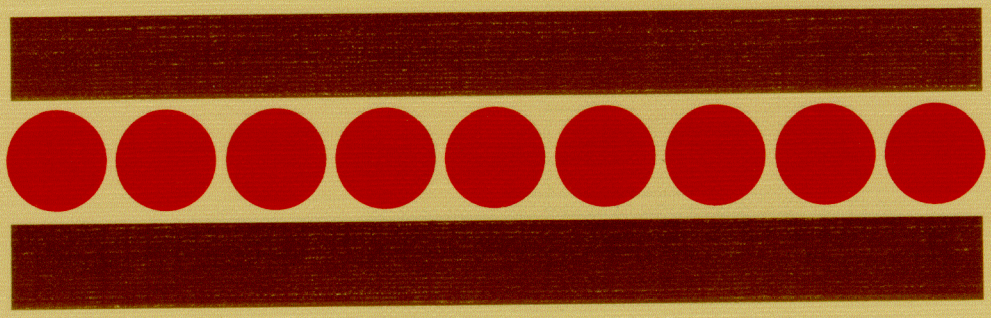
NAZIOARTEKO ESTATISTIKA MINTEGIA EUSKADIN

1986

SEMINARIO INTERNACIONAL DE ESTADISTICA EN EUSKADI



# ESTATISTIKA ERAGIKETAK LAGINKETAZ

•

# OPERACIONES ESTADISTICAS POR MUESTREO

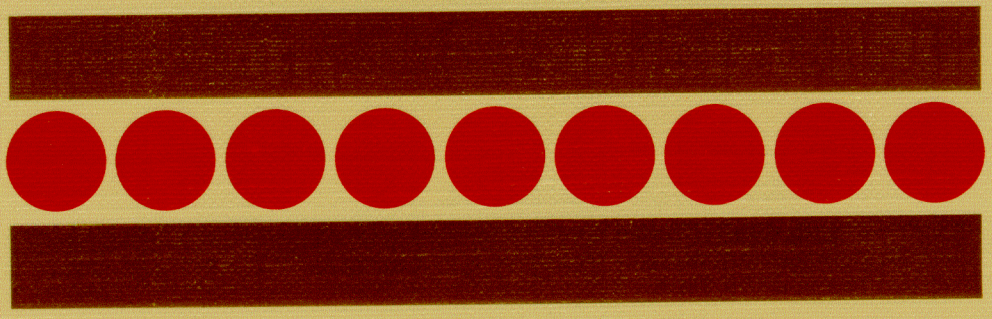
•

# OPERATIONS STATISTIQUES PAR ECHANTILLONNAGE

•

# STATISTICAL OPERATIONS BY SAMPLING

L. KISH



EUSKAL ESTATISTIKA - ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

EUSTAT

***Lanketa** / Elaboración:*
**Euskal Estatistika-Erakundea** /
Instituto Vasco de Estadística

***Argitalpena** / Edición:*
**Euskal Estatistika-Erakundea** /
Instituto Vasco de Estadística
C/ Dato 14-16 - 01005 Vitoria-Gasteiz

***Botaldia** / Tirada:*
1.000 ejemplares
VI-1989



***Inprimaketa eta Koadernaketa:***
*Impresión y Encuadernación:*
Itxaropena, S.A.
Araba kalea, 45 - Zarautz

***Lege-gordailua** / Depósito legal:* S.S. 963/89

ISBN: 84-7542-127-10 Obra completa
ISBN: 84-7749-05-1 Cuaderno 11

NAZIOARTEKO ESTATISTIKA MINTEGIA EUSKADIN **1986** SEMINARIO INTERNACIONAL DE ESTADISTICA EN EUSKADI

# ESTATISTIKA ERAGIKETAK LAGINKETAZ

•

# OPERACIONES ESTADISTICAS POR MUESTREO

•

# OPERATIONS STATISTIQUES PAR ECHANTILLONNAGE

•

# STATISTICAL OPERATIONS BY SAMPLING

L. KISH

KOADERNOA 11 CUADERNO

EUSTAT

## AURKEZPENA

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

- Unibertsitatearekiko eta estatistikaren alorrean interesaturik leudekeen guztien birziklapen profesionala erraztu.
- Funtzionari, irakasle, ikasle eta estatistikaren alorrean interesaturik leudeken guztien birziklapen profesionala erraztu.
- Estatistikako alorrean eta mundu-mailan irakasle prestu eta aqbangoardiako ikerlari diren pertsonaiak Euskadira ekarr, guzti horrek zuzeneko harremanei eta esperientzien ezagupenei dagokienez suposatzen duen ondorio positiboarekin.

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

Vitoria-Gasteiz, 1989ko maiatza

JOSE LUIZ NARVAIZA SOLIS
Zuzendari Orokorra

## PRESENTACION

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

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

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

Vitoria-Gasteiz, mayo de 1989

JOSE LUIS NARVAIZA SOLIS
Director General

## PRESENTATION

L'Institut Basque de la Statisque se propose d'atteindre plusieurs objectifs par la promotion des Séminaires Internationaux de la Statistique:

- Encourager la collaboration avec l'Université et spécialement avec les départements de la statistique.
- Faciliter le recyclage professionnel des fonctionnaires, professeurs, élèves, et tous ceux qui pourraient etre intéressés par la statistique.
- Inviter en Euskadi des professeurs mondialement renommés et des rechercheurs 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 recontres 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.

Vitoria-Gasteiz, mai 1989

JOSE LUIS NARVAIZA SOLIS
Directeur Général

## PRESENTATION

In promoting the International Seminars on Statistics, the Basque Statistical Office 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, 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 so 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-Gasteiz, may 1989

JOSE LUIS NARVAIZA SOLIS
Director General

# SARRERA

Liburu honek, Euskal Estatistika Erakundeak eta Euskal Herriko Unibretsitateak autolatutako IV Nazioarteko Estatistika Mintegiaren barnean L. Kish-ek «Estatistika eragiketak laginketaz» gaiari buruz emandako ikastaroa laburbiltzen du. Mintegi honek kontatzen du baita Kandan Director of the Methods Division of Census and Household Survey den R. Platek-en partaidetzarekin, «Methodology and Treatment of the non-response» gaiarekin eta I. Gallastegi, Euskal Herriko Unibertsitateko Ekonometriako Katedradunaren «Denborazko serieen azterketa. Aurresate metodo batzuk», kurtsoarekin.

# INTRODUCCION

Este libro resume el curso que sobre «Operaciones estadísticas por muestreo» ha impartido L. Kish dentro del IV Seminario Internacional de Estadística en Euskadi, organizado por el Instituto Vasco de Estadística y la Universidad del País Vasco. Este IV Seminario cuenta además con la participación de R. Platek, miembro honorario del ISI y Director de la División de Metodología de Censos y Encuestas en Hogares, de Statistics Canada, con un curso sobre «Metodología y Tratamiento de la no-respuesta», e I. Gallastegi, catedrática de Econometría, de la Universidad del País Vasco, con un curso sobre «Análisis de series temporales. Algunas técnicas de predicción».

# INTRODUCTION

Ce libre résume le cours donné par L. Kish au sein du IV Séminaire International de Statistique à Euskadi, organisé par l'Institut Basque de Statistique et l'Université du Pays Basque. Ce IV Seminaire compte aussi avec la participation de R. Platek, membre honoraire de l'ISI et Director of the Methods Division of Census and Household Survey à Statistics Canada, sur le sujet «Méthodologie et Traîtement de la non-réponse» et I. Gallastegi, Professeur d'Econométrie avec un cours sur «Analyse des séries temporaires. Quelques techniques de prédictiou».

# INTRODUCTION

This book summarizes the seminar on «Statistical operations by sampling» which L. Kish gave as part of the IV International Statistics Seminar in the Basque Country, organized by the Basque Statistical Institute and the University of the Basque Country. This IV Seminar also included the participation of R. Platek, honorary member of ISI and Director of the Methodology Division of Census and Household Survey in Statistics Canada, with a seminar on «Methodology and Treatment of the non-response», and of I. Gallastegi, University Professor of Econometry, in the University of the Basque Country, with a seminar on «Time series analysis. Some prediction techniques».

# BIOGRAFIA

***Leslie Kish*** *Hungrian jaioa eta orain Estatu Batuetako hiritartasuna du. Estatistika Matematikan formazioa dauka. 1971an International Statistical Institute-eko ohorezko bazkidea izendatu zuten. 101 urte dituen Erakunde honek, 98 herri eta estatuetako Estatistikako profesionariak elkartzen ditu. Azkeneko bi urte hauetan Leslie Kish, Inkesta Estatistiko saileko Lehendakaria izan da Erakunde honetan, Sail honek aktibitate eta dinamismo handiz ia 3.000 profesionari elkartzen ditu. Beste ohorezko titulu artean American Association for Advancement of Science-ko ohorezko bazkidea izendatua izan zen 1978an. Estatistikalari askok ezagutzen dituzte bere laginketari buruzko liburuak, eta nazioarteko metodologi buruzko aldizkari ospetsuenetan 100 artikulu baino gehiago idatzi ditu. ONU, UNESCO, OMS... eko estatistikako aholkularia izan da, eta anitz herrietako (Txina, Kuba, Brasil, Ekuardor, Txile etab.) inkesta nazionaleko proiektuetan parte hartu du.*

***Leslie Kish,*** *húngaro de nacimiento y con ciudadanía estadounidense en la actualidad. Estadístico matemático de formación. Nombrado en 1971 miembro honorario del International Statistical Institute. Instituto que reúne a profesionales de la Estadística de 98 países y estados de todo el mundo, con una historia de ya 101 años. En este Instituto, Leslie Kish ha sido en el último bienio Presidente de la Sección de Estadísticos de Encuestas, sección que con gran actividad y dinamismo reúne a casi 3.000 profesionales. Entre otros muchos títulos honorarios, se le nombró en 1978 miembro honorario de la American Association for Advancement of Science. Sus libros de muestreo son conocidos por muchas generaciones de estadísticos, y ha escrito más de 100 artículos en las más prestigiosas revistas de metodología estadística del ámbito internacional. Ha sido asesor estadístico en la ONU, UNESCO, OMS, ... y ha participado en proyectos de encuestas nacionales de muchos países (China, Cuba, Brasil, Ecuador, Chile, etc.).*

***Leslie Kish,*** *Né en Hongrie, il es actuellement nationalisé aux Etats Unis. Il es Statisticien Mathématicien de formation. Il fut nomé en 1971 membre honoraire de l'International Statistical Institute. Cet Institut réunit des professionels de 98 pays et états de tout le monde, il a une vie de 101 ans déjà. C'es dans cet Institut que Leslie Kish a été dans ces deux dernières années Président de la Section de Statisticiens d'Enquêtes, cette section réunit avec une grande activité et une grande dynamique presque 3.000 profesionels. Entre d'autre titres honoraires, il fut nomé en 1978 membre honoraire de l'Americain Association for Advancemente of Science. Ses livre d'échantillonnage sont connus par plusieurs générations de statisticiens, il a écrit plus de 100 articles sur des revues les plus prestigieuses de méthodologie statistique dans tout le monde. Il a été assesseur statistique à l'ONU, l'UNESCO, l'OMS, ... et il a participé dans des projets d'enquêtes internationales de plusieurs pays (Chine, Cuba, Brésil, Equateur, Chili, etc.).*

***Leslie Kish,*** *hungarian born and at present with United States citizenship. Statistical mathematician by profession. In 1971 he was named honorary member of ISI. The Institute which unites professional statisticians from 98 countries and states throughout the world, which has now existed 101 years. In this Institute, Leslie Kish has been President of the Survey Statisticians Section for the past two years. This section actively and dinamically unites nearly 3.000 professional people. Among many other honorary titles, he was named honorary member of the American Association for the advancement of science in 1978. His books on sampling are known by many generations of statisticians, and he has written over 100 articles in the best international statistical methodology journals. He has been statistical assesor in the UNO, UNESCO, WHO... and he has participated in national survey projects in many countries (China, Cuba, Brazil, Ecuador, Chile, etc,).*

# ESTATISTIKA ERAGIKETAK LAGINKETAZ

L. KISH

## LABURPENA

Lehen kapituluak 1984eko 7. Australian Statistical Conference-en inaugurazioko hitzaldiari buruz dihardu, non, labur-labur, gobernu-politikarako inkesta garrantzitsuak hobetzeko lagin-diseinuetan egindako zazpi aldaketa azaltzen bait dira. Aldaketa horiek dira: arlo txikietarako estimazio hobeak, lagin pirikari metatuak, panel-inkesta gehiago, diseinu asmoanitzak lagin periodikoetarako, panelak lagin ez-teilakatuekin konbinatzen dituzten panel-diseinu azpizatituak (SPD), eta bilketa usu metatua informazioaldi gutxiago usuetarako.

Bigarren kapituluak lagin-diseinu asmonitzez tratatzez du. Inkesta gehienak asmo askorekin egiten dira eta hemen sei mailako hierarkia bat proposatzen dugu. Gaur egun oraindik, teoria eta testuliburu gehienak teoria asmobakarrean oinarritzen dira, konplexutasuna eta diseinu asmonitzen gatazkak eragozteko. Hamar gatazka-arlo seinalatzen dira asmoen artean, eta gero horietako bakoitzarentzako problemak eta soluzioak ematen dira. Konpromisu eta soluzio bateratuak problemak eta soluzioak ematen dira. Konpromisu eta soluzio bateratuak bideragarriak dira zorionez, optimo asko oso lauak direlako; eta baita zehaztasunerako behar diren baldintza gehienak oso malguak direlako ere. Asmo asko ezarri eta horiei aurre egitea hobe da, artifizialki ezarritako asmo bat bakarraren atzean ezkutatzea baino. Gero, noski, bideragarriago egiten da hau ordenadore modernoak erabiliz.

Eta hirugarren kapituluan, argitalpen hau dagokion Mintegian zabalago tratatu izan ziren inkesta-laginketako oinarrizko hamahiru alderdi azaltzen dira. Alderdi horiek dira; lagin-diseinua eta inkesta-diseinua, biztanleriaren definizioa. Inkesta-elementuak eta —unitateak, probabilitate-ahutespena, hautespeneko berdinak— EPSEM, langiketa-tasa finkoak (f), laginaren ez-tamaina (n), pisuak kasuei batzea = elementuak, iturriak eta pisuen efektuak, markoaren arazoak, lagin-tamainen diseinua, SRSko desbiderapenak, elementu geruzatuen laginketa, konglomeratu-laginketa eta laginketa-erroreen kalkulua eta aurkezpena.

# OPERACIONES ESTADISTICAS POR MUESTREO

L. KISH

## RESUMEN

El primer capítulo trata sobre el discurso de apertura de la 7.ª Australian Statistical Conference (1984) que expone brevemente siete modificaciones en diseños muestrales para mejorar la utilidad y la oportunidad de encuestas importantes para la política gubernamental. Estas son: mejores estimaciones para dominios pequeños, muestras rodantes acumuladas, más encuestas de panel, diseños multipropósitos para muestras periódicas, diseños subdivididos de panel (SPD) que combinen paneles con muestras no solapadas, y recopilación frecuente acumulada para períodos de información menos frecuentes.

El segundo capítulo trata de los diseños muestrales multipropósito. La mayoría de las encuestas se realizan con muchos propósitos y aquí proponemos una jerarquía de seis niveles. Todavía, la mayor parte de la teoría y de los libros de texto se basan en la teoría unipropósito para evitar la complejidad y los conflictos de los diseños multipropósito. Se indican diez áreas de conflicto entre los propósitos, después se dan los problemas y soluciones para cada uno. Los compromisos y las soluciones conjuntas son afortunadamente factibles porque muchos óptimos son muy planos; también porque la mayoría de los requisitos para la precisión son muy flexibles. Establecer y hacer frente a muchos propósitos es preferible a esconderlos detrás de un único propósito establecido artificialmente. También se hace esto más factible mediante el uso de los modernos computadores.

En el tercer capítulo se exponen trece aspectos básicos del muestreo de encuestas que fueron tratados más ampliamente en el Seminario al que corresponde esta publicación. Estos aspectos son: diseño muestral y diseño de encuestas, definición de población. Elementos y unidades de encuestas, selección de probabilidad, probabilidades iguales de selección —EPSEM, tasas de muestreo fijas (f), no tamaño de muestra (n), unir pesos a casos = elementos, fuentes y efectos de los pesos, problemas del marco, diseño de los tamaños muestrales, desviaciones de SRS, muestreo de elementos estratificados, muestreo por conglomerados y cálculo y presentación de los errores de muestreo.

# INDEX



# INDICE

# STATISTICAL OPERATIONS BY SAMPLING

# 1. TIMING OF SURVEYS FOR PUBLIC POLICY

**Summary**

This keynote address at the 7th Australian Statistical Conference (1984) discusses briefly seven modifications of sample design for improving the usefulness and timeliness of surveys relevant to public policy. These are: better estimates for small domains, cumulating rolling samples, more panel surveys, multipurpose designs for periodic samples, split panel designs (SPD) combining panels with nonoverlapping samples, and frequent collections cumulated for less frequent reporting periods.

## 1.1. Introduction

These notes concern several distinct aspects of the timing of surveys and their relevance for public policy. I developed them separately in consulting and reports, for diverse institutions in several countries and continents. Yet they have some mutual coherence, some relationships, I believe, so that this joint presentation has some heuristic value.

Each of the seven points I direct toward some modifications of present practices, with the aim of higher satisfactions of some aims, closer approach to or higher efficiency for some criteria. Since criteria, aims and values belong to the realm of public policy, these aspects address the interaction of technical statisticians with persons and offices concerned with public policy.

I am advocating some changes, modifications in several aspects of the timing of surveys. But suggesting changes implies criticism of present practices, and that invokes the most natural defensive reactions of *flight or fight*. Diversity and disagreements without violence are the very staff of life; Australians are famous for them, I also love disagreements and hope we have room and time for them. But there is no need for trivial and artificial arguments, and let me hold out three branches of olives.

First, the changes I advocate are not drastic. Rather they propose evolutionary modifications of present practices, and often adaptations of methods already used in other situations. Actually their utility may be underestimated because they lack the novelty of *powerful new tools*, of brilliant technical innovations. Thus they may be caught between two sources of inertia: fear of changes and the underevaluation of new uses for known methods.

Second, my remarks are general and not directed specifically at Australia. My knowledge of Australia's needs and practices is not specific enough nor deep enough. But I know that Australia's statisticians are in touch with the world's leaders, in close touch because they are right among the leaders. Many of the problems and practices among them are similar, and that goes both for advances and for unmet needs, hence for possible changes. Inertia, like progress, can be contagious. Australia has one distinct feature for samples: its population spread over an area comparable to the world's largest countries, though smaller by factors of ten to seventy; Canada with a factor of two is the closest.

Third, these remarks concern only sample surveys, because that subject is enough for me today. They impinge indirectly on related subjects like the timing and scope of censures, the future possibilities for registers, records and other administrative sources. But discussing them is beyond the limits of my time and my abilities.

Lastly, I express my optimism that better methods will prevail, that scientific advances create positive feedbacks, that success breeds. The advances of sample surveys have not satisfied a fixed demand, but created even more demands awaiting

solutions, which should be not only theoretically sound but also operationally feasible.

### 1.2. **Estimation methods for small domains**

Estimates for local areas and other small domains have been of general interest for a long time, but have been unavailable except for estimates from population censuses, special surveys, or administrative registers. These interests, however, have been superseded by increasing demands for more diverse, rich and current data for small domains, which are required for the planning of reforms, welfare and administration in many fields. Very recently the demands for data to be used directly for apportioning money and resources have been added (at least in the U.S.A.) to the needs of planners and to the curiosity of scientists. These demands are now greatly increasing research interests and efforts in this area.

Estimates for small domains have been largely neglected, until recently, by statistical and sampling theory. As exceptions we note that, for population counts of local areas, statistical demographers have developed several competing methods...

Only in recent years have estimates for small domains become an active area for research. This has resulted in investigations of a variety of statistical techniques for application to problems of estimation for small domains. (Purcell & Kish, 1979).

Recently a new method of *iterative proportional fitting* (IPF) has been developed, as a method using categorized data analysis. These methods of *structure preserving estimation* (SPREE) yield new tools for estimation of small domains, including small areas but also other small populations. Their development needed not only a new theoretical view but also the development of high speed modern computers (Purcel & Kish, 1980).

For methods of small domain estimation Australia is a leader in the *development* of new methods, with several contributions during the past decade. Australia is also a leader in the *application* of the new methods, with recognition and use in public policy.

This section began with the new demands from public policy for timely data for small domains, demands arising from the needs for public planning and for distributing public funds. However I believe that demands for data are almost limitless and the existence of feasible methods may be the chief constraints and facilitators.

### 1.3. **Cumulating rolling samples**

Over the years we have gotten used (in the USA) to two methods for collecting information from the public: the decennial census and the Current Population Survey... The contrasting natures of the efforts-large, rare and singular for the census, but small, frequent and regular for the surveys-explain in large part the common acceptance of these large differences in reporting periods; monthly for surveys and once every 120 months for censuses. After all, filling the gaps between the 120 months depends on assumptions of smoothness, of stability of the observed variables. On the other hand, the relatively small sizes of monthly samples depend on accepting averages over larger spatial domains. A good case can be made that many, perhaps most, of our needs for data could be better met with yearly reports, and some with quarterly reports, but with broader sample bases than are feasible from our current monthly samples.

A good case is now being made by demands for more current data for local areas. The mid-decade census would be a response to some demands, but with 60 month gaps it would still not serve many other purposes. Yearly censuses have been mentioned in some situations, but even 1 percent samples would need large, sporadic efforts tha would tend to lower quality and raise unit costs.

A solution to the seemingly conflicting demands of economy, timeliness and detail may come from cumulated rolling samples (Kish, 1981). (See also Kish 1979, 1965, 12.51; NCHS 1958). These quotations refer to the situation in the USA, whereas Australia and Canada (and perhaps others) have quinquennial censuses. For these countries the question changes: should yearly cumulated samples supplement or replace the quinquennial censuses?

Rolling (rotating) samples refer to cumulated samples, discussed in 6 below, that have been designed especially to *roll over* the entire population in space and over designed periods. But the cumulation may be designed for a large (e.g. 10 percent) sample census rather than for a complete census (Section 6).

Rolling samples (2) and improved estimates (1) both promise to yield more timely estimates than the periodic censuses, plus more detailed data and better statistics for small domains than sample surveys. Is there much redundancy here so that one would need only one of the two methods and disregard the other? No. Happily the two methods reinforce each other with synergistic effects. The joint use of better estimates from rolling samples can be found in the references.

