecuación de regresión es que según aumenta  $m$ , disminuye el coeficiente de cualquier  $x$  fija, y la suma del coeficiente es

$$a_0 \frac{\sum_{i=1}^m \sigma_i^2}{1 + \sum_{i=1}^m \sigma_i^2} \quad (4.7)$$

que se approxima a  $a_0$  según aumenta  $m$ . El cuadrado de la correlación múltiple de  $y$  sobre  $x_1, \dots, x_m$  es

$$a_0^2 = \left( 1 + \frac{1}{\sum_{i=1}^m \sigma_i^2} \right) \left( a_0^2 + \frac{1}{a_0^2} \right) \quad (4.8)$$

que tiene como cota superior

$$a_0^2 \leq \left( a_0^2 + \frac{1}{a_0^2} \right) \quad (4.9)$$

Esto implica que cuando  $m \rightarrow \infty$ , la tasa de reducción de la varianza residual de  $y$ , dado  $x_1, \dots, x_m$ , será pequeña.

Supongamos que ajustamos la regresión de  $y$  sobre  $x_1, \dots, x_m$  sobre la base de  $n$  conjuntos de observaciones sobre  $(y, x_1, \dots, x_m)$  de la forma habitual. Las magnitudes de los coeficientes de regresión parcial disminuyen según aumenta  $m$ , mientras que los errores típicos de estos coeficientes podrían aumentar a partir de un cierto valor de  $m$ . El resultado es que, si se docima la significación de estos coeficientes con un nivel del 5 ó del 1 por ciento, puede considerarse un gran número de veces como no significativa.

Cuando hay más de un factor común, surge una situación semejante. El número de los coeficientes de regresión aumenta, pero sus magnitudes están sujetas a restricciones lineales. Puesto que tales restricciones no se admiten al ajustar la regresión, la precisión de los/

estimadores individuales disminuye según aumenta  $m$ . Además, hay un límite natural para la correlación múltiple, según aumenta  $m$ , y, en consecuencia, las décimas sobre los coeficientes de regresión individual llegan a ser cada vez menos potentes, a no ser que se incremente simultáneamente el tamaño de la muestra.

El ejemplo clásico de Fisher (1938) para la predicción de la aclimatación a grandes altitudes, usando 7 características a nivel del mar, puede explicarse por medio del fenómeno descrito anteriormente. En la fórmula con 7 características, ningún coeficiente fue significativamente distinto de cero a un nivel de significación del 5%. Comentando este fenómeno, Fisher observó:

Esto no significa que la fórmula carezca de valor, sino que todos los coeficientes individuales podrían variarse en gran medida, y, siempre que los otros coeficientes se ajustasen de forma adecuada, el valor predictivo no variaría apreciablemente. La fórmula con 7 variables es, por tanto, muy arbitraria, y una o más de las variables utilizadas debe ser ciertamente redundante.

Fisher procedió a eliminar algunas variables, una a una, suprimiendo cada vez la variable cuyo coeficiente tenía la razón  $F$  (o  $t$ ) mínima, hasta que los coeficientes de la ecuación de regresión, con el número de variables reducido, eran significativos al nivel del 5%. Fisher dió también otra explicación para eliminar algunas de las variables, argumentando que la varianza del error de predicción mediante la fórmula de regresión no debía ser inferior a la de los errores independientes en la variable dependiente. El uso de más de cuatro caracteres parecía indicar una precisión espuria.

No está claro lo que Fisher pensaba al establecer que la ecuación de regresión es arbitraria, indicando que algunas variables eran redundantes. Quizás se estaba refiriendo a la multicolinealidad, estudiada en la Sección 3 de esta conferencia. Sin embargo, la situación, como en el ejemplo de Fisher, puede surgir cuando las variables tienen una estructura factorial, como se demostró en el caso sencillo de un único factor. En tal caso, se puede proceder como sigue:

Consideremos, de nuevo, la estructura dada en (4.3) y sea  $\bar{x} = (\bar{x}_1 + \dots + \bar{x}_m)/m$ . La regresión de  $y$  sobre  $\bar{x}$  es:

$$y = \frac{a_0}{1 + m^{-2} \sum_{i=1}^m \sigma_i^2} \bar{x} \quad (4.10)$$

Si el coeficiente de  $x$  se estima de la forma habitual a /

partir de  $n$  conjuntos de observaciones sobre  $(y, x)$  no surge ningún problema del tipo que encontramos en el caso de los coeficientes de regresión múltiple. El error de predicción disminuye cuando  $m \rightarrow \infty$  y ninguna variable necesita ser considerada como redundante. Observamos que  $\bar{x}$  es una estimación (no necesariamente óptima) de la variable factorial  $f$ , así que la regresión de  $y$  sobre  $\bar{x}$  es una buena aproximación para la de  $y$  sobre  $f$ . En la práctica, no se conoce la estructura factorial ni el número de factores. En tal caso, puede considerarse la posibilidad de extraer los factores usando la matriz de dispersión estimada de  $(x_1, \dots, x_m)$ . Sean  $f_1, \dots, f_k$  los factores estimados como combinaciones lineales de  $x_1, \dots, x_m$  usando cualquier método típico del análisis factorial. Si la estructura factorial se mantiene como en (4.1)-(4.2), la regresión de  $y$  sobre  $f_1, \dots, f_k$  puede proporcionar una fórmula de predicción mejor que la regresión de  $y$  sobre  $x_1, \dots, x_m$ .

La técnica descrita es, probablemente, bien conocida por los expertos en análisis factorial, pero no es popular en la literatura estadística. Sin embargo, los economistas han intentado utilizar las componentes principales dominantes de  $x_1, \dots, x_m$  en la fórmula de predicción, que no es lo mismo que usar los factores, a menos que sean casi iguales las varianzas  $\sigma_i^2 = V(\epsilon_i)$  en (4.2) (véase, por ejemplo, Rao (1955)). Incluso bajo una estructura factorial, las componentes principales menos importantes pueden tener correlaciones altas con la variable dependiente y despreciarlas puede dañar la predicción. Debe tenerse presente que el método sugerido sólo es aplicable cuando la estructura factorial del tipo (4.1)-(4.2) se mantiene tanto para la variable dependiente como para las variables explicativas.

## 5. ESTIMACION DE UNA MEDIDA VERDADERA-UNA ANOMALIA

Supongamos que  $x$  es una medida verdadera (no observable) de una característica de un individuo  $y$  que  $y$  es la medida observada sujeta a error, es decir,  $y = x + e$ , siendo  $e$  un error independiente de  $x$ . El problema consiste en estimar (o predecir)  $x$ , dado  $y$ .

Consideremos la distribución bivariante de  $(x, y)$ . Puede verse que:

$$E(y) = E(x) = \alpha \quad (5.1)$$

$$V(y) = \sigma_x^2 + \sigma_e^2, \text{ cov}(x, y) = \sigma_x^2 \quad (5.2)$$

siendo  $\sigma_x^2$  y  $\sigma_e^2$  las varianzas de las distribuciones marginales de  $x$  y  $e$ , respectivamente. Por hipótesis, la

regresión de  $y$  sobre  $x$  es lineal, lo que no quiere decir que la regresión de  $x$  sobre  $y$  sea lineal. Las condiciones exactas bajo las cuales ocurre esto pueden verse en Rao (1974). Si suponemos que la regresión de  $x$  sobre  $y$  es también lineal, la ecuación de regresión es

$$\begin{aligned} \hat{x} &= \alpha + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} (y - \alpha) \\ &= \frac{\sigma_e^2}{\sigma_x^2 + \sigma_e^2} \alpha + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_e^2} y \end{aligned} \quad (5.3)$$

siendo  $y$  un estimador insesgado de  $x$  (o de regresión inversa) y  $\hat{x}$  es el estimador de regresión (directa) de  $x$ , que son iguales sólo cuando  $\sigma_e^2$ , la varianza del error, es cero. Los errores cuadráticos medios de estos dos estimadores son:

$$E(y-x)^2 = \sigma_e^2 \quad (5.4)$$

$$E(\hat{x}-x)^2 = \frac{\sigma_x^2 \sigma_e^2}{\sigma_x^2 + \sigma_e^2} \leq \sigma_e^2 \quad (5.5)$$

de modo que juzgando por el criterio del error cuadrático medio (o pérdida cuadrática), el estimador de regresión directa  $x$  es mejor que el estimador de regresión inversa  $y$ .