### 1.4. **Panels of individuals**

Panels here denote samples in which the same elements are measured on two or more occasions to obtain *individual* changes $d_i = (x_{i1} - x_{i2})$. From a good sample of the $d_i$ the distributions of individual changes in the population (i=1, 2, ..., N) can be estimated for diverse variables. From the means of these *gross, individual, internal* changes we can estimate the *net, external* change: $\bar{d} = (\bar{x}_2 - \bar{x}_1) = \Sigma\,(x_{i2} - x_{i1})/n$. But from these net changes of means one cannot estimate (directly) the gross changes of individuals. Only panels can reveal the gross, micro changes behind the net macro changes-except for unusual situations of either dependable memory or strong monotonic models. To study the micro changes behind the macro changes we need panels. We need them not only for their absolute and relative frequencies, but also for the dynamics of relationships and of causation at the individual level (kish, 1958, ch. 6; Kish, 1965, 9.5, 12.5C; Duncan & Kalton, 1986). The word *longitudinal* o *strictly longitudinal* is used sometimes for panels, but that word is also used for *longitudinal data* in retrospective studies from one interview, also in *longitudinal studies* of single sites (city, institution) with populations changing over time (Duncan, Juster & Morgan, 1986).

Sometimes panels may seem too dificult and not even feasible due to mortality, mobility and refusals. However, often those obstacles are not insuperable, and panels should be investigated and incorporated into surveys. The evidence here comes from the many *overlapping* samples in use. Panels define special cases of complete overlaps, when the sampling units are the elements themselves. Overlapping samples based on stable area segments yield good current estimates and net changes, but due to the mobility and mortality of individuals they fail to provide panel data automatically. Hence, I propose that some portion of overlapping surveys can and should be converted to panels, in order to provide the micro data on the dynamics of individual behavior, so much needed for and so lacking from social statistics today.

### 1.5. **Purposes and designs for periodic samples**

Sample designs should follow their aims, uses, purposes-rather than dictate them-and they should follow them closely. We begin with listing five major purposes, then look at sample designs that seem best for each purpose (either necessary or most efficient). Last we look at what compromises, if any, can be made to *satisfice* or *proximize* for several purposes; instead of optimizing for one purpose and neglecting all others. The discussion refers to periodic samples, because they seem most relevant, but they can also serve for samples repeated *ad hoc* at irregular intervals.

**TABLE 1**
**PUROSES AND DESIGNS FOR PERIODIC SAMPLES**

| Purposes | Designs | Rotation scheme |
|---|---|---|
| A. Current levels | A. Partial overlaps $0 < P < 1$ | abc - cde - efg |
| B. Cumulations | B. Nonoverlaps $P = 0$ | aaa - bbb - ccc |
| C. Net changes (means) | C. Complete overlaps $P = 1$ | aaa - aaa - aaa |
| D. Gross changes (individual) | D. Panels | Same elements |
| E. Multipurpose time series | E. Combinations, SPD | |
| | F. Master frames | |

In Table I we present five purposes and six designs. The first four are paired with similar letters in the same four rows. These pairings call attention to the designs that best serve, with reduced variances and otherwise, each of the four purposes. Most periodic studies have several purposes, hence we should face the difficult problems of multipurpose design-though not necessarily solve them. Actually three of the purposes can be served with other designs also, but with increases in variances or in cost; the lone exception is gross changes for individuals that require panels, as noted in Section 3. Overlapping samples of area segments will not suffice because individual elements (persons) move between periods. On the contrary, for cumulations nonoverlaps are best. Thus reasonable compromises become possible and what we need are clear definitions of purposes and uses. These belong to the domain of public policy.

Extraneous considerations may rule out some designs, and overlaps may be either prohibited or, on the contrary, required; they would result in less efficient designs, but they can still be valid designs. The chief variation among these designs concerns the amount (and kind) of overlaps between periods. The rotation scheme of complete overlaps shows, with *aaa-aaa-* etc., that all parts are common to the periods. The nonoverlaps with *aaa-bbb-* etc., shows no common parts. The partial overlaps with *abc-cde-efg-* etc., shows *c* and *e* as 1/3 overlaps between succeeding periods only.

The approach used here can be made more general in several aspects, whereas here we have to be specific and simple. We concentrate on how varying the relative overlap (P=1 for complete overlaps, P=0 for nonoverlaps and O<P<1 for partial overlaps) affects the variances for means of diverse statistics representing three basic purposes. The sizes and structures of the periodic samples are assumed to remain constant.

There are other simplifications also. The comparisons of variances ignore possible differences in element costs; but reinterviews, especially with telephones, may cost much less than new interviews. Second, the variances, based on $S^2/n$, ignore the design effects of clustering; but these will often be less for overlaps than for nonoverlaps. Third, the discussion assumes positive correlations R which is most likely; but R may vary from low values for some variables (like unemployment) to high values for others (like occupation, economic status).

**TABLE 2**
**EFFECTS OF DIVERSE OVERLAPS ON VARIANCES OF DIFFERENCES OF TWO MEANS**

| Designs | Specified Sample Sizes for Special Cases | $1/n_x + 1/n_y - 2P_xP_yn_c/n_xn_y$ |
|---|---|---|
| A. Partial overlap | $n = n_x = n_y$, $n_c = P_n$ | 2 (1 − PR) |
| B. Nonoverlap | $n = n_x = n_y$, $n_c = P = 0$ | 2 |
| C. Complete overlap | $n = n_x = n_y$, $n_c = P = 1$ | 2 (1 − R) |
| D. Subset | $n = n_x = n_y$, $n_c = P_n$ | (1/P + 1 − 2R) |

Samples with nonoverlaps, P=0, are best for cumulations, because they are simpler and also yield lowest variances: $S^2/2n$ for means of two periods and $S^2/Jn$ for J periods. These variances would be higher in overlapping samples due to positive correlations. The variance for the difference of two periods would be $2S^2/n$ for nonoverlaps.

On the other hand, the variance is only (1 - R) $2S^2/n$ for measuring net changes with complete overlaps, P=1. Thus the effect of complete overlaps is (1 - R) on the variances of differences of means. This is the optimal limit of (1 - PR) of the effect of partial overlaps. But with improved weighted estimates this can be reduced greatly, so that the effect on the variances becomes (1 - R)/

[1 - (1 - P) R], which for higher P values approaches (1 - R). Thus estimation can almost replace high overlaps. These gains on the variances of net changes can be high with large values of P and R (Kish, 1965, 12.4-12.5).

For measuring gross (macro) changes of individuals, panel studies are needed; these are discussed in Section 4. The conflict between the needs for panels and for cumulations can be resolved with Split Panel Designs, discussed in Section 6. These appear as primary examples of combinations of designs for fulfilling the needs of multipurpose designs. The needs of time series are not well defined in terms of sample surveys, but we can begin with the four basic designs above (for further discussions of the above see Kish (1985), 6.2).

### 1.6. **Split panel designs**

With this name (SPD) I hope to distinguish this new type of design that aims to combine the advantages of all the four basic designs (partial overlaps, nonoverlaps, overlaps, panels) though it is distinct from all four of them.

The basic notion is to add to a panel $p$ a series of nonoverlapping samples, denoted as *a-b-c-d* etc. Thus the periodic SPD is symbolized by *pa-pb-pc-pd*. The panel $p$ yields individual changes, and the nonoverlaps *a-b-c-d* can be cumulated into larger samples. The combined sample provides the partial overlaps best for current estimates, and good for net changes. Thus, SPD satisfies for all five purposes. The sample designs for the panel and for the nonoverlaps can differ and be optimized separately; but the population and measurements used must be similar and compatible for the combination.

SPD differs from the classical designs of partical overlaps, which have the eye-appeal of symmetry: all sample groups are in the sample for similar patterns of waves, all respondents subject to the same (small?) fatigue. Classical designs are rigidly designed only for mean changes between prespecified intervals, but critical intervals discovered later are neglected (though they are most common). On the contrary SPD has the overlap $p$ for *all* intervals. Another advantage of SPD comes from the higher correlations of covering the same elements, when these exhibit great mobility between the units used in ordinary overlaps. Furthermore, because SPD has overlaps for all intervals, it has advantages in flexibility in estimation over rigidly prespecified partial overlaps; since these may not have the interval with the highest correlations $R$ for measuring changes. For example the overlaps may be yearly but the highest $R$ may come monthly (or vice versa).

The chief difference and principal advantage of SPD come from having a proper panel $p$ that classical overlaps lack. Panels are necessary for measuring individual changes. But they also have unique problems. Yet SPD also holds a great advantage over a panel alone: it allows for cumulating evidence to check against and to correct the biases which panels may incur. Another advantage lies in the possibility of using the changing samples to recruit from them replacements for panel mortality and for panel renewals.

Other advantages of the flexibility of SPD concern the relative sizes of the panel $p$ and the nonoverlap *a-b-c-d*. In favor of high $p$ are statistics on individual changes, also reduced variances for mean changes; also perhaps lower cost per interview especially with telephone reinterviews. But lower values of $p$ (around 1/3) are better for current levels. They also reduce the fear of possible biases in panels. A great advantage of SPD is that the sizes of nonoverlaps can be varied to suit changing budgets and changing needs.

However, we must recognize that panels face great obstacles in the mobility in birth/death processes, in respondent fatigue and reluctance. For these reasons rotating designs with partial overlaps may be more appropriate for many purposes.

SPD have several features that result in their several advantages over other designs. If some feature seems unfeasible it may be omitted and we can still retain some of the advantages of SPD. For example, if $p$ becomes a permanent overlap of area segments, the panel is sacrificed, but other advantages can be retained. Further this overlap may be rotated slowly with some sacrifice of correlations. On the other hand a slow rotation may also be built into a true panel, and thus retain most of the panel's advantages but reduce the feared fatigue and deterioration of panels. the flexibility of the changing sizes of the nonoverlaps *(a-b-c-d)* is another feature that can be introduced into designs that are not pure SPD (Kish, 1985, Section 6.7; Kish, 1981b).

## 1.7. Frequent collections cumulated for less frequent reporting

We need to distinguish three kinds of periods for sample surveys: *collection* periods during which data are collected; *reference* periods that may differ that may consist of one reference period or cumulated for several. For example, the monthly surveys of the CPS of the USA use one week for collecting and the previous week as reference; and with judgmental inference to the whole month, which also serves for reporting. There exists a strong and needless tendency to fuse these three periods together and to fail to consider them separately. I proposed that the reporting period should be less frequent (quarterly and yearly) and the collection period more frequent (weekly) (Report to the National Commission on Employment and Unemployment Statistics, 1978).

We now have so much information on the economy, in such detail, so frequent and so relatively accurate, that there is a temptation for policy-makers to tune too finely, to change too frequently, to look at every last figure and be thrown by it. We do our forecasting as well as anybody in the world, but there is a tendency to miss the wood for the trees. (Sir Claus Moser, Director of the Central Statistical Office, UK, 1978). I also believe that our society and government are not geared to react to monthly changes.

Second, monthly fluctuations (of employment, etc.) as compared to quarterly or yearly changes, are often subject to *random shocks* that are not components of reasonable models. With most monthly reports we also receive *ad hoc* explanations involving the calendar, storms, strikes, or news events.

Third, the effects of sampling errors would also be reduced for quarterly reports. These reductions would be most welcome for domains: regions, geographic, demographic and economic domains that may be as important as national totals. For domains the sampling errors in monthly reports often swamp any meaningful fluctuations, and these would especially benefit from cumulations into quarterly and yearly reports.

Fourth, the monthly reports can continue to be prepared for governmental, scientific and other uses but their unreliability should be stressed realistically. However, the sample design, emphasis and presentation should shift to quarterly and yearly reports. I gladly note that the monthly Swedish Labour Force Surveys (AKU) use three independent monthly samples for each quarter. But there is a 7/8 overlap from quarter and every new sample is used for 8 consecutive quarters (Dahlström & Wahlström, 1973). *The selected rotation system implies a preference for precision in the estimates of quarterly averages and of differences between quarterly averages. However, at the present time the Swedish news media seem above all to discuss changes from one month to the next. Such changes could have been estimated better, for certain variables considerably better, with a different rotation system. Still, the system chosen is based on the preferences of other customers.* Well said and well done! In 1984 the UK has also shifted its Labor Force Surveys to cumulating three non overlapping monthly samples in each of its quarterly reports.

There is a 4/5 overlap between quarters and a 1/5 overlap between yearly reports. Each dwelling remains in the sample for 5 successive quarterly samples.

Fifth, statistical strategy should dictate less frequent reporting especially for small but non-negligible domains. Too often such domains either remain unreported, or they are reported with unduly large errors or at too great a cost, or both. As prime examples consider the vast but underpopulated areas with small populations which appear in many countries and for which provincial authorities demand separate reports. Other examples come from demographic, ethnic, occupation groups, etc., for which separate data are needed. Instead of the usual rigid practice prevailing now, it would be preferable to report at, to cumulate for, and to design for longer periods for these smaller domains. The tables for these statistics should indicate the different designs used for those statistics.

On the other hand there is no need to confine the collection period to a single week within the month, designated by arbitrary selection and subject to the vagaries of chance, storms and the calendar. If spreading the sample within each Primary Sampling Unit seems too expensive, the four weeks can be spread in balanced manner among the PSU's. Actually most sampling areas are *self-representing* metropolitan areas, and the PSU's are blocks or enumeration areas, and these can be spread over the four weeks.

## 1.8. **Timing and institutions**

Great differences can be observed between countries both in the sources of funds for social and survey research and in the nature and structure of operating survey centers. Some of these differences of organization may be traced to country sizes in area and population, some to development, both economic and statistical, and some to political and social structure. I also suspect that much of it is arbitrary, idiosyncratic, random, due to accidents of history and personalities. In any case countries could learn from each other more often, rather than simply accepting a Pangloss theory that *their* situation represents an optimal fit.

Funds come from four sources, roughly grouped: government offices; foundations, public, private and international; business and industry; and universities. Survey centers are of three major types: governmental; private, business and market research; universities and research institutes. Of the $4 \times 3 = 12$ possible flows of funds from sources to centers, most are used, but governmental centers seldom receive nongovernmenal funds except from international bodies. And universities seldom give funds to outsiders. These remarks are based on qualitative personal experience and I have no statistics to back them up, nor for the following remarks.

Three concerns about periodic studies link this section with the title and the paper. First the sources of funds for most periodic surveys are governmental. Other sources have had neither funds for nor interest in the long term results. Second, most periodic surveys are done by governmental research centers; university centers have shown less interest, capacity or initiative for periodic work. But there are notable exceptions in long term studies of my Institute at Michigan, of NORC at Chicago, of SCPR now in London, and probably elsewhere. Third, the social sciences and public policy both need many more periodic studies. Some of these should be done at universities and in other nongovernmental research centers.

Finally and briefly let me exhort you all to help the development of sample surveys, both periodic and *ad hoc* both inside and outside government offices. Also for more cooperation between the three kinds of centers, rather than competition. There is need and room for more help going both ways. Universšity researchers can help with better methods- though many *methodologists* are too far removed from reality to be of much help. Nongovernmental researchers need help from public sources: financial, methodological and policy guidance. A specific suggestion concerning periodic studies: the pilot, tryout and beginning periods may be better done outside governmental offices, perhaps at university centers; then the proven, stabilized, accepted studies can be taken over by the government.

At the close, I have nothing to add to the widespread discussions on how surveys (or statistics in general) should be organized between government departments within a country. There are many who have more to say on this topic, and so I may become silent at last.

# II. MULTIPURPOSE SAMPLE DESIGNS

**Abstract**

Most surveys have many purposes and a hierarchy of six levels is proposed here. Yet most theory and textbooks are based on unipurpose theory, in order to avoid the complexity and conflicts of multipurpose designs. Ten areas of conflict between purposes are shown, then problems and solutions are fortunately feasible because most optima are very flat; also because most *requirements* for precision are actually very flexible. To state and to face the many purposes are preferable to hiding behind some artificially picked single purpose. It has also become more feasible with modern computers.

## 2.1. Introduction

Most studies have several purposes during the planning stages and then typically many more purposes emerge later during the analyses of data and during their interpretation and utilization. However, the real multipurpose nature of most studies today tend to remain hidden under the surface of oversimplified, univariate discusions to study designs. This seems most clearly evident for sample surveys, which I shall discuss here, but I believe that this discrepancy also holds for other statistical designs.

In practice, surveys are usually multipurpose. Why then are multipurpose designs neglected in sampling theory? Because multipurpose theory would be too complex and difficult, and sampling theory is rather complex already. Even the descriptions we read of actual sample designs tend to follow and to borrow the prestige of univariate and unipurpose sampling theory, rather than to portray faithfully the many compromises of complex reality. Many common designs (especially epsem selections) probably serve robustly a variety of purposes, but explicit planning of multipurpose designs seems to be rare. Therefore this topic seems to merit the highest rating on a scale of need/availability ratio.

There are several aspects to the multipurpose nature of survey samples, and these are displayed in a hierarchy of six *levels* in Section 2. Then ten areas on *conflict* between purposes are specified in Section 3. Sections 4 to 9 deal with specific areas of conflict, presenting approaches to and solutions for them. Some of these are attributed to widely dispersed articles or survey sampling, but others are more and less fully developed, derived and referenced.

In this overall review I aim first and foremost to serve practitioners with a handy reference on approaches, methods and procedures for multipurpose designs; to alert them both to the importance and to the feasibility of such designs. Second, I wish to provide a framework for further integrated work on the many problems and conflicts of multipurpose designs. Imperfections of my methods can serve as stimuli to others for better derivations for them as well as for developing new methods.

## 2.2. A hierarchy for levels of purpose

To begin with, we need some clarification of the meaning of *multipurpose*, because too many concepts are confused under this term in our literature. Most of the time either levels 3 or 4 in Table 2.1. are meant, and for that reason *multisubject* has been distinguished from multipurpose for the same or closely related variables (Murthy, 1967, Section 9.22, 14.13). Note that each of these levels can have several specific manifestations, which would be useful to discuss when time is available.

Integrated survey operations on level 5 are related to, but should be distinguished from multisubject surveys, because they refer to organizations and institutions that conduct many surveys in diverse fields over longer periods of time (UN 1980, Foreman 1983). An earlier name was *continuing survey operations*, when it was recognized that most large-scale, widespread sample surveys were conducted by continuing survey organizations like the U.S. Census Bureau, or our Survey Research Center. Such continuity has large advantages in costs and quality, with profound effects on the sample design (Kish 1965, Section 12.6).

Master frames or master samples on level 6 refer to further extensions and specializations of multipurpose approaches. They may refer simply to using the same maps, or block listings, or area segments for several different surveys; or to the large-scale example of the *Master Sample of Agriculture* (King and Jessen 1945), where rural areas on the maps of all the counties of the USA were divided into segments of about four farms each; or to the firm that sells current listings of dwellings for most samples used in Western Germany. These very diverse examples have common bases in the savings from sharing the *startup* costs (of design, stratification, listing, etc.) for constructing sampling frames.

Levels 1 and 2 have been typically neglected in the literature of multipurpose sampling, although they are the most common and they can have the most drastic effects and cause the most dramatic conflicts, as we shall see later. The effect of designs can be very different for statistics like medians and quantiles or regression coefficients than for means and for aggregates (Kish 1961; Kish 1965, Sections 3.5, 12.9; Kish and Frankel 1974). Furthermore, designing for periodic samples brings on new considerations (Section 8) (Kish 1985). But dramatic effects can be seen simply for the means of small *subclasses* (e.g., as small as 0.10 or 0.01) of the entire sample, representing similar *domains* in the population (Section 5).

Each of the six levels of purposes presents different aspects for designs and each level can be fruitfully explored for more specific meanings and examples, some of which are listed in Table 2.1.

The difficulties of multipurpose designs, which have caused them to be neglected and avoided, are of several kinds. First, the different purposes must be formulated explicitly in statistical terms, so that these may serve in formulas for their comparisons and compromises; obtaining a complete list of such explicit, formal terms and maybe the principal obstacle. Second, estimates of variance and cost factors are needed for each purpose. Third, values must be obtained for and assigned to the *required* precisions for all the purposes, but this is necessary only for some methods (Section 5). Fourth, the above values and estimates must be combined in a mathematical formulation in order to arrive at the solution of a single *optimal* design to be actually used. The computational tasks of such solution have been eased by electronic computers, but the conceptual and theoretical tasks remain (Section 5).

The difficulties of these tasks help to explain why discussions of multipurpose designs are avoided in textbooks, and also in descriptions of actual surveys. Often a single statistic (e.g. the mean) of a single principal variable is presented as the only purpose for the study. In the framework of multipurpose design this is equivalent to assigning zero importance to all other purposes. The unreality of this pretense may be softened by another: that other principal purposes would result in similar allocations; but this pretense should be buttressed with calculations of the four steps above.

## 2.3. **An overall view of areas of conflict**

A brief overall view of ten areas of conflict, listed, in Table 3.1, should be useful before we look at specific problems and possible solutions. The list will probably not prove exhaustive, and readers are invited to find areas that I have overlooked. Even more likely, I hope that they will find within the ten areas other problems and other solutions that I failed to explore here.

Of this long list of ten areas of conflict fortunately not all need to be formulated for each sample design. I believe that possible conflicts about a) the sample sizes $m_g$ and about b) the relation of biases to sampling errors should always be considered, at least informally, because they are ubiquitous. Also c) allocation among domains and d) allocation among strata should receive at least a brief discussion, and often more. Computing sampling errors (j) should also be present on most surveys. On the other hand, in a continuing operation with a continuing sampling frame,

the decisions about e), f), g), and h) may have been made a long time ago for a fixed design. However, the cluster sizes (f) used in intermediate stages (blocks and segments) may be open to flexible operational changes. In the common case of one-time surveys, conflicts i) about design over time need not be considered.

It is also reassuring to know that compromises based on statistical methods can yield quite acceptable results, for several reasons. First, because moderate departures from optimal allocation result in only small or negligible increases of variance (Section 5). Curves of efficiency tend to be flat within broad areas around the optimal points; thus great accuracy in designs, which would not be feasible, are not needed. Second, however, it is also true that wide departures from optimal allocations can cause moderate to large increases in variances. Thus, ignoring important purposes can result substantial losses of efficiency for them, and therefore those purposes should be included in compromise designs. Third, compromise designs, in accord with statistical methods, can reduce drastically the losses for every purpose, with only small increases over the separate optimal designs for each purpose (Section 5).

### 2.4. **Sample sizes and bias ratios (B/$\sigma$)**

These two areas of conflict, a and b in Table 3.1, should perhaps be considered most important overall, because they can be most dramatic. We treat them together here only because they may be closely related through the effects of subclasses. Let us begin with the familiar **srs** sample size $m = S^2/V^2$ needed to yield the sample mean y, with element variance $= S^2$, for *required* precision $= V^2$. However, the $SS_g^2$ depend greatly on the variables g and the *required* $V_g^2$ may vary even more. We also include design effects $D_g^2$ that also vary, and thus $m_g = S_g^2 D_g^2/V_g^2$ expresses the sample size needed for the mean of the variable g. For the mean $\bar{y}_g$ of a domain g, comprising only the proportion $P_g$ in the population, the needed overall sample size becomes $n_g = m_g/P_g$, and it is safer to formulate the needed sampling fraction $f_g = n_g/N = S_g^2 D_g^2/V_g^2 P_g N$. The factor (1-f) may be neglected or included in $D_g^2$. The $P_g$ become most important if high precisions are *required* for small subclasses.

For comparisons of subclasses the variances increase even more: $m_g = (m_a^{-1} + m_b^{-1})^{-1} = n(P_a^{-1} + P_b^{-1})^{-1}$, with the $P_a$ and $P_b$ denoting proportions in the sample n. E.g., for the comparison of two subclass means of 0.01n and 0.10n, we have the *effective size* $m_g = n(0.01^{-1} + 0.10^{-1})^{-1} = n/110$.

For other statistics, such as medians and regression coefficients, formulating *required* sample sizes would become complex, but some numbers may be hopefully specified.

Considerations for subclass statistics become greatly modified if, in addition to variances $\sigma^2$, we also include biases $B^2$ in the Root-Mean-Square-Error = RMSE = $\sqrt{(\sigma^2 + B^2)}$ as a measure of accuracy. Figure 4.1 is meant to portray a common tendency in the accuracy of survey data, although great differences in the relations of biases to sampling errors are possible. On the horizontal axis the standard error $\sigma_1$ is shown to increase by a factor of about 3 for $\sigma_2$ of a subclass of about 1/10 of total sample. For comparisons (differences) of two such subclasses $\sigma_3$ increases by about 1.4 more.

However, the hypotenuse denoting the RMSE are shown to increase much less. In $RMSE_1$ the bias $B_1$ is shown to dominate, and this may happen for some variables in large total samples. However, the subclass $RMSE_2$, because the bias was kept constant at $B_2 = B_1$, has hardly increased at all and is dominated by $\sigma_2$. This is even more true for $RMSE_3$, where the $\sigma_3$ has increased, but the biases were assumed to tend to cancel in the difference of means, because that is a common tendency.

These relations assumed here are not mathematical but empirical and common. I propose them as practical answers to some common questions, such as: Why do we spend money for large samples and on rigorous sampling methods in the face of large measurement biases? Why bother computing sampling errors when response biases dominate the total error? The implicit answer lies in the domination of sampling errors in the subclasses, and even more in the comparisons. Let us make these implicit answers explicit in the sampling design.

### 2.5. **Allocation among domains**

This most important and frequent area of conflict has several aspects. First, consider the allo-

cation of a total sample size (or effort or cost) among the domains that constitute a partition of the total population. A common example is allocation among the several (5, 10, 20 or 50) provinces or regions or states of a country; those domains typically have very unequal populations $N_d$, with ranges of 1 to 100 in relative sizes perhaps, though they may cover roughly equal surface areas. Often the question takes this form: Should the sample sizes $d_d$ be roughly equal; or should the $n_d$ be proportional to the $N_d$, with constant sampling rates $f_{d = f?\ Equal\ nd}$ tends to yield roughly equal errors, ste($\bar{y}_d$) for the means (proportions, rates) of each region, thus tends to be near optimal for the separate domain means. On the other hand, constant $f_d = f$ tends to yield the lowest set($\bar{y}_w$), for the overall mean $\bar{y}_w = \Sigma W_d \bar{y}_d$ because it yields lower errors for the larger domains. This error may be lower than needed for $\bar{y}_w$, especially in view of potential biases (see Fig. 4), and this is the contention of proponents of equal sizes $n_d$ for provinces. However those increased errors for $\bar{y}_w$ are also suffered by most other subclasses, especially *crossclasses* like age, sex, socioeconomic classes, etc.; and those are common disadvantages of the unequal $f_d = n_d/N_d$ for provinces resulting from the equal $n_d$ values.

For example, in the Current Population Survey of the USA, larger $f_d$ are assigned to the smaller (western) states. the resulting weighting increases the variances (for a fixed total cost) of the overall means, and especially of *Black teenage boys and girls (with crucially and cruelly high unemployment rates).*

To reduce the usual confusión, I distinguish *domains* to denote partitions of the population, from *subclasses*, the corresponding partitions of the sample. Then I distinguish *design domains* (and subclasses) to refer to partitions (like provinces and regions) that are contained in strata defined by the sample design, from crossclasses (like age, sex, occupation, income, etc.) that cut across the sample design and strata, often almost randomly. The design effects differ for these two types of subclasses (Kish 1961, 1980).

In addition, other sources of conflict may arise from domain differences in the distribution of variables, also from differences either in variances $S_d^2$ or in *required* precisions. But we need not enter into those complexities here. Beyond calling attention to the problems, we must refer to two distinct technical methods for the joint solution of the conflicts in allocation, (the fourth step noted at the end of Section 2).

One approach uses iterative nonlinear programming in order to satisfy for **minimal cost** the *required* precisions **jointly for all** stated purposes. These elegant solutions to diverse problems exploit modern computers and have been published in many articles since 1963 (see reviews in Bean and Burmeister 1978, Rodriguez-Vera 1982, Cochran 1977, 5A.3-.6). The *required minimal* cost often turns out much too high, because the *required* precisions were unrealistic. Then the solutions are drastically rescaled downwards. But such rescaling exposes the false pretensions (in my view) of this elegant approach that depends on unrealistic *required* precisions. I also disbelieve in the reality of *step functions* for *required* precisions that assign a constant value to any variance below the required $V^2$ and zero value to variances above it.

A very different approach calls for some form of averaging between all the *optimal* (preferred) allocations for various purposes, by minimizing the combined (weighted) variance either for fixed cost or fixed sample size. Of course, if the resulting combined variances turn out to be too high (or low), the solutions can be scaled up (or down) in total fixed cost or sample size. I prefer this solution, which compromises between different allocations, each of which would optimize for only one purpose. It involves assigning relative values of imortance $I_g$ to all the listed statistics and this may seem difficult (though an *ignorant* decision-maker can assign equal $I_g$ to all of them). But the other two alternatives are more extreme and they should be even more difficult: either to specify the *required* precisions of all statistics for the first approach, which then assigns arbitrarily equal weights of importance to all of them; or to specify one statistic for the total weight of one, and thus zero weights for all other statistics.

Furthermore, compromises can be shown to be generally feasible and worthwhile, because the allocations are insensitive to moderate changes of weights of importance (as is often true in statistics). After all, changing the relative importance by ratios of (e.g.,) 2 or 5 should be less drastic than assigning the total weight 1 to one variable and 0 to all others, a process that implies infinite ratios of importance.

First, we denote with $\Sigma_i V_{gi}^2/n_i$ the variance attainable for a statistic g with the allocations of sample sizes $n_i$ for the ith component of variation. Then let $1 + L_g(n_i) = (\Sigma_i V_{gi}^2/n_i)/V_g^2(\min) = \Sigma_i C_{gi}^2/n_i$ denote the ratio of increase (with the allocation $n_i$) in the variance of the gth statistic over its own minimal variance; thus $L_g(n_i)$ is the relative loss over the minimal value of 1. Accepting the relative variances $C_{gi}^2/n_i$ as the functions to be minimized is a critical decision. They seem to me more reasonable than any others that I can imagine for the functions to be combined in (5.1). For example, I prefer them to the $V_{gi}^2$ which depend on arbitrary units of measurement, which are removed by the $V_{gi}^2(\min)$. But in rare cases we may be faced with $V_{gi}^2(\min = 0$ or very small and this may make $C_{gi}^2$ wildly large and unstable; in these cases assign arbitrary values to the $C_{gi}^2$ or to the ig.

Then with the weights $I_g$ assigned for relative importance of the gth statistic for any set of allocations $n_i$ of the sample sizes,

$$\begin{aligned} 1 + L(n_i) &= \\ = \Sigma_g I_g\{1 + L_g(n_i)\} &= \\ = \Sigma_g I_g \Sigma_i C_{gi}^2/n_i &= \\ = \Sigma_i \Sigma_g I_g C_{gi}^2/n_i &= \\ = \Sigma_i Z_i^2/n_i & \qquad (5.1) \end{aligned}$$

After we merely changed the order of summation, we created the new variables $Z_i^2 = \Sigma_{gz} C_{gi}^2$. This function may be minimized to give a compromise solution for fixed total cost $\Sigma C_i n_i$. For the conflict between $n_d = n/H$ of equal sample sizes for domains versus $n_d = nW_d$ proportional to domain sizes, $W_d$ the optimal compromise allocations are found to be propotional to $\sqrt{(W_d^2 + H^{-2})}$.

Two examples in Table 5.1 illustrate the surprisingly good compromises between conflicting allocations yielded by the method of weighted averaging. First, its results on the fourth row of Table 5.1 compare very favorably with the others. The reasons for the surprisingly excellent results from the compromise come from the very broad flat surfaces for the optimal allocations, as discussed in Section 2 and shown in Table 5.2 (from Kish 1976 and Kish 1987, 7.3).

An uncommon but meaningful (negative) example was provided by the (otherwise excellent) World Fertility Surveys. They decided on roughly equal sample sizes for small and large countries, and actual sample sizes varied only within the range of 3 to 10 thousand and with no discernible correlation with population size (despite my advice). Consequently, there was a two-or three-fold increase of variance in their *main contributions to knowledge:*

*So far, the main contribution to knowledge has been to confirm the downward trend in fertility that characterized much of Asia and Latin America in the 1970's and to highlight the contrast with Africa where both fertility and the desire for large numbers of children remain high.* (Macura and Cleland 1985).

## 2.6. Allocations to strata and choice of stratifiers

Domains and strata often get confused in discussions, but the two aspects are kept distinct in practical work on designs. Domains refer to subpopulations for which separate estimates are sought, whereas strata are usually smaller partitions created for decreasing variances. For example, within provinces as domains more strata may be created to reduce province variances: but cross-domains like age, sex and economic status tend to straddle across the strata. Allocations of sample sizes to strata, though often not as crucial as allocations to domains, may be important in case of efficient disproportionate optimal allocations. The two methods of Section 5 for allocating sample sizes to domains can also be applied to allocations to strata, although the aims differ. Some of the references on nonlinear programming refer to domains and others to strata.

The presence of several survey variables and statistics among the purposes have clear implications for using more stratifying variables. Different survey variables will tend to have diverse optimal relations with the stratifiers; Then it is best to use many stratifiers, even if each stratifier is used with only few stratum divisions (categories). Multipurpose design is the best reason for multivariate stratification (Kish and Anderson 1978). It may also best justify the need for *controlled selection* methods. The choice of stratum boundaries, called *optimal stratification*, is a related to-

pic, but of less importance in this condensed presentation.

## 2.7. Cluster sizes; measures of size; retaining units

In descriptions of sample designs we read often that the design effect has been approximated with $D_g^2 = (1 + roh_g(\bar{b}_t - 1)$, where **roh** stands for a synthetic intraclass correlation of the *most important* variable g and $\bar{b}_g = n/a$, the average cluster size. This would yield the effective element variance $S_g^2 D_g^2$ and the variance $S_g^2 D_g^2/n$ for the variable g. However, we must question the contents of n and of $\bar{b}_t$. If our population consists of married women of childbearing age, they may be only 10 percent of total persons and in only 30 percent of dewellings; and they may be much less than that for some rare populations. This situation has been treated in sampling for rare traits (Kish 1965, 11.4): *Ordinarily we avoid large clusters, because of their adverse effects on the variance. But even large clusters of the entire population will yield only small clusters of a rare trait, if this is widely spread. For example, entire blocks may be sampled for persons over 65 years of age; entire villages may be searched for persons with an identifiable disease. If, on the contrary, the trait is concentrated in small areas, those areas often can be recognized and stratified accordingly.*

In multipurpose designs, the crossclasses of the sample will be of variable sizes $m_g$ sizes of the total sample size $n_t$, with $\bar{M}_g$ as their different proportions in the populations. Thus we want to estimate in the design not only $\{1 + roh_g(\bar{b}_t - 1)\}$ for diverse variables g for the total sample $n_t$, but also $\{1 + roh_g(\bar{b}_g - 1)\}$ for many crossclasses. Then we make use of some conjectures that have been shown to be good approximations in thousands of empirical computations:

$$\{1 + roh_g(\bar{b}_g - 1)\} = \{1 + roh_g(\bar{M}_g\bar{b}_t - 1)\} = \{1 + roh_t(\bar{M}_g\bar{b}_t - 1)\} \qquad (7.1)$$

That is, we use $\bar{b}_g = \bar{M}_g\bar{b}_t$ and $roh_g \simeq roh_t$ as rough approximations. Although this somewhat underestimates the average values of $D_g^2$ for crossclasses, because of variations in cluster sizes of crossclasses, that is a small factor compared to the large variations of $roh_g$ between variables (Kish 1987, 7.1; Verma et al. 1980, Kish et al. 1976). That underestimate has small effects on the efficiency of the design. But it is important to consider efficiencies of estimates for subclasses as well as for the entire sample, considerations and tasks which are often neglected. This will tend strongly toward higher efficiencies for larger clusters than would be shown for $\bar{b}_t$ and n for the total sample only.

Measures of size are related to cluster sizes, but are not entirely the same because of errors in the available measures, due especially to different population contents and to obsolescence. This is also a good time to note problems concerning measures of size for *integrated survey operations* for different populations which may especially need drastic compromises. For example, consider integrated designs for total populations and for agriculture; also perhaps for ethnic subpopulations; also perhaps for industrial or business activities. The measures of size for each of these may differ greatly. Yet some compromise solution may be feasible that would yield reasonable efficiencies for each. The asymptotic validity of using selection rates is unaffected, of course, by those differences.

Measures of size are also closely related to problems for *Retaining units after changing strata and probabilities* (Kish and Scott 1971). Those methods were designed to deal with changes over time for measures of size and of strata, but they are also relevant for changes between variables. *Unequal selection probabilities are often assigned to sampling units. Our methods, though more generally applicable, are especially needed for the selection of primary sampling units for surveys. Often these are selected separately from many strata, with one selection from each stratum.*

*After the initial selection the units may be used for many surveys over several years. But as time passes, the needs of new surveys may be better served by new strata and new selection probabilities, based on new data, than by those used for the initial selection. The difference between initial and new data may be due to differential changes among the sampling units as revealed by the latest Census. Or the differences may be due to changes in survey objetives and populations; for example, a sample initially designed for households and persons may later be required to serve a survey of farmers, or college students.* ***Obviously our methods are also applicable to designing si-***

***multaneously a related group of samples with differing objectives.***

This method would allow for using the best measures (for size and for strata) separately for each sample purpose, but maximizing the retention of the overlap of sampling units between the samples for separate purpose (especially PSU's). At another extreme it would be possible to design a compromise that would average the measures in order to achieve a complete overlap of units, but sacrificing some efficiency for each of the purposes. A compromise between the two extremes may be even better than either: increase the overlap with small sacrifices of separate efficiencies by recognizing only differences of measures that surpass some arbitrary minimal criteria (Kish and Scott 1971, 3a).

## 2.8. **Purposes and designs for periodic studies**

Periodic studies provide areas of conflict with great and growing importance as their numbers and sizes increase. It is untrue that those expensive and influential surveys have only one of the five purposes listed in Table 8.1, because usually they are needed for several or all of them would be if the design permitted them.

In Table 8.1. we note five purposes and six designs. The first four are paired with letters on the same four lines. These pairings call attention to designs that best serve, with reduced variances, each of the four purposes. Most periodic studies have several purposes and thus we should face —not necessarily solve— the difficult problems of multipurpose designs. Actually current levels (A) and net changes (C) can be served with any of the six listed designs, even if with some increase in variances or in costs. But individual (gross, micro) changes (D) need panels, and cumulations (B) usually need some changes. Often reasonable compromises become possible, when purpose can be defined. Furthermore, extraneous considerations may rule out some designs (e.g., overlaps may be either prohibited or enforced) and thus force the use of less efficient —but still valid— designs. The chief variation in these six designs concerns the amount (and kind) of overlaps between periods. The rotation scheme of complete overlaps shows, with aaa-aaa, that the periods have all common parts; the nonoverlap with aaa-bbb shows none; and the partial overlap abc-cde-efg shows c and e as 1/3 overlaps between succeeding periods only.

This section concentrates on the effects of varying proportions of overlaps P in diverse designs on different purposes; in complete overlaps P = 1, in nonoverlaps P = 0, and in partial overlaps $0<P<1$. The purposes are discussed in terms of variances for estimated means, because means (and percentages, rates, proportions) are both the most used and the simplest estimates to treat. Effects on other estimates will not be entirely different but they are too many, diverse, and difficult to be explored here.

For more details consult references (e.g., Kish 1987, Section 6.2 or 1965 Sections 12.4-.5). More discussion of panels is also available there, with its advantages, disadvantages, problems and solutions. I call attention to SPD, or Split Panel Designs that I am trying to promote as multipurpose designs. These would combine a panel sample P with new rotating of *rolling* samples, so that Pa-Pb-Pc-Pd would symbolize the periodic samples. The a, b, c, d, etc., could be cumulated into larger samples. The panel P would serve as the partial overlap for better estimates of both current levels and macro (mean, net) changes ***for any pair of periods***. The panel serves primarily to provide micro (individual gross changes).

## 2.9. **Computing and presenting sampling errors**

It seems questionable to include this topic under design, but I have no doubt that it is a multipurpose problem. The strategies for computing and presenting sampling errors deserve separate listing as an area of conflict among the many statistics given generally for the results of surveys. It is not enough to present standard errors for only one or a few of the most important statistics: they are too many and too diverse. Because of that diversity, the practice has grown up to compute from the variances other functions of sampling errors, especially estimates of the *design effects* $d_g^2$; also sometimes for the $d_g^2 = 1 + roh_g(\bar{b}_g - 1)$ also estimates of the synthetic intraclass correlation $roh_g$.

Briefly, I advise: a) Compute sampling errors for many variables, because the variances, the design effects ($d_g^2$) and the intraclass coefficients

($roh_g$) can and do differ greatly between variables. b) You may have to do some averaging of sampling errors, because it may be inconvenient or confusing to present them all. c) It may be neither feasible nor necessary to compute sampling errors for all subclasses, because they can often be approximated with reasonable models. d) It is necessary to present sampling errors for subclasses and for other statistics to guide the readers of the reports (Kish 1965, 14.1-2; Kish 1987, 7.1; Verma et al 1980). I hope that this topic will receive in the future from theorists and methodologists some of the attention it lacks and richly deserves.

## 2.10. **Conclusions**

The solutions to the ten areas of conflicts of Section 3 that I propose in Sections 4 to 9 have not been uniform. Averaging allocations among domains in Section 5 seems to give surprisingly good compromise solutions. The chief advice in Section 6 is to use more stratifiers. In Sections 4 and 8 I propose the joint consideration of all the several important purposes for sample designs. We looked at the different levels of purposes and at the various areas of conflicts jointly.

Asking the right question is the core of most problems. I propose multipurpose design as a new paradigm, to replace *optimal* solutions to artificially partial questions such as: What is the optimal allocation for the mean $\hat{y}$ of the most important variable?

TABLE 2.1

**Hierarchy of purposes**

1. Diverse statistics from the same variable
   - — Totals or means or medians and quantiles, distributions.
   - — Analytical statistics: regressions, categorical analysis.
   - — Time aspects: static, macro-change, micro-change, cumulative.
2. Diverse populations and domains (subclasses)
   - — Proper classes and crossclasses.
   - — Comparisons of subclasses.
3. Multiple variables on the same subject
   - — Several measures of one variable; e.g. of income, or unemployment.
   - — Diverse periods - per day, week, month, year.
   - — Several aspects of one subject: income, savings, wealth.
4. Multisubject surveys
   - — Several subjects on same schedule, interview, operation.
   - — Health surveys of many diseases.
   - — Market research for several clients, many goods.
   - — Agricultural surveys of many crops.
   - — «Omnibus» social surveys.
5. Continuing, integrated survey operations
   - — NSS in India, CPA in USA, NHSCP of UN.
   - — Separate surveys from one office and field staff.
   - — Common source of surveys.
   - — Diverse methods, costs, operations, allocations, respondents.
6. Master frames
   - — Several samples from one frame or set of listings.
   - — Separate institutions, organizations.
   - — Separate field staffs? Same PSU's?

TABLE 3.1

**Areas of conflicts**

a. Sizes $m_g$ and rates $f_g$ of needed samples for purposes g

$$m_g = S_g^2 D_g^2 / V_g^2 \text{ for } n_g = m_g / P_g \text{ or } f_g =$$

$$= S_g^2 D_g^2 / V_g^2 P_g N$$

b. Relation of biases to sampling errors in RMSE $= \sqrt{(\sigma^2 + B^2}$

— The bias ratio $B/\sigma$ decreases as $\sigma$ increases for subclasses.

— For comparisons $B/\sigma$ tends to vanish as B decreases, $\sigma$ increases.

c. Allocation of the $m_g$ among domains

$$m_t = \Sigma_g m_g$$

Computation and presentation of sampling errors.

d. Allocation of $m_{gh}$ among strata h

$$m_g = \Sigma_h m_{gh}$$

e. Choice of variables for stratifications
Multivariate stratification

f. Optimal cluster sizes

$$D_g^2 = \{1 + roh_g(\bar{b}_g - 1)\}\ \bar{b}_g =$$

$$P_g n_t / a$$

for crossclasses

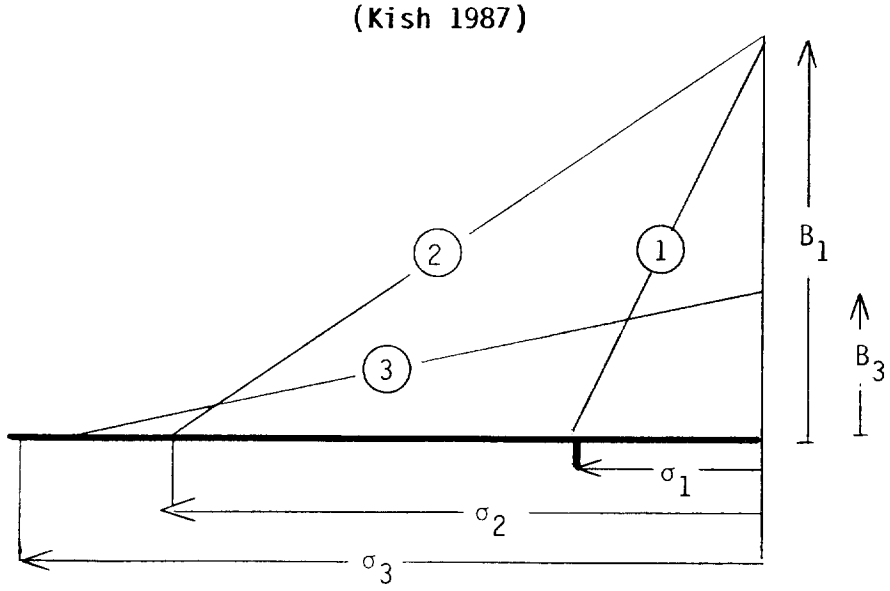
g. Measures for cluster sizes

h. Retaining sampling units (PSU's) for changed subjects, measures and strata.

i. Design over time
How much overlap? Panels? Change versus cumulation.

j. Computing and presenting sampling errors

**FIG. 4.1.**
**VARIABLE ERRORS ($\sigma$) AND BIASES (b) IN ROOT-MEAN-SQUARE ERRORS (RMSE)**




The bases represent sampling errors and other variable errors ($\sigma$). For example $\sigma_1$ may be the ste($\bar{y}_t$) for the mean y of the entire sample and $\sigma_2$ may be a larger ste($\bar{y}_c$) for a subclass mean, and $\sigma_3$ may be the ste($\bar{y}_c - \bar{y}_b$) for the difference between two subclass means.