En la práctica  $\alpha$ ,  $\sigma_x$  y  $\sigma_e$  pueden no conocerse. Sin embargo, si los datos anteriores proporcionan determinaciones múltiples de la característica específica sobre, al menos, algunos individuos observados, pueden estimarse los parámetros  $\alpha$ ,  $\sigma_x$  y  $\sigma_e$  como se indican en Rao (1965). Usando los estimadores  $\hat{\alpha}$ ,  $\hat{\sigma}_x$ ,  $\hat{\sigma}_e$ , el estimador de la regresión puede escribirse:

$$\hat{x} = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} \hat{\alpha} + \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} y \quad (5.6)$$

que será todavía mejor que  $y$ , siempre que los datos anteriores sean suficientemente numerosos y se dispongan de estimaciones estables de  $\hat{\alpha}$ ,  $\hat{\sigma}_x$  y  $\hat{\sigma}_e$ .

Consideremos una posible aplicación del procedimiento de estimación anterior. Supóngase que una persona va a un hospital y el médico pide un análisis de colesterol en sangre al laboratorio clínico, que da un resultado de  $y$  unidades. Se sabe que la determinación del colesterol en sangre está sujeta a un error, cuya varianza es, por ejemplo,  $\sigma_e^2$ . Si se dispone de datos anteriores del análisis

del colesterol en sangre de personas, se pueden tener estimaciones de la media general,  $\alpha$ , de la varianza del verdadero colesterol en sangre entre las personas,  $\sigma_x^2$  y de la varianza del error en una determinación individual  $\sigma_e^2$ . ¿Debe el médico, por rutina, usar el estimador  $y$  o usar el estimador de regresión alternativo, posiblemente con error cuadrático medio menor:

$$\hat{x} = \frac{\hat{\alpha}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} \hat{\alpha} + \frac{\hat{\sigma}_e^2}{\hat{\sigma}_x^2 + \hat{\sigma}_e^2} y \quad (5.7)$$

con fines de diagnóstico?. Puede observarse que  $x$  es un estimador sesgado de  $y$  y  $|\hat{x} - \hat{\alpha}| < |y - \hat{\alpha}|$ , es decir,  $x$  se desvia algo menos de la media general que  $y$ , lo que indica una situación patológica menor que la revelada por  $y$ .

Para responder a esta pregunta, calculamos los errores cuadráticos medios de  $y$  y  $\hat{x}$ , dado  $x$ , suponiendo que se conocen  $\alpha, \sigma_x, \sigma_e$ , que son:

$$E[(y-x)^2/x] = \sigma_e^2 \quad (5.8)$$

$$E[(\hat{x}-x)^2/x] = \frac{\sigma_e^2 \sigma_x^4}{(\sigma_x^2 + \sigma_e^2)^2} + \frac{\sigma_e^4}{(\sigma_x^2 + \sigma_e^2)^2} (x-\alpha)^2 \quad (5.9)$$

Para  $x$  dado (una persona determinada), el error cuadrático medio de  $\hat{x}$  es mayor que el de  $y$  cuando

$$(x-\alpha)^2 > 2\sigma_x^2 + \sigma_e^2 \quad (5.10)$$

aunque el error cuadrático medio total (personas) de  $\hat{x}$  es más pequeño que el de  $y$ . En consecuencia, para grandes desviaciones del valor verdadero de  $x$  de la media general,  $y$  es mejor estimador que  $\hat{x}$ , y al contrario, para pequeñas desviaciones. Puesto que  $\hat{x}$  subestima (sesgo por defecto) y la subestimación, en el caso de grandes desviaciones, puede ser no deseable, desde el punto de vista del diagnóstico y tratamiento médicos, y parece ser una elección más segura que  $x$ , como estimador del desconocido  $x$  en un caso individual.

Además, supongamos que a una persona se le somete a tratamiento a causa de una gran desviación de la media general de su colesterol en sangre, y cada semana se le hace un análisis. Sean  $y_1, y_2, \dots$  las estimaciones correspondientes de la regresión. Puesto que las estimaciones de la regresión son sesgadas con respecto a la media general, la tasa de recuperación indicada por la serie de  $\hat{x}$  será más baja (sesgada) que la de la serie de  $y$ , que proporciona una estimación insesgada de la tendencia. En consecuencia, en la situación considerada,  $y$  parece que sirve un mejor propósito que  $\hat{x}$ .

Sin embargo, puede haber otras situaciones en que las consideraciones individuales no son importantes y en que la función de pérdida es cuadrática, lo que favorece la elección de  $\hat{x}$  e  $y$ .