The heights represent biases (B) and the hypotenuse denotes the RMSE = $\sqrt{(\sigma + B2)}$; (see 7.2F). 1) For the entire sample the bias $B_2$ may be large compared to the variable error $\sigma_1$, thus taking larger samples would not decrease the $RMSE_t$ by much. 2) However, with the same bias $B_1$, but

with a smaller sample in the subclass, the ratio changes and the $\sigma_2$ dominates the $RMSE_2$; and this is not much larger than for 1) despite a much smaller sample. 3) Furthermore, for the difference of means, the net bias $B_3$ may be much smaller; so that even with a larger $\sigma_3$, the $RMSE_3$ for the difference is but little greater than $RMSE_2$. This drastic change in the bias ratio $B/\sigma$ tends to appear not only for differences between subclasses within the same sample, but also for differences between repeated surveys.

**TABLE 5.1.**
**LOSS FUNCTIONS (1 + L) FOR TWO POPULATIONS**
**(Kish 1976)**

| Allocations $m_i$ | (A) (1 + L) for $W_1/W_2 = 4$ | | | (1 + l) for 133 countries: 0.2 to 100 mm | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | Joint with weights | |
| | $\Sigma W_i\bar{y}_i$ | $\Sigma\bar{y}_i/2$ | Joint | $\Sigma W_i\bar{y}_i$ | $\Sigma\bar{y}_i/133$ | 1 : 1 | $I_c/I_d$ : 1 |
| | 1 | 1·56 | 1·28 | 1 | 6·86 | 3·93 | |
| $mW_i$ | 1·36 | 1 | 1·18 | 3·34 | 1 | 2·17 | |
| m/H | 1·08 | 1·125 | 1·102 | 1·35 | 1·54 | 1·44 | |
| $\propto\sqrt{W_i}$ | 1·116 | 1·080 | 1·098 | 1·31 | 1·28 | 1·295 | |
| $\propto\sqrt{(W_i^2 +$ | | | | 1·47 | 1·17 | (1·32) | 1·27 |
| $H^{-2}$ | | | | 1·20 | 1·44 | (1·32) | 1·28 |
| $\propto\sqrt{(0{\cdot}5W_i^2 +$ | | | | 1·12 | 1·66 | (1·39) | 1·23 |
| $H^{-2})$ | | | | | | | |
| $\propto\sqrt{2W_i^2 +$ | | | | | | | |
| $H^{-2})$ | | | | | | | |
| $\propto\sqrt{(4W_i^2 +$ | | | | | | | |
| $H^{-2})$ | | | | | | | |

In (A) there are two strata and domains ($W_1 = 0{\cdot}8$ and $W_2 = 0{\cdot}2$); note that the allocation $m_i = \sqrt{W_i}$ does most as well for the joint loss as the optimal.

In (B) we have the populations of 133 countries, ranging in size from 0·2 to over 100 millions, a range of 500 in relative sizes. From this problem of allocation (for the World Fertility Survey) we omitted, for practical reasons, the four largest countries and a few under 0·2 millions. Their inclusions would raise the variance of relative sizes, $W_i$, from 2·5 to 12, and would make the results more dramatic. Note that the $\sqrt{W_i}$ allocation reduces losses quite well. Some compromise is better than none. But the optimal allocation, $\sqrt{(W_i^2 = H^{-2})}$, is considerably better. Different values of $I_c/I_d$ ( = 1/2, 2/1 and 4/1) increase slightly the variance of the joint loss function with (1 : 1) weights; but they remain steady for joint loss functions with their own weights $I_c/I_d$ : 1.

**TABLE 5.2.**
**RELATIVE LOSSES (L) FOR SIX MODELS OF POPULATION WEIGHTS ($U_i$); FOR DISCRETE ($L_d$) AND CONTINOUS ($L_c$) WEIGHTS: FOR RELATIVE DEPARTURES ($k_i$) IN THE RANGE FROM 1 TO K**
**(Kish 1976)**

| Models | K | 1·3 | 1·5 | 2 | 3 | 4 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Dichotomous U (1 − U) | | | | | | | | | |
| (0·5) (0·5) | | 0·017 | 0·042 | 0·125 | 0·333 | 0·562 | 0·800 | 2·025 | 4·512 |
| (0·2) (0·8) | | 0·011 | 0·027 | 0·080 | 0·213 | 0·360 | 0·512 | 1·296 | 2·888 |
| (0·1) (0·9) | | 0·006 | 0·015 | 0·045 | 0·120 | 0·202 | 0·288 | 0·729 | 1·624 |
| Rectangular | $L_d$ | 0·017* | 0·042* | 0·125* | 0·222 | 0·302 | 0·370 | 0·611 | 0·889 |
| $U_i \propto 1/K$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·099 | 0·155 | 0·207 | 0·407 | 0·656 |
| Linear decrease | $L_d$ | 0·017* | 0·040* | 0·111* | 0·203 | 0·283 | 0·353 | 0·616 | 0·940 |
| $U_i \propto K+1-K_i$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·097 | 0·153 | 0·205 | 0·409 | 0·680 |
| Hyperbolic decrease | $L_d$ | 0·017* | 0·040* | 0·111* | 0·215 | 0·312 | 0·404 | 0·807 | 1·466 |
| $U_i \propto 1/k_i$ | $L_c$ | 0·006 | 0·014 | 0·041 | 0·103 | 0·171 | 0·235 | 0·528 | 0·011 |
| Quadratic decrease | $L_d$ | 0·016* | 0·036* | 0·080* | 0·150 | 0·211 | 0·264 | 0·460 | 0·696 |
| $U_i \propto 1/k_i^2$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·099 | 0·155 | 0·207 | 0·407 | 0·656 |
| Linear increase | $L_d$ | 0·017* | 0·040* | 0·111* | 0·167 | 0·200 | 0·222 | 0·273 | 0·302 |
| $U_i \propto k_i$ | $L_c$ | 0·006 | 0·013 | 0·037 | 0·088 | 0·120 | 0·148 | 0·223 | 0·273 |

Dichotomous $1 + L = U(1 - U)(K - 1)^2/K$, also discrete with*

Discrete $1 + L_d = (\Sigma U_i k_i)(\Sigma U_i/k_i)$, with $K_i = i = 1, 2, 3, \ldots, K$

Continuous $1 + L_c = \geqq Uk.dk. \geqq (U/k)dk$, with $1 \leqq k \leqq K$

Only two values, 1 and K, were used for $L_d$ for K = 1·3, 1·5 and 2

**TABLE 8.1.**
**PURPOSES AND DESIGN FOR PERIODIC SAMPLES**

| Purposes | Designs | Rotation scheme |
|---|---|---|
| A. Current levels | A. Partial overlaps $0 < P < 1$ | abc - cde - efg |
| B. Cumulations | B. Nonoverlaps $P = 0$ | aaa - bbb - ccc |
| C. Net changes (means) | C. Complete overlaps $P = 1$ | aaa - aaa - aaa |
| D. Gross changes (individual) | D. Panels | Same elements |
| E. Multipurpose time series | E. Combinations, SPD | |
| | F. Master frames | |

**TABLE 8.2.**
**EFFECTS OF DIVERSE OVERLAPS ON VARIANCES OF DIFFERENCES OF MEANS ASSUMES $S_x^2 = S_y^2 = S^2$ AND Srs or Deff = 1**

| Design types | Sample sizes for specific designs | Effects on $S^2/n$ |
|---|---|---|
| A. Partial overlap | $n = n_x = n_y$, $n_c = P_n$ | $2 (1 - PR)$ |
| B. Nonoverlap | $n = n_x = n_y$, $n_c = P = 0$ | 2 |
| C. Complete overlap | $n = n_x = n_y$, $n_c = P = 1$ | $2 (1 - R)$ |
| D. Subset | $n = n_x = n_y$, $n_c = P_n$ | $(1/P + 1 - 2R)$ |

# III. THIRTEEN BASIC ASPECTS OF SURVEY SAMPLING

The thirteen aspects which are shown in this chapter were dealt with more extensively in the seminar corresponding to this publication

## 3.1. Sample design and survey design

### Sampling design

Selection methods and procedures.
Allocation to strata.
Choice of sampling units.
Allocation to clusters.
Stratification.

### Joint design

Estimation.
Statistical analysis.
Sampling errors.
Nonresponse, noncoverage, imputation.
Population and elements.

### Survey design

Survey variables, substantive analysis.
Measurements, observations.
Response errors and biases.
Choice of domains of analysis.
Presentation of data and results.
Utilization of results.

## 3.2. Definition of population, elements and survey units

Definition of population and elements.
Extent and units of analysis.
Domains = subpopulations of analysis.
Design domains and crossclasses.

Observational units.
Respondents, units of observation.

Sampling units.
Hierarchical nestling, identification in multistage sampling. A) $PSU_s$ – B.–C.–...– Elements.

Populations: sampled → frame → target → inferential.

Probability samples underlie the achieved Survey population, but two discrepancies come even between them: sampling errors and item nonresponses. Both of these differ greatly among variables and the amount of item nonresponse is shown as differing greatly among variables. For both of these discrepancies the sample responses serve as bases; sampling errors are computed from them; and they are used for «imputing» or weighting for item nonresponses.

Thus probability samples are shown as a broad and solid foundation for the Survey population, on which to build the structure of the inference above it. For the discrepancies beyond the Survey population one must go beyond the sample data, with the help of implicit or explicit data. The span to the Frame population is due to *total nonresponses* of diverse kinds (refusals, not-at-homes etc.); the size of nonresponses may be estimated from sample records (with effort and care), but estimating their effects needs models and auxiliary data. The size of *noncoverage* can only be estimated with models or from checks with outside sources, yet this portion also belongs to the Target population. This may also include a defined and deliberate *exclusion* from the coverage.

Furthermore, sample data are also used for inferences beyond the Target populations and these are many, various and ill-defined. «Superpopulations» of sampling theory are not only among these, but behind all these inferential populations. These model-dependent inferences are too often merely implicit. Even more vagueness describes

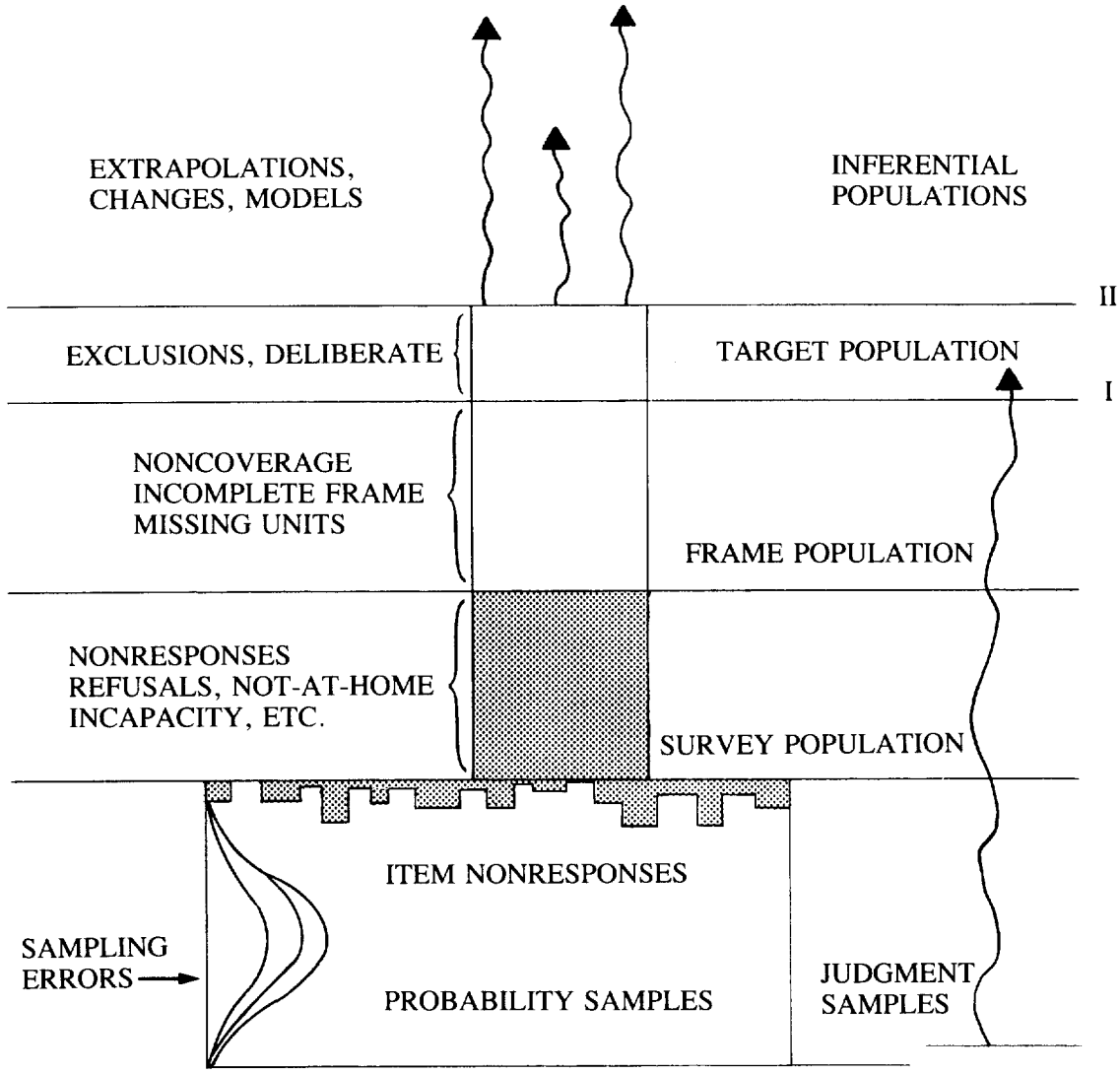



**FIG. 2.1.1**
**DISCREPANCIES BETWEEN FOUR POPULATIONS (from «Statistical Design for Research» by L. Kish, Wiley, 1987)**

the path of judgment samples directly to the Target population; and such vagueness is indicated by the thin, wavy, population line, as for the extrapolations to vague inferential populations.

### 3.3. **Probability selection**

Implications.
Statistical inference.
Measurability.
Probability selection.
Mechanical selection, random numbers.
Lists or frames.

### **Alternatives to probability sampling**

Haphazard, convenient, fortuitous.
Judgmental selection.
Population dependent.
Quota sampling.
Restricted sites, community studies.

Internal replications - extreme probes.

Known (knownable) $P_i>0$ for all elements in $p_o$ assured with selection operation.

### 3.4. Equal probabilities of selection -«epsem»

Self-weighting data form epsem selections.
Constant P = f for all elements in population (frame).
Overall f achieved through unequal rates in stage.
PPS selections for clusters.

**Reasons for epsem**

Weighting is not simple.
Mistakes of man-machine systems.
Secondary analysis by distant researchers.
Complex analytical statistics.
Increased costs.
Increase of variances from haphazard weights.
Public relations easier with epsem.

### 3.5. Fix sampling rates (f), not sample size (n)

Often N unknown and N' variable.
Then n' = fN' with f fixed and n' variable is **better** than f' = N'/n with n fixed and f' variable.

Examples: total sample size n fixed
fixed b from clusters (e.g.: blocks) with variable $N_i$.

Disadvantages of variable f'.

Weights increase costs, variances, complexity.
When $b>N_i$.
Field selection of $b/N_i$ difficult.
Exact counts of $N_i$ in field.

### 3.6. Attach weights to cases = elements

Compute element (case) weights $w_i$ from all sources, all stages of selection, all adjustments.

Use in all moments: $\Sigma w_j y_j \quad \Sigma w_j y_j^2 \quad \Sigma w_j y_j x_j$

also ratios $\dfrac{\Sigma w_j y_j}{\Sigma w_j} \quad \dfrac{\Sigma w_j y_j}{\Sigma w_j x_j} \quad \dfrac{\Sigma w_j y_j x_j \Sigma w_j}{\Sigma w_j y_j \Sigma w_j x_j}$

relative weights $W_j = w_j/\Sigma w_j \quad _{\Sigma w_j} = 1$

Random replication to produce self-weighting data.
Multiple replications to reduce variances.

### 3.7. Sources and effects of weights

**Sources**

a) Disproportionate allocation - large differences
Domain estimates.
Optimal allocation to strata.

b) Inequalities in frames and procedures - moderate differences - rare?

c) Adjustments for nonresponse and noncoverage-small replications for item nonresponses?

d) Statistical adjustments-ratios, poststratification.

Haphazard weights increase variances, b, c, d above? not a)?

Examples:

1) Single dwellings from buildings with 1 to 62:2

2) b = 3 from segments with 1 to 200, increase great

4) Weights h/4 for h calls for nonesponse, increase 1, 3;

5) Adults in U.S. households (telephone), increase 1.0

3) Equal samples for 19 provinces: increase of 2.

Moderate departures have small effects on variances.
Large departures have moderate to large effects.
Increases tend to be inherited in crossclasses.
Multipurpose allocation may be close to epsem.

Increase $= (\Sigma w_i k_i)\,(\Sigma w_i/k_i) = (\Sigma w_i k_i)^2/(\Sigma w_i)^2 = 1 + C_R^2$

**TABLE 1**
**RATIO OF VARIANCE OF THE MEAN DUE TO DEPARTURES FROM OPTIMAL ALLOCATION. FOR RELATIVE DEPARTURES ($k_i$) IN THE RANGE 1 TO K. FOR RECTANGULAR DISTRIBUTION OF THE POPULATION FREQUENCY $U_i$ IN THE POPULATION. WITH $k_i$ = 1, 2, 3, ... K FOR DISCRETE, AND 1 < k < K FOR CONTINUOUS**

| K | 1.3 | 1.5 | 2 | 3 | 4 | 5 | 10 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Discrete | 1.017 | 1.042 | 1.125 | 1.22 | 1.30 | 1.37 | 1.61 | 1.89 | 2.30 | 2.62 |
| Continuous | 1.006 | 1.014 | 1.040 | 1.10 | 1.16 | 1.21 | 1.41 | 1.66 | 2.14 | 2.35 |

### 3.8. **Frame problems**

Most important skill **craft** (act) of survey samples.

We want listing = element L-E or L-U U = Sampling Unit.

**Imperfections**

L-O Blanks and foreign elements non-members, non-members of population, or of subclass, nonresponse.

L-E-L Duplicate, replicate, multiple listings, dual multiple frames.

E-L-E Small clusters of elements.

O-L Missing elements, incomplete frame, noncoverage (O-L)-(L-O) = **net** undercoverage.

Avoid problems?

1) Correct lists.
2) Redefine population.
3) Ignore small problem.

Specific procedures for maintaining f for the E's. Also ordinary mistakes for each.

### 3.9. **Design of sample sizes**

**Costs** are real constraints in practice.
Variances too many and diverse.
Purposes too many and ill-defined.
Field collection costs and methods are decisive.

Actual decision:

$$C = C_o + cn \rightarrow n = (C - C_o)/c \rightarrow Var(\bar{y}_g) = S_g D_g^2/n$$

$D_g^2$ = design effect

Theoretical: $n = S^2/Var(\bar{y})$

**Multipurposes**

Different statistics mean, median, corr coefficient.
Different variables.
Different subjects.
Different domains.

**Domains** in population reflected by subclasses in sample.
Design domains: provinces, states, urban, rural.
Crossclasses: sex, age, education, occupation, etc.
Major domains: 5 or 10, well represented by most samples.
Minor domains: 20 to 100 poorly represented by most samples.
Mini domains: 200 - 3000 not represented by most samples.

### 3.10. **Departures from SRS**

I.I.D.: Identically and independently distributed.
«given n I.I.D. random variables, all false independence a powerful tool assumption sometimes approximated by condition.
Seldom actually selected.
In survey sampling lack of independence.

**Descriptive** statistics depend only on $p_j$ of probabilities sample.
Include $\hat{Y}$, $\bar{y}$, $s^2$, r, b, etc.
Robust for complex selections $E(\bar{y}) \simeq \bar{Y}$ $E(s^2) = S^2$, etc.

**Inferential** statistics, sampling errors, conf. intervals.

Depend on selection methods.

Are sensitive to departures from srs.

Depend on:variables estimation methods (statistics selection methods)

$$\text{design effects} = \text{deft}^2 = \frac{\text{actual variance}}{\text{srs variance}}$$

### 3.11. Stratified element sampling

srs selection within strata = subpopulations.

Proportionate, random element sampling (pres).

$f_h = n_h/N_h = f = n/N \Rightarrow n_h/n = N_h/N$

same sampling rates $\Rightarrow$ sample in «miniature» of pop.

easy to do with systematic selection through strata common, easy, understandable.

$\text{deft}^2 < 1$, but little

for crossclasses $\bar{M}_c$ $\text{deft}^2 \to 1$ with $\bar{M}_c$

for differences $(\bar{y}_c - \bar{y}_b)$ $\text{deft}^2 \simeq 1$

and analytical statistics.

Disproportionate (optimal) allocation

$$f_h = \frac{n_h}{N_h} \propto \frac{S_h}{\sqrt{C_h}}$$

### 3.12. Cluster sampling

Sampling units are clusters of elements.

Clusters are usually stratified for selection efficient.

Selecting compact clusters simplest.

When cluster sizes are small and equal.

Subsampling when clusters are large and unequal.

Two stage sampling: subsampling of selected PSU's.

Multistage sampling.

Selections in stages from hierarchical subdivisions.

PPS with MOS to create approx equal subsample sizes from grossly unequal clusters, for fixed ultimate f.

$\text{deft}^2 >$ often $\text{deft}^2 >> 1$

$\text{deft}^2 = 1 + \text{roh}\,(\bar{b} - 1)$ where $\bar{b}$ is subsample size.

$\text{deft}^2 = 1 + \text{roh}_s\,(\bar{M}_s b - 1)$ decreases with $\bar{M}_s$ for crossclasses.

$\text{deft}^2$ is even closer to 1 fo $(\bar{y}_s - \bar{y}_t')$.

$\text{deft}^2$ is also lower for analytical statistics r and B.

but it exists $> 1$ for all sampling errors.

**Area sampling** is procedure for *identification* of elements with sampling units to construct frames for selection on hierarchical sampling basis.

Households, farms, etc., can be identified with area segments.

### 3.13. Computing and presenting sampling errors

**Measurability** almost as important as probability selection.

**Sampling errors** include variances and functions.
design effects = act var/srs var
$\text{roh} = (\text{deft}^2 - 1)/(\bar{b} - 1)$ $CV^2$ (X) for stability

**Pooling** and generalizing needed because:

1) lack of precision, not enough PSU, esp. in design domains.
2) too many statistics in multipurpose surveys.
3) use for other surveys and designs.

**Measurability** of variances *needs only*
Ultimate clusters
identification for all cases of PSU's and strata

**Strategy** for computations

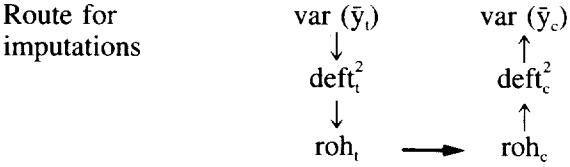
**Compute** $\text{var}(\bar{y}_t)$ for *many* variables
$\text{deft}^2(\bar{y}_t) = \text{var}(\bar{y}_t)/\text{srs var}$ for *many*
*Average* $\bar{y}_t$ for classes of variables?

Compute or impute $\text{var}(\bar{y}_c)$ and $\text{deft}^2(\bar{y}_c)$ for crossclasses

Compute or impute $\text{var}(\bar{y}_d)$ and $\text{deft}^2(\bar{y}_d)$ for design classes

Compute or impute var and deft for analytical statistics

Route for imputations




Presentations in tables of sampling errors for classes standard errors.

# BIBLIOGRAPHY


KING, A.J. and JESSEN, R.J. (1945) The master sample of agriculture, JASA, 38-56.

KISH, L. A Procedure for Objective Respondent Selection within the Household, *Journal of the American Statistical Association*, 44, (September, 1949), 380-387).

KISH, L. and LANSING, J.B.; DENT, J., and KATONA, G., Methods of the Survey of Consumer Finances, Federal Reserve Bulletin, 36 (July 1950), 759-809.

GOODMAN, R. and KISH, L., Controlled Selection - A Technique in Probability Sampling, *Journal of the American Association*, 45, (September, 1950), 350-372.

KISH, L., A Two-Stage Sample of a City, *American Sociological Review*, 17, (December, 1952), 761-769.

KISH, L. On the Differentiation of Ecological Units, *Ph. D. Thesis*, The University of Michigan, (1952).

KISH, L. Selection of the Sample, *Book*. Chapter 5 in Festinger and Katz (eds.), *Research Methods in the Behavioral Sciences*, New York: Dryden Press (now Holt, Rinehart and Winston), (1953), 175-240.

KISH, L. Differentiation in Metropolitan Areas, *American Sociological Review*, 19, (August, 1954), 388-398. Also in J.P. Gibbs (ed.) Urban Research Methods, New York: Van Nostrand (1960), 309-328. S-435 in the *Bobbs-Merrill Reprint Series in the* Social Sciences.

KISH, L. and LANSING, J.B., Response Error in Estimating the Value of Homes, *Journal of the American Statistical Association*, 49 (September, 1954), 520-538.

KISH, L. Confidence Intervals for Clustered Samples, *American Sociological Review*, 22, (April, 1957), 1954-1965.

LANSING, J.B., and KISH, L., Family Life Cycle as an Independent Variable, *American Sociological Review*, 22, (October, 1957), 512-519. Also in Martin, W.T., Schrag, C.C. and O'Brien, R.W., (eds.), Readings in Sociology, Boston: Houghton Mifflin Co. (1964). Also in Newman, J.W. and Boyd, H. (eds.), Advertising Mangement, Homewood, I11.: R.D. Irwin, Inc. (1965), also Britt, S.H. and Boyd, H.W., Marketing Management, McGraw-Hill, 1968.

KISH, L. and HESS, I., On Noncoverage of Sample Dwellings, *Journal of the American Statistical Association*, 53, (June, 1958), 509-524.

KISH, L. Some Statistical Problems in Research Design *American Sociological Review*, 24, (June, 1959), 328-338. S-436 in the *Boss-Merrill Reprint Series in the Social Sciences*. Also reprinted as «Problemas Estadísticos del Diseño de Investigaciones», *Estadística*, 21, (March, 1963), 52-57. In Tufte FR, *The Quantitative Analysis of Social Problems*, Addison-Wesley, 1970, *Warner Modular Publications*. Several others.

KISH, L. and HESS, I., A «Replacement» Procedure for Reducing the Bias of Nonresponse, *The American Statistician*, 13 (October 1959), 17-19.

KISH, L. and HESS, I., Some Sampling Techniques for Continuing Survey Operations, *Proceedings of the Social Statistics Section*, American Statistical Association, (1959), 139-143.

KISH, L., SLATER, C.W., Two Studies of Interviewer Variance of Socio-Psychological Variables, *Proceedings of the Social Statistics Section*, American Statistical Association: Stanford, California, (1960), 66-70.

KISH, L. (1961) Efficient allocation for multi-purpose samples, *Econometrica*, 29, 363-385.

KISH, L. A measurement of Homogeneity in Areal Units, *Bulletin of the International Statistical Institute*, 33rd Session, (1961) Vol. 4, 201-209.

KISH, L. The Design of Sample Surveys, Chapter 2 in Collier, P.O. and Elam, S.N. (eds.) *Research Design and Analysis*, Bloomington, Indiana: Phi Delta Kappa (1961), 45064. (offprints available).

KISH, L., LOVEJOY, W., and RACKOW, P., A Multi-Stage Probability Sample for Continuous Traffic Surveys, *Proceedings of the Social Statistics Section*, American Statistical Association, (1961), 227-230.

KISH, L. Studies of Interviewer Variance for Attitudinal Variables, *Journal of the American Statistical Association*, 57, (March 1962), 92-115.

KISH, L., NAMBOODIRI, N.K., and PILAI, R.K., The Ratio Bias in Surveys, *Journal of the American*

*Statistical Association*, 57, (December, 1962), 92-115.

KISH, L. Changing Strata and Selection Probabilities, *Proceedings of the Social Statistics Section*, American Statistical Association, (1963), 124-131.

KISH, L. Generalizations for Complex Probality Sampling, *Proceedings of the Social Statistics Section*, American Statistical Association, (1964), 63-67.

KISH, L. (1965). Survey Sampling, New York and Sydney: John Wiley and Sons.

KISH, L. Selection Techniques for Rare Traits, Chapter in *Genetics, Epidemiology, and Chronic Diseases*, Public Health Service Publication, No. 1173 (February, 1965). (offprints available).

KISH, L. Sampling Organizations and Groups of Unequal Sizes, *American Sociological Review*, 20, (August, 1965), 564-572. Also in *Bobbs - Merrill Reprint Series in the Social Sciences* (S-595).

KISH, L. *Survey Sampling*, New York and London: John Wiley and Sons, (1965), XVI, 635 pages.

KISH, L., and HESS, I., *The Survey Research Center's National Sample of Dwellings*, Ann Arbor: Institute for Social Research, (1965), 63 pages (multilith, $1.00).

KISH, L. «Standard Errors for Indexes from Complex Samples», *Journal of the American Statistical Association*, (June, 1968) 63, 512-529.

REIFF, G., KISH, L., and HARTER, J., «Selecting a Probability Sample of School Children and in the Coterminous United States», *The Research Quarterly*, AAHPER, (May, 1968), 39, 409-414.

MURTHY, M.N. (1967) *Sampling Theory and Methods*, Calcutta: Statistical Publishing Society, Sections 9.11, 14.13, 15.1e.

KISH, L. «Design and Estimation for Subclasses, Comparisons, and Analytical Statistics», in Johnson, N.L. and Smith, H. (eds.) *New Developments in Survey Sampling*, New York: John Wiley and Sons, (1969).

KISH, L., and FRANKEL, M. «Balanced Repeated Replications for Standard Errors», *Journal of the American Statistical Association* (September, 1970), 65, 1071-1094.

KISH, L. and SCOTT, A., «Retaining Units After Changing Strata and Probabilities», *Journal of the American Statistical Association*, (September, 1971), 66, 461-470.

KISH, L. «Multipurpose Programs for Sampling Errors», *Proceedings of the International Statistical Institute*, 1971, Vol. II, 216-221.

KISH, L. and SCOTT, A.J. (1971) Retaining units after changing strata and probabilities, JASA, 66, 461-470.

KISH, L. Muestreo de Encuestas, México: Editorial Trillas, Spanish translation of *Survey Sampling* by J. Nieto de Pascual, 1972, 739 pages.

KISH, L. FRANKEL, M.R., and VAN ECK, M. *SEPP: Sampling Error Programs Package*, Ann Arbor: Institute for Social Research, 1972, 184 p. (out of print).

DAHLSTROM, P., JOS, O. & WAHLSTROM, S. (1973). The Swedish labour force surveys. Statistical Tidskrift 3, 189-203.

KISH, L. and FRANKEL, M.R., «Inference from Complex Samples», *Journal Royal Statistical Society*, (B), 36, (1974), 1-37.

MURTHY, M.N. (1974) Evaluation of multi-subject sample survey systems, *Int. Statistical Review*, 42.

KISH, L. Representation, Randomization and Control», Chapter 9 in *Quantitative Sociology*, ed., by Blalock, Borodkin, Boudon, Capecchi, 1975. Academic Press.

KISH, L. (1976) Optima and proxima in linear sample designs, JRSS, A, 139, 80-95.

KISH, L., GROVES, R.M., and KROTKI, K., Sampling Errors for Fertility Surveys, Occasional paper No. 17, World Fertility Survey (1976), 61 pages.

ANDERSON, D.W., KISH, L. and CORNELL, R.G., Quantifying Gains from Stratification for Optimum and Approximately Optimum Strata Using a Bivariate Normal Model. *Journal of the American Statistical Association*, (December, 1976), 71, 887-892.

COCHRAN, W.G. (1977) *Sampling Techniques*, 3rd ed., New York, John Wiley and Sons, Sections 5A.3-.6.

KISH, L. Chance, Statistics and Statisticians, (1977) Presidential Address), *Journal of the American Statistical Association* 73 (1978), 1-6.

KISH, L. Robustness in Survey Sampling, *Proceeding of the International Statistical Institute*, 1977, 47, Vol. 3, 515-528.

GROVES, R., Krotki, K., and KISH, L. The Analysis of Sampling Errors, *Proceeding of the American Statistical Association*, 1977.

BEAN, J.A. and BURMEISTER, L.F. (1978) A review of optimal sample allocation for multipurpose surveys, *Biometrika*, 20, 3-14.

KISH, L. On the Future of Survey Sampling, in N.K. Namboodiri ed. *Survey Sampling and Measurement*, New York: Academic Press, 1978, 13-21.

KISH, L. and ANDERSON, D.W. (1978) Multivariate and multipurpose stratification, JASA, 73, 24-34.

KISH, L. (1979). Rotating samples instead of censuses, Census Forum, Honolulu: East-West Center 6, 1-13.

KISH, L. Samples and Censuses, *Internacional Statistical Review*, (1979), 47, 99-109. Also as «Muestras y Censos» in Estadística (1981) 35, 37-51.

PURCELL, N.J., and LESLIE KISH, Estimation for Small Domains, *Biometrics*, (1979), 35, 365-384.

KISH, L. Diverse Adjustments for Missing Data, *Conference on Census Undercounts*, (1980), U.S. Census Bureau, 83-87.

KISH, L. Design and Estimation for Domains, London,

Institute of Statisticians, *The Statistician*, 29, 209-222.

PURCELL, N.J. and KISH, L., Postcensal Estimates for Local Areas (Small Domains), *International Statistical Review*, (1980), 48, 3-18.

KISH, L. Using Cumulated Rolling Samples, Washington: Library of Congress, (1981), 78 pages, U.S. govt. Printing Office No. 80-528-0.

KISH, L. Split Panel Designs, *SCPR Newsletter* (London) p. 2.

VERMA, V., SCOTT, C. and O'MUIRCHEARTAIGH, C. (1980) Sample designs and sampling errors for the World Fertility Survey, *JRSS*, A, 143, 431-473.

UNITED NATIONS (1980) *National Household Survey Capability Programme* (NHSCP), New York, UN, ST/ESAÅTAT. 92/Rev. 2.

PANEL ON DATA COLLECTION: *Collecting Data for Estimation of Fertility and Mortality*, Comm. on Population and Demography, Report No. 6, *National Academy of Sciences* (1981) 291 pages.

KISH, L. Sampling Methods for Demography, *International Encyclopedia of Population*, John A. Ross (ed.), (1982), McGraw Hill.

KISH, L. Design Effects for *Encyclopedia of Statistics*, John Wiley and Sons, (1982).

KISH, L. Chance, Statistics, Sampling, The Henry Russel Lecture of 1981 of the University of Michigan, *Statistica*, (Italy) Vol. 42 (1982), 3-20, also in *Journal of Official Statistics* (Sweden) (1985) 1, 35-47.

KISH, L. Rensis Likert 1903-1981, *The American Statistician* (1982) May 124-5.

RODRIGUEZ-VERA, A. (1982) Multipurpose Optimal Sample Allocation Using Mathematical Programming, Ann Arbor, Mi: The University of Michigan, Ph. D. in Dept. Biostatistics.

FOREMAN, E.K. (1983) Integrated programmes of household surveys: design aspects, *Bulletin of the International Statistical Institute*.

KISH, L. Data Colletion for Details over Space and Time, in *Statistical Methods and the Improvement of Data Quality*, T. Wright, ed., Academic Press (1983).

KISH, L., and VERMA, V., Censuses Plus Samples: Combined Uses and Designs, *International Statistical Institute*, Madrid, 44 th Session (1983).

KALTON, G. and KISH, L., Two Efficient Random Imputation Methods, Communications in Statistics, (1984), 13, 1919-39.

KISH, L. Analytical Statistics from Complex Samples, *Survey Methodology* (Statistics, Canada) (1984).

KIREGYERA, B. and GACHUKI, P. (1985) Experiences in panel surveys: examples from an integrated sample survey programme in Kenya, *Bulletin of the Int. Statistical Inst.*

MACURA, M. and CLELAND, J. (1985) Reflections on the World Fertility Surveys, Ch. 18 in Atkinson, AC and Fienberg, SE, *A Celebration of Statistics*, Berlin, Springer Verlag.

KISH, L. Sample Surveys versus Experiments, Controlled Observation, Censuses, Registers and Local Studies, (Kribbs Lecture) *Australian journal of Statistics* (1985).

KISH, L. Timing of Surveys for Public Policy, (Keynote address, 84) *Australian Journal of Statistics* (1985), also in *Statisztiai Szemle* (Budapest) (January 1985).

KISH, L., AGER, T., BAUMHART, W., III, WILLIAMS, D., Variance and Inference in a Judicial Decision, *Proc. Survey Research Methods*, Am. Stat. Assn. (1985).

DUNCAN, G.J., JUSTER, F.T., & MORGAN, J.N. (1986). The role of panel studies in research on economic behavior, Transportation Research (submitted).

DUNCAN, G.J. & KALTON, G. (1986). Issues of design and analysis of surveys across time, Bulletin of the International Statistical Institute, 45th session, (forthcoming).

KISH, L. (1986) Timing of surveys for public policy, *Australian Journal of Statistics*, 28, 1-12.

KISH, L. Statistical Design for Social Research, New York: John Wiley & Sons (1987) (about 300 pages).

# OPERACIONES ESTADISTICAS POR MUESTREO

# I. TIMING[1] DE ENCUESTAS PARA LA POLITICA GUBERNAMENTAL

**Resumen**

Este discurso de apertura de la 7th Australian Statistical Conference (1984) expone brevemente siete modificaciones al diseño muestral para mejorar la utilidad y la oportunidad de encuestas importantes para la política gubernamental. Estas son: mejores estimaciones para dominios pequeños, muestras rodantes acumuladas, más encuestas de panel, diseños multipropósitos para muestras periódicas, diseños subdivididos de panel (SPD) que combinen paneles con muestras, no solapadas, y recopilación frecuente acumulada para períodos de información menos frecuentes.

## 1.1. Introducción

Estas notas se refieren a distintos aspectos del timing de encuestas y a su importancia para la política gubernamental. Las he expuesto por separado en consultoría y en informes, para diferentes instituciones en diversos países y continentes. Como tienen alguna coherencia mutua, algunas relaciones, creo que esta presentación conjunta tiene un valor heurístico.

En cada uno de los siete puntos voy a hacer algunas modificaciones de los métodos actuales con el fin de satisfacer mejor algunos objetivos, lograr una mayor aproximación, o una mayor eficiencia en algunos criterios. Puesto que los criterios, objetivos y valores pertenecen al reino de la política gubernamental, estos aspectos se refieren a la interacción de los estadísticos técnicos con las personas y las oficinas implicadas en la política gubernamental.

Abogo por algunos cambios, por algunas modificaciones en diversos aspectos del timing de encuestas. Pero el sugerir cambios implica una crítica de las prácticas actuales, y eso provoca reacciones defensivas, muy naturales, de «huye o lucha». La diversidad de opiniones y las discrepancias sin violencia son la sal de la vida; los australianos son famosos por ello. A mí también me encantan las discrepancias y espero tener espacio y tiempo para ellas. Como no hay necesidad de argumentos triviales y artificiales, déjenme hacer tres ofertas de paz.

Primero, los cambios por los que abogo no son drásticos. Más bien se refieren a modificaciones evolutivas de las prácticas actuales, y muchas veces son adaptaciones de los métodos ya utilizados en otras situaciones. Realmente, su utilidad puede estar subestimada, porque les falta la novedad de las «nuevas poderosas herramientas», de brillantes innovaciones técnicas. Así, pueden quedar atrapadas entre dos fuentes de inercia: miedo al cambio e infravaloración de nuevos usos de métodos conocidos.

Segundo, mis observaciones son generales y no se dirigen específicamente a Australia. Mi conocimiento de las necesidades y experiencias de Australia no es ni lo bastante específica ni lo bastante profunda. Pero sé que los estadísticos de Australia están en contacto con los líderes del mundo, en estrecho contacto porque están justamente entre los líderes. Muchos de sus problemas y experiencias son semejantes, y lo mismo con los posibles cambios. La inercia, como el progreso puede ser contagiosa. Australia posee una característica peculiar para los que toman las muestras: su población se extiende sobre un área comparable a la de los países más grandes del mundo, aunque más pequeña en una proporción de diez a setenta; Canadá con un factor de dos es el país que más se aproxima.

(1) Se ha considerado conveniente no traducir la palabra TIMING por no existir su equivalente en castellano. TIMING se usa en inglés para indicar EL MOMENTO ADECUADO PARA LA REALIZACION DE UNA ENCUESTA.

Tercero, estas observaciones se refieren solamente a las encuestas muestrales, porque dicho tema es suficiente para mí hoy. Tocan indirectamente temas relacionados como el timing y forma de los censos, los registros, informes y otras fuentes administrativas. Entrar en discusión sobre ellas excede de los límites de mi tiempo y mis habilidades.

Finalmente, expreso mi optimismo de que los mejores métodos prevalecerán, de que los avances científicos crearán realimentaciones positivas, que el éxito alimenta. Los avances de las encuestas muestrales no han satisfecho una determinada demanda, pero han creado aún más demandas que esperan soluciones, que deben ser, no sólo perfectas desde el punto de vista teórico, sino también factibles operacionalmente.

### 1.2. **Métodos de Estimación para Dominios Pequeños**

«Estimaciones para áreas locales y otros pequeños dominios han sido de interés general durante mucho tiempo, pero no han estado disponibles excepto para estimaciones de censos de población, encuestas especiales, o registros administrativos. Este interés, sin embargo, ha sido sobrepasado por una demanda creciente de una mayor riqueza, diversidad y actualidad de los datos para pequeños dominios, necesarios para la planificación de reformas, bienestar, y administración en muchos campos... Muy recientemente, las demandas de datos a ser utilizados directamente para destinar dinero y recursos se han sumado (al menos en Estados Unidos) a las necesidades de los planificadores y a la curiosidad de los científicos. Esta demanda está ahora aumentando mucho los intereses y esfuerzos de la investigación en este área.

«Las estimaciones para pequeños dominios han estado muy abandonadas, hasta hace poco, en favor de la teoría de muestras y la teoría estadística. Como excepciones observamos que los estadísticos especializados en demografía han desarrollado diversos métodos competitivos para los recuentos de población de áreas locales...

«Sólo en los últimos años, el tener estimaciones para dominios pequeños se ha convertido en un campo activo para la investigación. Esto ha dado como resultado una variedad de técnicas estadísticas de aplicación a problemas de estimación en dominios pequeños» [Purcell & Kish, 1979].

Recientemente, se ha desarrollado un nuevo método de «ajuste proporcional iterativo» (IPF), como método de análisis de clasificación de datos. Estos métodos de «estimación que conservan la estructura (SPREE) han originado nuevas herramientas para la estimación en dominios pequeños, que no sólo incluyen áreas pequeñas, sino también otras poblaciones pequeñas. Su desarrollo necesitó no sólo un nuevo punto de vista teórico, sino también el desarrollo de modernos computadores ultra-rápidos [Purcel & Kish, 1980].

Con respecto a los métodos de estimación de dominios pequeños, Australia es líder en el *desarrollo* de nuevos métodos, con diversas contribuciones durante la pasada década. Australia es también líder en la *aplicación* de los nuevos métodos, con su reconocimiento y utilización en la política gubernamental.

Esta sección empezó con las nuevas demandas por el sector público de datos de dominios pequeños en el momento oportuno, demandas que surgen de las necesidades de la planificación gubernamental y de la distribución de los fondos públicos. Sin embargo, creo que las demandas de datos son casi ilimitadas y que la existencia de métodos factibles puede ser la principal causa de constricción y facilitación.

### 1.3. **Muestras Rodantes Acumuladas**

«Con el transcurso de los años nos hemos acostumbrado (en los Estados Unidos) a dos métodos para la recogida de información del público: el censo decenal y la encuesta de población actual... El contraste en la naturaleza de los esfuerzos —grandes, raros y singulares para el censo, pequeños, frecuentes y regulares para las encuestas— explican en gran parte la común aceptación de estas grandes diferencias en los períodos de información; mensualmente para las encuestas y una vez cada 120 meses para los censos. Después de todo llenar las lagunas entre los 120 meses depende de las suposiciones de suavidad, de estabilidad de las variables observadas. Por otra parte, los tamaños relativamente pequeños de las muestras mensuales dependen de aceptar promedios sobre dominios espaciales más grandes. Un buen ejemplo puede ser el de que muchas, quizás la mayor parte, de nuestras necesidades de datos podrían satisfacerse mejor con informes anuales, y algunas con informes trimestrales, pero con ba-

ses muestrales más amplias que las que en la actualidad son factibles para nuestras muestras mensuales.

«Hoy se está planteando un buen ejemplo con las demandas de datos más actuales para áreas locales. Los censos quinquenales podrían ser una respuesta a algunas demandas, pero con 60 lagunas mensuales tampoco serviría para otros muchos fines. Los censos anuales se han mencionado en algunas situaciones, pero incluso muestras de un 1 por ciento necesitarían esfuerzos grandes y esporádicos, que tenderían a incrementar los costes unitarios y a disminuir la calidad.

«Una solución a lo que parecen conflictivas demandas de economía, oportunidad y detalle pueden surgir de las muestras rodantes acumuladas» [Kish, 1981]. [Ver también Kish, 1979; 1965, 12.51; NCHS 1958]. Esta cita se refiere a la situación en los Estados Unidos, mientras que Australia y Canadá (y quizás otros) tienen censos cada cinco años. Para estos países la cuestión cambia: ¿suplementarían o reemplazarían muestras acumuladas anualmente a los censos quinquenales?

Las muestras rodantes (rotativas) se refieren a las muestras acumuladas, que expondremos en el punto 6, y que han sido diseñadas especialmente para «rotar sobre» toda la población en el espacio y sobre períodos designados. Pero la acumulación puede diseñarse para un censo muestral grande (por ejemplo, 10 por ciento) más que para un censo completo (Sección 6).

Las muestras rodantes (2) y las estimaciones mejoradas (1) prometen dar estimaciones más oportunas que los censos periódicos, datos más detallados y mejores estadísticos para dominios pequeños que las encuestas muestrales. ¿Existe tanta redundancia aquí que sólo se necesitaría uno de los métodos y habría que desechar el otro? No. Felizmente los dos métodos se refuerzan mutuamente con efectos sinérgicos. El uso conjunto de mejores estimaciones a partir de muestras rodantes puede estudiarse en las referencias citadas en la bibliografía.

### 1.4. **Paneles de Individuos**

Los paneles aquí representan muestras en las que los mismos elementos se miden en dos o más ocasiones para obtener cambios *individuales* $d_i = (x_{i1} - x_{i2})$. A partir de una buena muestra de las $d_i$, pueden estimarse las distribuciones de los cambios individuales en la población ($i = 1, 2, \ldots, N$) para diversas variables. A partir de las medias de estos cambios *internos, individuales y globales* podemos estimar los cambios *externos, medios y netos:* $d = (x_2 - x_1) = \Sigma\,(x_{i2} - x_{i1})/n$. Pero a partir de estos cambios en las medidas no se pueden estimar (directamente) los cambios totales en los individuos. Sólo los paneles pueden revelar los microcambios globales después de los macrocambios netos —excepto para situaciones inusuales bien de memoria confiable o bien de modelos monotónicos fuertes. Para estudiar los microcambios después de los macrocambios se necesitan paneles. Los necesitamos no sólo por sus frecuencias absolutas y relativas, sino también por las dinámicas de las relaciones y de la causalidad a nivel individual [Kish, 1985, Cap. 6; Kish, 1965, 9.5, 12.5C; Duncan & Kalton, 1986]. La palabra «longitudinal» o «estrictamente longitudinal» se utiliza también para «datos longitudinales» en estudios retrospectivos a partir de una entrevista, también en «estudios longitudinales» de sitios únicos (ciudad, institución) con poblaciones que cambian a través del tiempo [Duncan, Juster & Morgan 1986].

Algunas veces los paneles pueden parecer demasiado difíciles y ni siquiera factibles debido a la mortalidad, movilidad y a las negativas. Sin embargo, muchas veces esos obstáculos no son insuperables, y los paneles deben investigarse e incorporarse a las encuestas. La evidencia viene aquí de las muchas muestras solapadas» que se utilizan. Los paneles definen casos especiales de solapamientos completos, cuando las unidades muestrales son los elementos mismos. Muestras solapadas basadas en segmentos de área estables producen buenas estimaciones actuales y cambios netos, pero debido a la movilidad y mortalidad de los individuos dejan de proporcionar datos de panel automáticamente. Por consiguiente, propongo que algunas partes de las encuestas solapadas puedan y deban convertirse en paneles, para proporcionar los microdatos sobre las dinámicas del comportamiento individual, hoy tan necesarios y tan ausentes de las estadísticas sociales.

### 1.5. **Diseños Multipropósito para Muestras Periódicas**

Los diseños muestrales deben perseguir sus fines, usos y propósitos —más que dictarlos— y

deben seguirlos muy de cerca. Empezamos dando una lista de cinco propósitos principales, después buscamos los diseños muestrales que sirven mejor a cada propósito (o los más necesarios o los más eficientes). Por último, miramos qué compromisos, si los hay, pueden hacerse para «satisfacer» o «aproximar» diversos propósitos; en lugar de optimizar para un propósito y despreciar todos los demás. La exposición se refiere a muestras periódicas, porque parecen las más relevantes, pero también pueden servir para muestras repetidas *ad hoc* a intervalos irregulares.

**TABLA 1**
**PROPOSITOS Y DISEÑOS PARA MUESTRAS PERIODICAS**

| Propósitos | Diseños | Esquema de Rotación |
|---|---|---|
| A. Niveles actuales | A. Solapamientos parciales $0 < P < 1$ | abc - cde - efg |
| B. Acumulaciones | B. No solapamientos $P = 0$ | aaa - bbb - ccc |
| C. Cambios netos (medias) | C. Solapamientos $P = 1$ | aaa - aaa - aaa |
| D. Cambios globales (individuales) | D. Paneles | Los mismos elementos |
| E. Series cronológicas multipropósito | E. Combinaciones, SPD | |
| | F. Marcos maestros | |

En la Tabla 1 presentamos cinco propósitos y seis diseños. Los cuatro primeros están emparejados con letras semejantes en las cuatro filas. Estos emparejamientos distinguen a los diseños que mejor sirven, con varianzas reducidas y de otra manera, a cada uno de los cuatro propósitos. La mayoría de los estudios periódicos tienen varios propósitos, de aquí que afrontemos los difíciles problemas del diseño multipropósito —aunque no los resolvamos necesariamente. Realmente tres de los propósitos pueden servir también para otros diseños, pero con incrementos en las varianzas o en el coste; la única excepción es la de los cambios globales para individuos que requieren paneles, como observamos en la Sección 3. Muestras solapadas de segmentos de área no serán suficiente porque los elementos individuales (personas) se mueven entre los períodos. Por el contrario, para acumulaciones son mejores los no solapamientos. Así se hacen posibles los compromisos razonables y lo que necesitamos son definiciones claras de los propósitos y usos. Estos pertenecen al dominio de la política gubernamental.

Consideraciones extrínsecas pueden excluir algunos diseños, y los solapamientos pueden estar prohibidos o, por el contrario, ser necesarios; esto daría como resultado diseños menos eficientes, pero aún podrían ser diseños válidos. La principal variación entre estos diseños se refiere a la cantidad (y a la clase) de solapamiento entre períodos. El esquema de rotación de solapamientos completos muestra, con *aaa-aaa* etc., que todas las partes son comunes a los períodos. Los no solapamientos con *aaa-bbb* etc., muestran partes no comunes. Los solapamientos parciales con *abc-cde-efg* etc., muestran a la *c* y a la *e* como solapamientos de 1/3 sólo entre períodos sucesivos.

El enfoque utilizado aquí puede hacerse más general en varios aspectos, a pesar de que aquí hemos sido específicos y sencillos. Nos concentramos en cómo la variación del solapamiento relativo ($P = 1$ para solapamientos completos, $P = 0$ para no solapamientos y $0 < P < 1$ solapamientos parciales) afecta a las varianzas de las medias de diversos estadísticos que representan los tres propósitos básicos. Los tamaños y las estructuras de las muestras periódicas se supone que permanecen constantes.

También hay otras simplificaciones. Las comparaciones de las varianzas ignoran posibles diferencias en los costes de los elementos; pero las entrevistas, especialmente las telefónicas, pueden costar mucho menos que nuevas entrevistas. Segundo, las varianzas, basadas en $S^2/n$, ignoran los efectos de conglomeración del diseño; pero muchas veces éstos serán menores para los solapa-

mientos que para los no solapamientos. Tercero, la discusión supone correlaciones positivas $R$, que es lo más normal; pero $R$ puede variar desde valores pequeños para algunas variables (como sin trabajo) a valores grandes para otras (profesión, status económico).

**TABLA 2**
**EFECTOS DE LOS DIFERENTES SOLAPAMIENTOS SOBRE LAS VARIANZAS DE LAS DIFERENCIAS DE DOS MEDIAS**

| Diseños | Tamaños de Muestra para casos especiales | $1/n_x + 1/n_y - 2P_xP_yn_c/n_xn_y$ |
|---|---|---|
| A. Solapamiento parcial | $n = n_x = n_y$, $n_c = P_n$ | $2(1 - PR)$ |
| B. No solapamiento | $n = n_x = n_y$, $n_c = P = 0$ | 2 |
| C. Solapamiento completo | $n = n_x = n_y$, $n_c = P = 1$ | $2(1 - R)$ |
| D. Subconjunto | $n = n_x = n_y$, $n_c = P_n$ | $(1/P + 1 - 2R)$ |

Las muestras sin solapamientos, $P = 0$, son mejores para las acumulaciones, porque son más sencillas y también porque producen varianzas más bajas: $S^2/2n$ para medias de dos períodos y $S^2/Jn$ para $J$ períodos. Estas varianzas serían más altas al solapar muestras debido a correlaciones positivas. La varianza para la diferencia de dos períodos sería $2S^2/n$ para no solapamientos.

Por otra parte, la varianza vale solamente $(1 - R)2S^2/n$ cuando se miden cambios netos con solapamientos completos $P = 1$. Por consiguiente, el efecto de los solapamientos completos sobre las varianzas de las diferencias de las medias es $(1 - R)$. Este es el límite óptimo de $(1 - PR)$ del efecto de solapamientos parciales. Pero con estimaciones ponderadas mejoradas éste puede reducirse mucho, de modo que el efecto sobre la varianza sea $(1 - R)/\{1 - (1 - P)R\}$, que para valores grandes de $P$ se aproxima a $(1 - R)$. Por consiguiente, la estimación puede casi reemplazar grandes solapamientos. Estas ganancias en las varianzas de los cambios netos pueden ser mayores con valores grandes de $P$ y de $R$ *(Kish, 1965, 12.4-12.5)*.

Para medir cambios globales (macros) de individuos, se necesitan estudios de panel; éstos se exponen en la Sección 4. El conflicto entre las necesidades de paneles y de acumulaciones puede resolverse con los Diseños Subdivididos de Panel, expuestos en la Sección 6. Estos aparecen como ejemplos primarios de combinaciones de diseños para satisfacer las necesidades de diseños multipropósito. Las necesidades de series cronológicas no están bien definidas en las encuestas muestrales, pero podemos empezar con los cuatro diseños básicos anteriores (para profundizar en los conceptos anteriores, ver Kish (1985), 6.2.).

### 1.6. **Diseños Subdivididos de Panel**

Con este nombre (SPD) espero que se distinga este nuevo tipo de diseño que ayuda a combinar las ventajas de los otros cuatro diseños básicos (solapamientos parciales, no solapamientos, solapamientos, paneles), aunque es distinto a todos ellos.

El concepto básico es añadir a un panel $p$ una serie de muestras no solapadas, designadas por *a-b-c-d*, etc. Así los SPD periódicos se representan por *pa-pb-pc-pd*. El panel $p$ da los cambios individuales, y los no solapamientos *a-b-c-d* pueden acumularse en muestras más grandes. La muestra combinada proporciona los mejores solapamientos parciales para estimaciones actuales y es buena para los cambios netos. Por consiguiente, los SPD satisfacen los cinco propósitos. Los diseños muestrales para el panel y para los no solapamientos pueden diferir y optimizarse separadamente; pero la población y las medidas utilizadas deben ser semejantes y compatibles para la combinación.

Los SPD difieren de los diseños clásicos de solapamientos parciales, que tienen la apariencia de simétricos: todos los grupos muestrales están

en la muestra para patrones de ondas semejantes, todos los entrevistados están sujetos a la misma fatiga (¿pequeña?). Los diseños clásicos se diseñan rígidamente sólo para cambios de la media entre intervalos predeterminados, pero los intervalos críticos descubiertos más tarde se desprecian (aunque son los más corrientes). Por el contrario, los SPD tienen el solapamiento $p$ para *todos* los intervalos. Otra ventaja de los SPD procede de las correlaciones más grandes de recubrimiento de los mismos elementos, cuando éstos exhiben gran movilidad entre las unidades utilizadas en solapamientos ordinarios. Además, como los SPD tienen solapamientos para todos los intervalos, tiene la ventaja de la flexibilidad en la estimación sobre los solapamientos parciales rígidamente predeterminados; ya que éstos pueden no tener el intervalo con las correlaciones $R$ más grandes para medir los cambios. Por ejemplo, los solapamientos pueden ser anuales, pero las $R$ más grandes pueden presentarse mensualmente (o viceversa).

La diferencia fundamental del SPD y su principal ventaja reside en tener un panel propio $p$ del que carecen los solapamientos clásicos. Los paneles son necesarios para medir los cambios individuales. Pero también tienen problemas únicos. Aún así, los SPD mantienen una gran ventaja sobre el panel único: permiten a partir de la evidencia acumulada moderar y corregir el sesgo en que los paneles pueden incurrir. Otra ventaja se encuentra en la posibilidad de utilizar las muestras que cambian para reclutar de las mismas los reemplazamientos por caso de muerte o por renovación del panel.

Sin embargo, debemos reconocer que los paneles se enfrentan a grandes obstáculos de movilidad en los procesos de nacimiento/muerte, en los de fatiga y desgana del entrevistado. Por estas razones, en muchos casos, pueden ser más apropiados los diseños rodantes con solapamientos parciales.

Los SPD poseen varias características que dan como resultado sus ventajas sobre otros diseños. Si algún rasgo no parece factible puede omitirse y podemos retener todavía algunas de las ventajas de los SPD. Por ejemplo, si $p$ se convierte en un solapamiento permanente de segmentos de área, el panel se sacrifica, pero se pueden conservar otras ventajas. Además, este solapamiento puede girarse lentamente sacrificando algunas correlaciones. Por otra parte, una rotación lenta se puede efectuar dentro de un verdadero panel, y de esta manera conservar la mayoría de las ventajas del panel pero reducir la temida fatiga y el deterioro de los paneles. La flexibilidad de los tamaños variables de los no solapamientos (*a-b-c-d*) es otra característica que puede introducirse en los diseños que no son puros SPD (Kish, 1985, Sección 6.7; Kish, 1981b).

## 1.7. Recopilaciones Frecuentes Acumuladas Para Información Menos Frecuente

Es necesario distinguir tres clases de períodos en las encuestas muestrales: períodos de recogida durante los cuales se recogen los datos; períodos de *referencia* que pueden diferir entre variables y estadísticos; y períodos de *información* que pueden constar de un período de referencia o de varios acumulados. Por ejemplo, las encuestas semanales de los CPS (Current Population Surveys) de los Estados Unidos utilizan una semana para la recogida de datos y la semana anterior como referencia; y con inferencia discreta a todo el mes, que también sirve de informe. Existe una fuerte e innecesaria tendencia a fusionar estos tres períodos y a dejar de considerarlos separadamente. Propuse que el período de información debía ser menos frecuente (trimestral y anual) y el período de recogida más frecuente (semanal) (Informe para la National Commission on Employment and Unemployment Statistics, 1978).

Ahora tenemos tanta información sobre la economía, con tanto detalle, tan frecuente y tan exacta relativamente, que para los políticos existe la tentación de hilar demasiado fino, cambiar con demasiada frecuencia, mirar cada última cifra y ser empujado por ella. Nosotros hacemos nuestra predicción tan bien como cualquiera en el mundo, pero hay una tendencia a que los árboles no dejen ver el bosque (Sir Claus Moser, Director of the Central Statistical Office, UK, 1978). Yo también creo que nuestra sociedad y nuestro gobierno no están preparados para reaccionar ante cambios mensuales.

Segundo, las fluctuaciones mensuales (de empleo, etc.) comparadas con los cambios trimestrales o anuales, están sujetas, a menudo, a «shocks aleatorios» que no son componentes de modelos razonables. Con la mayoría de los informes mensuales también recibimos explicaciones *ad hoc* que implican el calendario, las tormentas, las huelgas y otros nuevos sucesos.

Tercero, los efectos de los errores muestrales deberían reducirse también para los informes trimestrales. Estas reducciones deberían ser muy bien recibidas para los dominios: regiones, dominios geográficos, demográficos y económicos que pueden ser tan importantes como los totales nacionales. Para los dominios, los errores muestrales en los informes mensuales desfiguran cualesquiera fluctuaciones significativas, y éstas se beneficiarían especialmente de las acumulaciones dentro de los informes trimestrales y anuales.

Cuarto, los informes semanales pueden continuar preparándose para usos científicos, gubernamentales y otros, pero hay que hacer hincapié en su falta de confianza. Sin embargo, el diseño muestral, énfasis y presentación deben trasladarse a los informes trimestrales y anuales. Observo con alegría que la Swedish Labour Force Surveys mensual (AKU) utiliza tres muestras mensuales independientes para cada trimestre. Pero hay un solapamiento de 7/8 de trimestre a trimestre, y cada nueva muestra se utiliza para 8 trimestres consecutivos (Dahkström & Wahlström, 1973). «El sistema de rotación elegido implica una preferencia por la precisión en las estimaciones de los promedios trimestrales y de las diferencias entre promedios trimestrales. Sin embargo, en la actualidad, los periódicos suecos parecen sobre todo exponer los cambios de un mes al siguiente. Tales cambios podrían haber sido mejor estimados, para ciertas variables considerablemente mejores, con un sistema de rotación diferente. Todavía el sistema elegido se basa en las preferencias de otros consumidores». ¡Bien dicho y bien hecho!.En 1984 el Reino Unido trasladó también su Labor Force Surveys a tres muestras mensuales no solapadas en cada uno de sus informes trimestrales. Hay un solapamiento de 4/5 entre los informes trimestrales y de 1/5 entre los anuales. Cada unidad de vivienda permanece en la muestra durante 5 muestras sucesivas trimestrales.

Quinto, la estrategia estadística debe exigir información menos frecuente especialmente para dominios pequeños pero no despreciables. Muy a menudo o no se ofrece información sobre dichos dominios o esta información contiene grandes errores o se obtiene a un coste demasiado alto, o ambas cosas. Como ejemplos importantes considérense las vastas pero escasamente pobladas zonas con pequeñas poblaciones que existen en muchos países y para los cuales las autoridades provinciales exigen informes separados. Otros ejemplos provienen de grupos de ocupación, demográficos, étnicos, etc., para los que son necesarios datos separados. En lugar de la rígida práctica que prevalece ahora, sería preferible informar, acumular y diseñar para períodos más largos en el caso de estos dominios más pequeños. Las tablas para estas estadísticas deberían indicar los diferentes diseños utilizados para dichas estadísticas.

Por último, no hay necesidad de confinar el período de recopilación a una única semana dentro del mes, designada mediante una selección arbitraria y sujeta a las vaguedades de la casualidad, de las tormentas y del calendario. Si extender la muestra dentro de cada Unidad Muestral Primaria (PSU) parece demasiado caro, las cuatro semanas se pueden extender de manera equilibrada entre las PSU. Realmente, la mayoría de las áreas muestrales son áreas metropolitanas auto-representativas, y las PSU son bloques o áreas de enumeración, y éstas se pueden extender a cuatro semanas.

### 1.8. **Timing e Instituciones**

Pueden observarse grandes diferencias entre los países tanto en las fuentes de fondos para la investigación social y de encuestas como en la naturaleza y estructura de los centros que realizan las encuestas. Algunas de estas diferencias de organización pueden deberse al tamaño de los países en área y población, al desarrollo, económico y estadístico, o la estructura política y social. También sospecho que hay mucho de arbitrario, idiosincrásico, aleatorio, debido a accidentes de la historia y a personalidades. En cualquier caso los países deberían aprender unos de otros más a menudo, más que simplemente aceptar la teoría de Pangloss de que *su* situación representa un ajuste óptimo.

Los fondos provienen de cuatro fuentes, aproximadamente agrupadas en: oficinas gubernamentales; fundaciones públicas, privadas e internacionales; empresas e industrias; y universidades. Los centros de encuestas son de los tres tipos principales: gubernamental; privado, empresas e investigación de mercados; universidades e institutos de investigación. De los $4 \times 3 = 12$ posibles flujos de fondos de las fuentes a los centros, la mayoría son utilizadas, pero los centros gubernamentales rara vez reciben fondos no gubernamentales, excepto de los organismos internacionales. Las uni-

versidades rara vez dan fondos a los de fuera. Estas observaciones se basan en mi experiencia personal cualitativa y no dispongo ni de estadísticas para respaldarlas ni para las observaciones siguientes.

Tres asuntos concernientes a estudios periódicos enlazan esta sección con el título y el trabajo. Primero las fuentes de los fondos para la mayoría de las encuestas periódicas son gubernamentales. Otras fuentes no han tenido ni fondos ni interés en los resultados a largo plazo. Segundo, la mayoría de las encuestas periódicas las hacen los centros de investigación gubernamentales; los centros universitarios han mostrado menos interés, capacidad o iniciativa para el trabajo periódico. Pero hay notables excepciones en estudios a largo plazo como los de mi Instituto de Michigan, del NORC de Chicago, del SCPR en Londres ahora, y probablemente de algunos más. Tercero, las ciencias sociales y la política gubernamental necesitan muchos más estudios periódicos. Algunos de éstos deberían hacerse en las universidades y en otros centros no gubernamentales.

Finalmente, y con brevedad, les exhorto a todos ustedes a contribuir al desarrollo de las encuestas muestrales, tanto periódicas como *ad hoc*, tanto dentro como fuera de las oficinas gubernamentales. También que esto se haga para una mayor competición. Hay necesidad y espacio para una mayor contribución entre ambas vias. Los investigadores de la Universidad pueden contribuir con mejores métodos —aunque muchos «metodologistas» están demasiado alejados de la realidad para ser de mucha ayuda. Los investigadores no gubernamentales necesitan ayuda de los fondos gubernamentales: guía política, metodológica y financiera. Una sugerencia específica concerniente a los estudios periódicos: los períodos piloto, de entrenamiento y del comienzo pueden hacerse mejor fuera de las oficinas gubernamentales, quizás en los centros universitarios; después el gobierno puede hacerse cargo de los estudios una vez aceptados, demostrados y estabilizados.

Para terminar, no tengo que añadir a las tan extendidas discusiones sobre cómo deberían organizarse las encuestas (o las estadísticas en general) entre los distintos departamentos gubernamentales de un país. Hay muchos que tienen más que decir sobre este tema y por lo tanto yo me callo por fin.

## Agradecimiento

El autor desea agradecer al editor y a Graham Kalton la ayuda que le han prestado con sus observaciones.

# II. DISEÑOS MUESTRALES MULTIPROPOSITO

**Resumen**

La mayoría de las encuestas se realizan con muchos propósitos y aquí proponemos una jerarquía de seis niveles. Todavía, la mayor parte de la teoría y de los libros de texto se basan en la teoría unipropósito, para evitar la complejidad y los conflictos de los diseños multipropósito. Se indican diez áreas de conflicto entre los propósitos, después se dan los problemas y soluciones para cada uno. Los compromisos y las soluciones conjuntas son afortundamente factibles porque muchos óptimos son muy planos; también porque la mayoría de los «requisitos» para la precisión son muy flexibles. Establecer y hacer frente a muchos propósitos es preferible a esconderlos detrás de un único propósito establecido artificialmente. También se hace más factible mediante el uso de los modernos computadores.

## 2.1. Introducción

La mayoría de los estudios se hacen con varios propósitos durante las etapas del diseño y luego es típico que surjan muchos más propósitos durante los análisis de datos y durante su interpretación y utilización. Sin embargo, la naturaleza real del multipropósito de la mayoría de los estudios tiende hoy a permanecer escondida bajo la superficie de discusiones univariantes y supersimplificadas de los estudios de diseño. Esto se hace mucho más evidente en las encuestas muestrales, que se expondrán aquí, pero creo que esta discrepancia también se mantiene para otros diseños estadísticos.

En la práctica, las encuestas suelen ser multipropósito. ¿Por qué entonces se desdeñan los diseños multipropósito en la teoría muestral? Porque la teoría del multipropósito sería demasiado compleja. Incluso las descripciones que leemos de los diseños muestrales actuales tienden a seguir y a tomar prestado el prestigio de la teoría muestral unipropósito y univariante, más que a retratar fielmente los numerosos compromisos de la compleja realidad. Muchos diseños corrientes, (especialmente las selecciones EPSEM) sirven probablemente y con rigor una variedad de propósitos, pero parece que es raro el diseño explícito de diseños multipropósito. Por eso, este tema parece ser merecedor del rango más alto en una escala de razones necesidad/disponibilidad.

Hay diversos aspectos en la naturaleza del multipropósito de las encuestas muestrales y éstas se exponen en la Sección 2.2 en una jerarquía de seis *niveles*. Luego, se especifican las diez áreas de *conflicto* entre propósitos en la Sección 2.3. En las Secciones 2.4 a 2.9 se tratan determinadas áreas de conflicto que presentan enfoques y soluciones a las mismas. Algunas de ellas se atribuyen a artículos ampliamente conocidos sobre encuestas muestrales, pero otras son más nuevas y menos desarrolladas, derivadas y referenciadas en su totalidad.

Al hacer esta revisión general mi primera y principal aspiración es proporcionar una referencia fácil de manejar sobre las aproximaciones, métodos y procedimientos de los diseños multipropósito; alertarles de la importancia y viabilidad de dichos diseños. Segundo, deseo proporcionar un marco para un trabajo integrado posterior sobre los numerosos problemas y conflictos de los diversos multipropósito. Las imperfecciones de mis métodos pueden servir de estímulo a otros para mejores derivaciones de ellos así como para desarrollar nuevos métodos.

## 2.2. Jerarquía para Niveles de Propósitos

Para empezar necesitamos clarificar el significado de «multipropósito», porque hay demasiados conceptos confusos bajo este vocablo en nuestra literatura. La mayor parte del tiempo se pro-

ponen en la Tabla 2.1 o el nivel 3 o el nivel 4, y por esa razón se ha distinguido el «multi-sujeto» del multipropósito para las mismas variables o las más estrechamente relacionadas. (Murthy, 1967, Sección 9.11, 14.13). Obsérvese que cada uno de estos niveles puede tener varias manifestaciones específicas, que sería útil exponer si se dispone de tiempo.

Las operaciones de encuesta integradas en el nivel 5 están relacionadas, pero deberían distinguirse de las encuestas multi-sujeto porque se refieren a organizaciones e instituciones que dirigen muchas encuestas en diversos campos durante períodos más largos de tiempo (UN 1980, Foreman 1983). El nombre primitivo fue «operaciones de encuesta contínua» cuando se admitía que las encuestas muestrales muy a gran escala, ampliamente extendidas se realizaban por organizaciones de encuesta continua como el U.S. Census Bureau, o nuestro Survey Research Center. Esta continuidad tenía grandes ventajas en costes y calidad, con efectos profundos en el diseño de la muestra (Kish 1965, Sección 12.6).

Marcos maestros o muestras maestras en el nivel 6 se refieren a extensiones y especializaciones posteriores de enfoques multipropósitos. Pueden referirse simplemente a utilizar los mismos mapas, o listados en bloque, o segmentos de área de varias encuestas diferentes; o al ejemplo en gran escala de la «Master Sample of Agriculture» (King and Jessen 1945), donde las áreas rurales de todos los condados de los Estados Unidos estaban divididas en segmentos de cuatro granjas cada uno; o a la empresa que vende listas actuales de unidades de vivienda para la mayoría de las muestras utilizadas en Alemania Occidental. Estos ejemplos muy diferentes tienen una base común en el ahorro de compartir los costes de «instalación» (diseño, estratificación, listado, etc.) para construir los marcos muestrales.

Los niveles 1 y 2 no se encuentran por lo regular en la literatura sobre muestreo multipropósito, aunque son los más corrientes y pueden tener los efectos más drásticos y causar los conflictos más dramáticos, como veremos posteriormente. El efecto de los diseños puede ser muy diferente para los estadísticos como medianas y cuantiles o coeficientes de regresión que para medias y para agregados (Kish 1961, Kish 1965, Secciones 3.5, 12.9; Kish and Frankel 1964). Además el diseño de las muestras periódicas nos lleva a nuevas consideraciones (Sección 2.8) (Kish 1985). Pero los efectos dramáticos pueden verse simplemente para las medias de pequeñas «subclases» (por ejemplo, tan pequeñas como 0.10 ó 0.01) de toda la muestra, que representan «dominios» semejantes de la población (Sección 2.5).

Cada uno de los seis niveles de propósitos presenta aspectos diferentes de los diseños y cada nivel puede explorarse provechosamente para significados y ejemplos más específicos, algunos de los cuales aparecen en la Tabla 2.1.

Las dificultades de los diseños multipropósito, que han sido la causa de que se hayan omitido y evitado son de varias clases. Primero, los diferentes propósitos deben formularse explícitamente en términos estadísticos, de modo que éstos puedan servir en las fórmulas para sus comparaciones y sus compromisos; obtener una lista completa de dichos términos, puede ser el principal obstáculo. Segundo, para cada propósito se necesitan estimaciones de la varianza y de los factores de coste. Tercero, los valores deben obtenerse y asignarse a las precisiones «requeridas» para todos los propósitos, pero esto sólo es necesario para algunos métodos (Sección 2.5). Cuarto, los valores y las estimaciones anteriores deben combinarse en una formulación matemática para llegar a la solución de un solo diseño «óptimo» que sea realmente utilizado. La tarea de calcular dicha solución puede resultar más fácil utilizando ordenadores, pero queda la comprensión de la teoría y de los conceptos (Sección 2.5).

La dificultad de esta tarea nos ayuda a comprender por qué se han evitado en los libros de texto, y también en las descripciones multipropósito. Muchas veces se presenta un único estadístico (por ejemplo, la media) de una única variable principal, como el único propósito del estudio. En el esquema del diseño multipropósito es equivalente asignar una importancia de cero a los demás propósitos. La falta de realismo de esta pretensión puede suavizarse por otra: que otros propósitos importantes darían como resultado afijaciones semejantes; pero esta pretensión debe afianzarse con los cálculos de los cuatro pasos anteriores.

## 2.3. **Revisión General de las Areas de Conflicto**

La breve lectura de diez áreas del conflicto, que aparecen en la Tabla 3.1, debería sernos útil

antes de fijarnos en problemas específicos y soluciones posibles. La lista probablemente no resultará exhaustiva, e invito a los lectores a hallar las áreas que he comentado. Aun más, espero que encuentren dentro de las diez áreas otros problemas y otras soluciones no comentadas aquí.

De esta larga lista de diez de conflicto afortunadamente no todas necesitan ser formuladas para cada diseño muestral. Creo que posibles conflictos acerca de a) los tamaños de muestra m y acerca de b) la relación de los sesgos con los errores muestrales debe considerarse siempre, al menos informalmente, porque son ubicuos. También c) afijación entre estratos debe recibir al menos una breve exposición, y a menudo más. El cálculo de los errores muestrales (j) debe también estar presente en la mayoría de las encuestas. Por otra parte, en una operación contínua con un marco muestral contínuo, las decisiones acerca de e), f), g), y h) podían haberse hecho hace mucho tiempo para un diseño fijo. Sin embargo, los tamaños del conglomerado (f) utilizados en etapas intermedias (bloques y segmentos) pueden estar abiertos a cambios operacionales flexibles. En el caso corriente de encuestas realizadas una vez, los conflictos i) acerca del diseño fuera de tiempo no necesitan considerarse.

También tranquiliza saber que los compromisos basados en los métodos estadísticos pueden producir resultados bastante aceptables, por varias razones. Primero, porque pequeñas desviaciones de la afijación óptima producen sólo pequeños o despreciables incrementos de la varianza (Sección 2.5). Las curvas de eficiencia tienden a ser planas dentro de amplias áreas alrededor de los puntos óptimos; por consiguiente, no se necesita una gran exactitud en los diseños que no sería factible. Segundo, sin embargo, también es verdad que grandes desviaciones de las afijaciones óptimas pueden originar de moderados a largos incrementos en las varianzas. Por tanto, ignorar propósitos importantes puede dar como resultado pérdidas substanciales de eficiencia para ellos, y por eso, aquellos propósitos deben incluirse en los diseños de compromiso. Tercero, los diseños de compromiso, de acuerdo con los métodos estadísticos, pueden reducir drásticamente las pérdidas para cada propósito, con sólo pequeños incrementos sobre los diseños óptimos separados para cada propósito (Sección 2.5).

## 2.4. Tamaños de Muestra y Razones de Sesgo ($B/\sigma$)

Estas dos áreas de conflicto, a y b de la Tabla 3.1, deben quizás considerarse como las más importantes de todas, porque pueden ser las más dramáticas. Aquí, las tratamos juntas sólo porque pueden estar estrechamente relacionadas a través de los efectos de las subclases. Empecemos con el familiar *srs* tamaño de muestra $m = S^2/V^2$ necesario para producir una media muestral $\bar{y}$, con varianza de elementos $= S^2$, para una «precisión requerida $= V$». Sin embargo, la $S_g^2$ depende en gran manera de las variables g y la $V_g^2$ «requerida» puede variar aún más. También incluimos los efectos de diseño $D_g^2$ que también varían, y por consiguiente, $m_g = S_g^2 D_g^2/V_g^2$ expresa el tamaño muestral necesario para la media de la variable g. Para la media $\bar{Y}_g$ de un dominio g, que comprende sólo la proporción $P_g$ de la población, el tamaño total de la muestra necesario se convierte en $n_g = m_g/P_g$, y es más seguro formular la fracción muestral necesaria $f_g = n_g/N = S_g^2 D_g^2/V_g^2 P_g N$. El factor $(1 - f)$ puede despreciarse o incluirse en $D_g^2$. La $P_g$ llega a ser la más importante si se «requieren» grandes precisiones para pequeñas subclases.

Para comparaciones de las subclases las varianzas se incrementan aún más:

$$m_g = (m_a^{-1} + m_b^{-1})^{-1} = n\,(P_a^{-1} + P_b^{-1})^{-1}$$

designado por $P_a$ y $P_b$ las proporciones en el tamaño n de la muestra. Por ejemplo, para la comparación de las dos medias de subclase de 0.01n y 0.10n, tenemos el «tamaño efectivo»

$$m_g = n(0.01^{-1} + 0.10^{-1})^{-1} = n/110.$$

Para otros estadísticos, tales como las medianas y los coeficientes de regresión, el formular «requeridos» tamaños de muestra se volvería complicado, pero con suerte se pueden determinar algunos números.

Las consideraciones para estadísticos de subclase puden llegar a modificarse en gran manera si, además de la varianza $\sigma^2$, incluimos también los sesgos $B^2$ en el error cuadrático medio $= \text{RMSE} = \sqrt{(\sigma^2 + B^2)}$ como una medida de exactitud. La Figura 4.1 se utiliza para retratar una tendencia común en la exactitud de los datos de la encuesta, aunque son posibles grandes diferen-

cias en las relaciones de los sesgos con los errores muestrales. En el eje horizontal, el error típico $\sigma_1$ aparece como incremento mediante un factor de aproximadamente 3 para $\sigma_2$ de una subclase de aproximadamente 1/10 de la muestra total. Para las comparaciones (diferencias) de dos de tales subclases se incrementa en 1.4 más aproximadamente.

Sin embargo, la hipotenusa que designa el RMSE aumenta mucho menos. En el $RMSE_1$ el sesgo $B_1$ aparece dominante, y esto puede ocurrir para algunas variables en muestras totales grandes. Sin embargo, el $RMSE_2$ de la subclase, a causa de que el sesgo se mantiene constante en $B_2 = B_1$, apenas se ha incrementado nada y está dominado por $\sigma_2$. Esto es aún más verdadero para $RMSE_3$, donde $\sigma_3$ ha aumentado, pero se supone que los sesgos tienden a desaparecer en la diferencia de medias, porque esta es una tendencia corriente.

Estas relaciones supuestas aquí no son matemáticas, sino empíricas y corrientes. Las propongo como una respuesta práctica a algunas preguntas corrientes tales como: ¿Por qué gastamos dinero para grandes muestras y en métodos muestrales rigurosos a la vista de sesgos de gran medida? ¿Por qué preocupan los errores muestrales de cálculo cuando los sesgos de la respuesta dominan el error total? La respuesta implícita está en la dominación de los errores muestrales en las subclases, y aún más en las comparaciones. Hagamos explícitas en el diseño muestral estas respuestas implícitas.

## 2.5. **Afijación entre Dominios**

Esta importantísima y frecuente área de conflicto tiene varios aspectos. Primero, considérense la afijación de un tamaño de muestra total (o esfuerzo o coste) entre los dominios que constituye una partición de la población total. Un ejemplo corriente es la afijación entre las diversas provincias (5,10,20 ó 50) o regiones o estados de un país; típicamente aquellos dominios tienen poblaciones muy desiguales $N_d$, con amplitudes de 1 a 100 quizás en los tamaños relativos, aunque pueden cubrir aproximadamente áreas de superficie iguales. Muchas veces las preguntas toman esta forma: ¿Deben ser los tamaños de muestra $n_d$ aproximadamente iguales, o debe ser $n_d$ proporcional a $N_d$, con tasas muestrales constantes $f_d = f$?. $n_d$ iguales tienden a producir errores aproximadamente iguales, $ste(\bar{y}_d)$ de las medias (proporciones, tasas) de cada región, por consiguiente, tiende a aproximarse al óptimo para las medias de los dominios separados. Por otra parte, la $f_d = f$ constante tiende a producir el $ste(\bar{Y}_w)$ más pequeño, para la media general $\bar{Y}_w = \Sigma\, W_d\, \bar{y}_d$ a causa de que produce errores más pequeños para dominios más grandes. Este error puede ser más pequeño de lo que se necesita para $\bar{Y}_w$, especialmente a la vista de los sesgos potenciales (ver Fig. 4), y esto es lo que retrae a los que proponen tamaños iguales $n_d$ para las provincias. Sin embargo, aquellos errores incrementados para $y_w$ también son soportados por la mayoría de las otras subclases, especialmente las «clases cruzadas» como edad, sexo, clase socioeconómica, etc.; y son inconvenientes comunes de la $f_d = n_d/N_d$ diferente para las provincias resultantes de valores iguales de n.

Por ejemplo, en la Current Population Survey de los Estados Unidos, se asignan las $f_d$ más grandes a los estados más pequeños (del oeste). El peso resultante aumenta las varianzas (para un coste total fijo) de las medias globales y también de las «clases cruzadas», tales como hombres y mujeres jóvenes, y especialmente de muchachos y muchachas «negros» entre los 13 y los 19 años (con tasas de desempleo muy elevadas).

Para reducir la confusión usual, distingo «dominios» para designar divisiones de la población, de «subclases», las particiones correspondientes de la muestra. Después distingo «dominios de diseño (y subclases) para referirme a las particiones (como provincias y regiones) que están contenidas en estratos definidos por el diseño muestral, de clases cruzadas (como edad, sexo, ocupación, ingresos, etc.) que cortan a través el diseño muestral y los estratos, muchas veces casi aleatoriamente. Los efectos del diseño difieren de estos dos tipos de subclases. (Kish 1961, 1980).

Además otras causas de conflicto pueden surgir de las diferencias de dominio en la distribución de variables, también de las diferencias o bien en las varianzas $S_d^2$ o bien en las precisiones «requeridas». Pero no necesitamos entrar aquí en estas complejidades. Además de llamar la atención sobre los problemas, debemos referirnos a dos métodos técnicos distintos para la solución conjunta de los conflictos en la afijación, (el cuarto paso indicado al final de la Sección 2.).

Un enfoque utiliza programación no lineal iterativa para satisfacer a *coste mínimo* las precisio-

nes «requeridas» *conjuntamente para todos* los propósitos establecidos. Estas elegantes soluciones a diversos problemas sacan partido a los modernos ordenadores y han sido publicadas en muchos artículos desde 1963 (ver revisiones en Bean y Burmeister 1978, Rodriguez Vera 1982, Cochran 1977, 5A.3-.6). El coste «mínimo requerido» a menudo se vuelve demasiado alto, porque las precisiones «requeridas» fueron poco realistas. Entonces a las soluciones se les sustituye drásticamente la escala con valores más bajos. Pero dicha sustitución expone las falsas pretensiones (bajo mi punto de vista) de esta elegante aproximación que depende de precisiones «requeridas» no realistas. Tampoco creo en la realidad de las «funciones en escalera» para las precisiones «requeridas» que asignan un valor constante a cualquier varianza inferior a la requerida $V^2$ y el valor cero a las varianzas superiores a ella.

Otro enfoque muy diferente recurre a alguna forma de *promedio* entre todas las afijaciones «óptimas» (preferidas) para diversos propósitos, minimizando la varianza combinada (ponderada) bien para un coste fijo o bien para un tamaño de muestra fijo. Desde luego, si las varianzas combinadas resultantes vuelven a ser demasiado altas (o bajas) las soluciones pueden escalarse hacia arriba (o hacia abajo) en el tamaño de muestra o en el coste total fijo. Prefiero esta solución, que arbitra entre afijaciones diferentes, cada una de las cuales optimizaría para un solo propósito. Esto implica asignar valores relativos de importancia $I_g$ para todos los estadísticos mencionados y esto puede parecer difícil (aunque si el que toma las decisiones es un ignorante puede asignar $I_g$ iguales a todos ellos). Pero las otras dos alternativas son más extremas y *serían* aún más difíciles: Bien determinar las precisiones «requeridas» de todos los estadísticos para el primer enfoque, asignando posteriormente pesos de igual importancia a todos ellos; bien especificar un estadístico con peso total uno, y por lo tanto pesos cero para todos los demás estadísticos.

Además, puede demostrarse que los compromisos son generalmente factibles y que merecen la pena, porque las afijaciones no son sensibles a cambios moderados de pesos de importancia (como muchas veces es verdad en estadística). Después de todo, cambiar la importancia relativa por razones de (por ejemplo) 2 ó 5 sería menos drástico que asignar el peso total 1 a una variable y 0 a las demás, un proceso que implica infinitas razones de importancia.

Primero, designamos con $\Sigma_i v_{gi}^2/n_i$ la varianza obtenida para un estadístico g con las afijaciones de los tamaños muestrales $n_i$ para el i-ésimo componente de variación. Si designamos por $1 + L_g(n_i) = [\Sigma_i v_{gi}^2/n_i]/Vg\ (\min) = \Sigma C_{gi}^2/n_i$ la razón del incremento (con afijación $n_i$) de la varianza del estadístico g-ésimo sobre su propia varianza mínima; resulta que $L_g(n_i)$ es la pérdida *relativa* sobre el valor mínimo de 1. Aceptar las varianzas relativas $C_{gi}/n_i$ como las funciones a ser minimizadas es una decisión crítica. Estas parecen ser más razonables que cualquier otra que se pueda imaginar para las funciones que van a combinarse en (5.1). Por ejemplo, las prefiero a las $V_{gi}^2$ que dependen de unidades arbitrarias de medida, que se sustituyen por la $V_{gi}^2$ (min). Aunque en algunos casos pueden estar recubiertas por $V_g^2\ (\min) = 0$ ó ser muy pequeñas, lo que pueden hacer $C_{gi}^2$ muy grande e inestable; en estos casos, asigno valores arbitrarios a las $C_{gi}^2$ o a los $I_g$.

Entonces con los pesos $I_g$ asignados según la importancia relativa del estadístico g-ésimo para cualquier conjunto de afijaciones $n_i$ de los tamaños muestrales.

$$\begin{aligned} 1 + L(n_i) &= \Sigma_g I_g\{1 + L_g(n_i)\} = \\ &= \Sigma_g I_g \Sigma_i C_{gi}^2/n_i = \\ &= \Sigma_i \Sigma_g I_g C_{gi}^2/n_i = \Sigma_i Z_i^2/n_i \end{aligned} \qquad (5.1)$$

Después de cambiar sencillamente el orden de los sumatorios, creamos las nuevas variables $Z_i^2 = \Sigma_g I_g C_{gi}^2$. Esta función puede minimizarse para obtener una solución de compromiso para un coste total fijo $\Sigma_i c_i n_i$. Para el conflicto entre $n_d = n/H$ de tamaños muestrales iguales para los dominios versus $n_d = nW_d$ proporcionales a los tamaños de dominio, se halla que las afijaciones de compromiso óptimas $W_d$ son proporcionales a $\sqrt{(W_d^2 + H^{-2})}$.

Los dos ejemplos de la Tabla 5.1 ilustran los sorprendentemente buenos compromisos entre afijaciones conflictivas producidos por el método del promediado ponderado. Primero, sus resultados en la cuarta fila de la Tabla 5.1 se comparan muy favorablemente con los otros. Las razones para los sorprendentemente excelentes resultados de los compromisos provienen de las superficies planas muy amplias para las afijaciones óptimas, como

se expuso en la Sección 2 y se mostró en la Tabla 5.2 (de Kish 1976 y Kish 1987, 7, 3).

Un ejemplo poco común pero significativo (negativo) nos lo proporcionan las World Fertility Surveys (excelentes por otra parte). Se eligieron tamaños muestrales iguales tanto para países grandes como pequeños, y tamaños muestrales reales que variaban sólo entre 3 y 10 mil y con una correlación no perceptible con el tamaño de población (a pesar de mi consejo). En consecuencia, hubo un incremento doble o triple de la varianza en sus «principales contribuciones al conocimiento»:

«Hasta aquí, la principal contribución al conocimiento ha sido confirmar la tendencia decreciente de la fertilidad que caracterizó a gran parte de Asia y Latino-América en la década de los 70 y hacer aún mayor el contraste con Africa donde la fertilidad y el deseo de tener muchos hijos permanecen altos» (Macura y Cleland 1985).

## 2.6. Afijaciones a Estratos y Elección de Estratificadores

Los dominios y los estratos se confunden a menudo al hablar de ellos, pero los dos aspectos se mantienen distintos en los problemas de diseños. Los dominios se refieren a las subpoblaciones para las que se buscan estimaciones separadas, mientras que los estratos son particiones generalmente más pequeñas creadas para reducir las varianzas. Por ejemplo, dentro de las provincias como dominios pueden crearse más estratos para reducir las varianzas de la provincia; pero dominios cruzados como edad, sexo y status económico tienden a distribuirse a través de los estratos. Aunque muchas veces las afijaciones de los tamaños muestrales para los estratos no son tan cruciales como las afijaciones para los dominios, pueden ser importantes en el caso de afijaciones óptimas desproporcionadas eficientes. Los dos métodos de la Sección 5 para asignar tamaños de muestra a los dominios pueden también aplicarse a las afijaciones a los estratos, aunque el objetivo difiere. Algunas de las referencias sobre programación no lineal se refieren a los dominios y otras a los estratos.

La presencia de varias variables y estadísticos de encuestas entre estos propósitos tienen claras implicaciones para utilizar más variables estratificadoras. Diferentes variables de encuestas tenderán a tener relaciones óptimas diversas con los estratificadores; es mejor, pues, usar muchos estratificadores, incluso si cada estratificador se utiliza con sólo pocas divisiones (categorías) por estrato. El diseño multipropósito es la mejor razón para la estratificación multivariante (Kish y Anderson 1978). Esto también puede justificar mejor la necesidad de métodos de «selección controlada». La elección de los límites del estrato, llamada «estratificación óptima», es un tema relacionado, pero de menor importancia, en esta condensada presentación.

## 2.7. Tamaños de Conglomerado; Medidas de Tamaño; Unidades de Retención

En las descripciones de los diseños muestrales se lee muchas veces que el efecto del diseño se ha aproximado con $D_g^2 = \{1 + roh_g\ (\bar{b}_t - 1)\}$, donde *roh* se mantiene para una correlación intraclase sintética de la variable «más importante» g y $\bar{b}_t = n/a$, el tamaño del conglomerado promedio. Esto produciría la varianza del elemento efectiva $s_g^2 D_g^2$ y la varianza $s_g^2 D_g^2/n$ para la variable g. Sin embargo, debemos cuestionar los contenidos de n y de $\bar{b}_t$. Si nuestra población consta de mujeres casadas en edad de tener hijos, pueden ser solamente el 10 por ciento del total de personas y en solamente el 30 por ciento de unidades de vivienda; y ellas son muchas menos que las de algunas poblaciones raras. Esta situación se ha tratado en el muestreo para elementos raros (Kish 1965, 11.4): «Por lo general, evitamos conglomerados grandes, a causa de sus efectos adversos sobre la varianza. Pero incluso los conglomerados grandes de toda la población producirán solo conglomerados pequeños de un elemento raro, si éste está ampliamente extendido. Por ejemplo, pueden muestrearse bloques enteros de personas de más de 65 años de edad; pueden buscarse personas con una enfermedad identificable en pueblos enteros. Si, por el contrario, el elemento está concentrado en áreas pequeñas, esas áreas pueden reconocerse y estratificarse en consecuencia.»

En diseños multipropósito, las clases cruzadas de la muestra serán de tamaños variables $m_g$, dimensiones del tamaño de la muestra total $n_t$, siendo $\bar{M}_g$ sus diferentes proporciones en la población. Por consiguiente, queremos estimar en el diseño no sólo $\{1 + roh_g(\bar{b}_t - 1)\}$ para las diversas variables g para la muestra total $n_t$, sino también $\{1+$

$roh_g(\bar{b}_g - 1)\}$ para muchas clases cruzadas. Entonces utilizamos algunas conjeturas que han demostrado ser buenas aproximaciones en miles de cálculos empíricos:

$$\{1+roh_g(\bar{b}_g-1)\} \cong \{1+roh_g(\bar{M}_g\bar{b}_t-1)\} \cong \{1+roh_t(\bar{M}_g\bar{b}_t-1)\}$$

Es decir, utilizamos $b_g = M_g b_t$ y $roh_g \cong roh_t$ como aproximaciones. Aunque esto subestima de alguna manera los valores promedio de $D_g^2$ para las clases cruzadas, a causa de las variaciones en los tamaños de los conglomerados de las clases cruzadas, es un pequeño factor comparado con las grandes variaciones de $roh_g$ entre las variables (Kish 1978, 7.1; Verma et alt 1980. Kish et alt 1976). Esta subestimación tiene pequeños efectos sobre la eficiencia del diseño. Pero es importante para considerar las eficiencias de las estimaciones de las subclases asi como para toda la muestra, consideraciones y tareas que suelen despreciarse. Esto tenderá fuertemente a eficiencias mayores para conglomerados más grandes que deberían mostrarse sólo para $b_t$ y n en la muestra total.

Las medidas del tamaño están relacionadas con los tamaños del conglomerado, pero no son enteramente lo mismo a causa de los errores en las medidas disponibles, debidos especialmente a contenidos de población diferentes y a obsolescencia. Esta es también una buena ocasión para observar problemas concernientes a las medidas de tamaño para «operaciones integradas de encuestas» de poblaciones diferentes que pueden necesitar especialmente compromisos drásticos. Por ejemplo, consideremos diseños integrados para poblaciones totales y para agricultura; quizás también para subpoblaciones étnicas; y quizás también para actividades comerciales o industriales. Las medidas de tamaño para cada una de éstas pueden diferir grandemente. Todavía puede ser factible alguna solución de compromiso que produciría eficiencias razonables para cada una. Desde luego, a la validez asintótica de utilizar tasas de selección uniforme no le afectan estas diferencias.

Las medidas de tamaño están también estrechamente relacionadas con los problemas de «Retención de unidades después de cambiar los estratos y las probabilidades» (Kish y Scott 1971). Aquellos métodos se diseñaron para tratar los cambios fuera del tiempo estipulado de las medidas de tamaño y de estratos, pero son también importantes en los cambios entre variables. «Muchas veces se asignan probabilidades de selección desiguales a unidades de muestreo. Nuestros métodos, aunque generalmente más aplicables, son necesarios especialmente en la selección de unidades de muestreo primarias para encuestas. A menudo éstas se eligen por separado de muchos estratos, con una selección de cada estrato.

«Después de la selección inicial las unidades pueden utilizarse para muchas encuestas durante varios años. Pero según pasa el tiempo, la necesidad de nuevas encuestas puede ser mejor atendida con nuevos estratos y una nueva selección de probabilidades, basadas en nuevos datos, que con los de la selección inicial. La diferencia entre los datos iniciales y los nuevos puede deberse a cambios diferenciales entre las unidades de muestreo como reveló el último censo, o a cambios en las poblaciones y en los objetivos de la encuesta. Por ejemplo, una muestra diseñada inicialmente para familias censales y personas puede necesitarse más tarde para una encuesta de agricultores o estudiantes de universidad». *Obviamente nuestros métodos se pueden aplicar también para diseñar simultáneamente un grupo de muestras relacionado con objetivos diferentes.»*

Este método nos permitiría usar por separado las mejores medidas (de tamaño y de estrato) para cada propósito, pero maximizando la retención del solapamiento de las unidades muestrales entre las muestras para propósitos separados (especialmente PSU). En el otro extremo sería posible diseñar un compromiso que promediase las medidas para lograr un solapamiento completo de unidades, pero sacrificando alguna eficiencia en cada uno de los propósitos. Un compromiso entre los dos extremos puede ser incluso mejor que ninguno: incrementar el solapamiento con pequeños sacrificios de eficiencias separadas reconociendo únicamente las diferencias de las medidas que superan algunos criterios mínimos arbitrarios (Kish y Scott 1971, 3a).

## 2.8. Propósitos y Diseños para Estudios Periódicos

Los estudios periódicos proporcionan áreas de conflicto de gran importancia creciente según se incrementan sus números y sus tamaños. No es verdad que aquellas encuestas influyentes y caras tengan sólo uno de los cinco propósitos expuestos en la Tabla 8.1, porque generalmente necesitan

varios o todos los que sean si el diseño se lo permite.

En la Tabla 8.1 indicamos cinco propósitos y seis diseños. Los cuatro primeros están emparejados con letras semejantes en las mismas cuatro líneas. Estos emparejamientos señalan los diseños que mejor sirven, con varianzas reducidas, cada uno de los cuatro propósitos. La mayoría de los estudios periódicos tienen varios propósitos y, por consiguiente, debemos hacer frente —no necesariamente resolver— los difíciles problemas de los diseños multipropósito. Realmente, los niveles actuales (A) y los cambios netos (C) pueden servir para cada uno de los seis diseños expuestos, incluso si algunos aumentan en varianzas o en costes. Pero los cambios individuales (brutos o micros) (D) necesitan paneles, y las acumulaciones (B) generalmente necesitan algunos cambios. A menudo pueden llegar a ser posibles compromisos razonables cuando pueden definirse los propósitos. Además, consideraciones extrañas pueden excluir algunos diseños (por ejemplo, los solapamientos pueden o prohibirse o reforzarse) y por consiguiente forzar el uso de diseños menos eficientes —pero todavia válidos—. A la variación principal en estos seis diseños concierne la cantidad (y clase) de solapamientos entre períodos. El esquema de rotación de solapamientos completos muestra, con aaa-aaa, que los períodos tienen todos partes comunes; el no solapamiento con aaa-bbb no muestra ninguno; y el solapamiento parcial abc-cde-efg muestra c y e como solapamientos de 1/3 entre períodos subsiguientes solamente.

Esta sección se concentra en los efectos de proporciones variables de solapamientos P en diversos diseños con propósitos diferentes; en solapamientos completos $P = 1$, en no solapamientos $P = 0$, y en solapamientos parciales $0 < P < 1$. Los propósitos se exponen en función de las varianzas para las medias estimadas, porque las medias (y porcentajes, tasas, proporciones) son las estimaciones más utilizadas y más sencillas de tratar. Los efectos sobre estas estimaciones no serán enteramente diferentes, pero son demasiados, diversos y difíciles para ser estudiados aquí.

Para una explicación más detallada consultar las referencias: (Kish 1987, Sección 6.2 o 1965 Secciones 12.4, 12.5). Allí también se pueden encontrar más estudios sobre paneles con sus ventajas, inconvenientes, problemas y soluciones. Llamo la atención sobre los SPD (SPD = Split Panel Designs), o Diseños Subdivididos de Panel que estoy tratando de promocionar como diseños multipropósito. Estos combinarían una muestra de panel P con nuevas muestras rotantes o «rodantes», de modo que Pa-Pb-Pc-Pd simbolizarían las muestras periódicas. Las a, b, c, d, etc., podrían acumularse en grandes muestras. El panel P se utilizaría como el solapamiento parcial para mejores estimaciones de los cambios en los niveles actuales y en los macro-cambios (media, neto) *para cualquier par de períodos*. El panel sirve primariamente para proporcionar micro-cambios (brutos, individuales).

## 2.9. Cálculo y Presentación de Errores Muestrales

Parece cuestionable incluir este tema bajo el concepto de diseño, pero no cabe duda de que es un problema multipropósito. Las estrategias para calcular y presentar los errores muestrales merecen hacerse por separado como área de conflicto entre los numerosos estadísticos dados generalmente para los resultados de las encuestas. Es suficiente presentar errores típicos para sólo uno o algunos de los estadísticos más importantes: Hay demasiados y muy diversos. A causa de esa diversidad, se ha desarrollado la práctica para calcular a partir de las varianzas otras funciones de los errores muestrales, especialmente las estimaciones de los «efectos de diseño» $d_g^2$; también algunas veces a partir de $d_g^2 = 1+roh_g(\bar{b}_g-1)$; también estimaciones de la correlación sintética intraclase $roh_g$.

Brevemente, aconsejo: a) Calcular los errores muestrales para muchas variables, porque las varianzas, los efectos de diseño ($d_g^2$). y los coeficientes intraclase ($roh_g$) pueden diferir, y lo hacen, grandemente entre las variables. b) Ustedes quizás tengan que hacer algunos promedios de los errores muestrales, porque quizás no sea conveniente o sea confuso presentarlos todos, c) Quizás no sea factible ni necesario calcular los errores muestrales de todas las subclases, porque éstas a menudo pueden aproximarse con modelos razonables. d) Es necesario presentar los errores muestrales de las subclases y de los otros estadísticos para guiar a los lectores de los informes (Kish 1965, 14.1, -2; Kish 1987, 7.1; Verma et al 1980). Espero que en el futuro este tema reciba de los teóricos y de los metodologistas la atención que no tiene y que desde luego merece.

## 2.10. **Conclusiones**

Las soluciones a las diez áreas de conflictos de la Sección 3 que propongo a partir de la Sección 4 hasta la 9 no han sido uniformes. El promedio de afijaciones entre los dominios en la Seccion 5 parece dar soluciones de compromiso sorprendentemente buenas. En la Sección 6 mi principal consejo es utilizar más estratificadores. En las Secciones 4 y 8 propongo considerar conjuntamente todos los propósitos importantes para los diseños muestrales. Hemos estudiado conjuntamente los diferentes niveles de propósitos y las diversas áreas de conflicto.

Hacer la pregunta correcta es el núcleo de la mayoría de los problemas. Propongo el diseño multipropósito como un nuevo paradigma, para reemplazar soluciones «óptimas» a preguntas artificialmente parciales tales como: ¿Cuál es la afijación óptima de la media $\bar{y}$ o la Y total de la variable más importante?.

TABLE 2.1

**Jerarquía de Propósitos**

1. Diversos estadísticos de la misma variable
   - — Totales o medias o medianas y cuantiles, distribuciones.
   - — Estadísticos analíticos: regresiones, análisis categórico.
   - — Aspectos de tiempo: estático, macro-cambio, micro-cambio, acumulativo.
2. Diversas poblaciones y dominios (subclases)
   - — Clases propias y clases cruzadas.
   - — Comparaciones de subclases.
3. Variables múltiples sobre el mismo sujeto
   - — Varias medidas de una variable; por ejemplo, de ingresos, o de desempleo.
   - — Diversos períodos: por día, semana, mes, año.
   - — Varios aspectos de un sujeto: ingresos, ahorros, bienestar.
4. Encuestas multisujeto
   - — Varios sujetos en el mismo plan de trabajo, entrevista, operación.
   - — Encuestas de salud de muchas enfermedades.
   - — Investigación de mercado para varios clientes, para muchos bienes.
   - — Encuestas de agricultura de muchas cosechas.
   - — Encuestas sociales «omnibus».
5. Operaciones de encuesta integradas, contínuas
   - — NSS en India, CPA, en Estados Unidos, NHCP en el Reino Unido.
   - — Encuestas separadas de una oficina y personal de campo.
   - — Fuente común de las encuestas.
   - — Métodos diversos, costes, operaciones, afijaciones, entrevistados.
6. Marcos principales
   - — Varias muestras de un marco o conjunto de listas.
   - — Instituciones separadas, organizaciones.
   - — ¿Personal de campo separado? ¿Las mismas PSU?

TABLA 3.1

**Areas de Conflictos**

a) Tamaños $m_g$ y tasas $f_g$ de muestras necesarias para los propósitos g

$$m_g = S_g^2 D_g^2 / V_g^2 \text{ para } n_g = m_g / P_g \text{ ó } f_g = S_g^2 D_g^2 / V_g^2 P_g N$$

b) Relación del sesgo con los errores muestrales en el RMSE $= \sqrt{(\sigma^2 + B^2)}$

- — La razón del sesgo $B/\sigma$ disminuye según $\sigma$ aumenta para las subclases.
- — Para las comparaciones $B/\sigma$ tiende a desaparecer según B disminuye, y $\sigma$ aumenta.

c) Afijación de $m_g$ entre dominios

$$m_t = \Sigma_g m_g$$

Cálculo y presentación de los errores muestrales

d) Afijación de $m_{gh}$ entre los estratos h

$$m_g = \Sigma_h m_g h$$

e) Elección de variables para estratificación

Estratificación Multivariante.

f) Tamaños óptimos de conglomerado

$D_g^2 = \{1 + roh_g(\bar{b}_g - 1)\}$ $\bar{b}_g = P_g n_t/a$ para las clases cruzadas.

g) Medidas para los tamaños de conglomerado

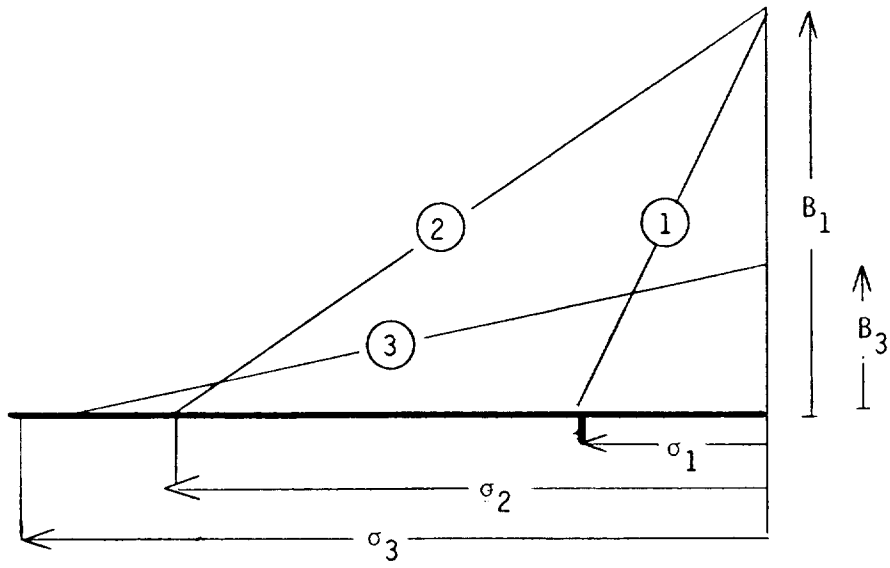
h) Unidades muestrales (PSU) de retención para sujetos medidas y estratos cambiados

i) Diseño a lo largo del tiempo (over time)

¿Cuánto solapamiento? ¿Paneles? Cambio versus acumulación.

j) Cálculo y presentación de los errores muestrales

**FIG. 4.1.**
**ERRORES VARIABLES ($\sigma$) Y SESGOS (B) EN ERRORES MEDIO-CUADRATICOS (RSME)**
**(Kish, 1987)**




Las bases representan errores muestrales y otros errores de la variable ($\sigma$). Por ejemplo, sea $\sigma_1$ el ste($\bar{y}_t$) de la media $\bar{y}$ de toda la muestra y sea $\sigma_2$ un ste($\bar{y}_c$) más grande de una media subclase, y sea $\sigma_3$ el ste($\bar{y}_c - \bar{y}_b$) de las diferencias entre las medias de dos subclases.

Las alturas representan los sesgos (B) y la hipotenusa el RMSE = $\sqrt{(\sigma^2 + B^2)}$ ; (véase 7.2F). 1) Para toda la muestra el sesgo $B_2$ puede ser grande comparado con el error de la variable $\sigma_1$, así tomando muestras más grandes disminuiría mucho el $RMSE_1$. 2) Sin embargo, con el mismo sesgo $B_1$, pero con una muestra más pequeña en la subclase, la razón cambia y $\sigma_2$ domina al $RMSE_2$; y éste no es mucho mayor que el de 1) a pesar de que la muestra es mucho más pequeña. 3) Además, para la diferencia de las medias, el sesgo neto $B_3$ puede ser mucho más pequeño; de modo que incluso con una mayor $\sigma_3$, el $RMSE_3$ para la diferencia es muy poco más grande que el $RMSE_2$. Este cambio drástico en la razón del sesgo $B/\sigma$ tiende a aparecer no sólo en las diferencias entre las subclases dentro de la misma muestra, sino también para diferencias entre encuestas repetidas.

**TABLA 5.1.**
**FUNCIONES DE PERDIDA (1 + L) PARA DOS POBLACIONES**
**(Kish 1976)**

| Afijaciones $m_i$ | (A) (1 + L) para $W_1/W_2 = 4$ | | | (1 + l) para 133 países: 0.2 to 100 mm | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | Conjunta con pesos | |
| | $\Sigma W_i \bar{y}_i$ | $\Sigma \bar{y}_i/2$ | Conjunta | $\Sigma W_i \bar{y}_i$ | $\Sigma \bar{y}_i/133$ | 1 : 1 | $I_c/I_d$ : 1 |
| | 1 | 1·56 | 1·28 | 1 | 6·86 | 3·93 | |
| $mW_i$ | 1·36 | 1 | 1·18 | 3·34 | 1 | 2·17 | |
| m/H | 1·08 | 1·125 | 1·102 | 1·35 | 1·54 | 1·44 | |
| $\propto \sqrt{W_i}$ | 1·116 | 1·080 | 1·098 | 1·31 | 1·28 | 1·295 | |
| $\propto \sqrt{(W_i^2 + H^{-2}}$ | | | | 1·47 | 1·17 | (1·32) | 1·27 |
| $\propto \sqrt{(0{\cdot}5W_i^2 + H^{-2})}$ | | | | 1·20 | 1·44 | (1·32) | 1·28 |
| $\propto \sqrt{2W_i^2 + H^{-2})}$ | | | | 1·12 | 1·66 | (1·39) | 1·23 |
| $\propto \sqrt{(4W_i^2 + H^{-2})}$ | | | | | | | |

En (A) hay dos estratos y dominios ($W_1 = 0.8$ y $W_2 = 0.2$); obsérvese que la afijación $m_i = \sqrt{W_i}$ trabaja con la misma eficacia para la pérdida conjunta (combinada) como para la óptima.

En (B) tenemos las poblaciones de 133 países, cuyo tamaño va desde 0.2 hasta más de 100 millones, una amplitud de 500 en tamaños relativos. De este problema de afijación (para la World Fertility Survey) omitimos, por razones prácticas, los cuatro países más grades y unos cuantos por debajo de los 0.2 millones. Su inclusión aumentaría la varianza de los tamaños relativos, $W_i$, de 2.5 a 12, y haría los resultados más dramáticos. Obsérvese que la afijación de $\sqrt{W_i}$ reduce las pérdidas bastante bien. Algún compromiso es mejor que ninguno. Pero la afijación óptima, $\sqrt{(W_i^2 + H^{-2})}$, es considerablemente mejor. Diferentes valores de $I_c/I_d$ = (1/2, 2/1 y 4/1) incrementan ligeramente la varianza de la función combinada de pérdida con pesos (1:1); pero permanecen constantes para las funciones combinadas de pérdida con sus propios pesos $I_c/I_d$ : 1.

## TABLA 5.2.
## PERDIDAS RELATIVAS (L) PARA SEIS MODELOS DE PESOS DE LA POBLACION ($U_i$): PARA PESOS DISCRETOS ($L_d$) Y CONTINUOS ($L_c$): PARA SALIDAS RELATIVAS ($K_i$) EN LA AMPLITUD DE 1 A K

(Kish 1976)

| Modelos | K | 1·3 | 1·5 | 2 | 3 | 4 | 5 | 10 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| Dicotómica U (1 − U) | | | | | | | | | |
| (0·5) (0·5) | | 0·017 | 0·042 | 0·125 | 0·333 | 0·562 | 0·800 | 2·025 | 4·512 |
| (0·2) (0·8) | | 0·011 | 0·027 | 0·080 | 0·213 | 0·360 | 0·512 | 1·296 | 2·888 |
| (0·1) (0·9) | | 0·006 | 0·015 | 0·045 | 0·120 | 0·202 | 0·288 | 0·729 | 1·624 |
| Rectangular | $L_d$ | 0·017* | 0·042* | 0·125* | 0·222 | 0·302 | 0·370 | 0·611 | 0·889 |
| $U_i \propto 1/K$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·099 | 0·155 | 0·207 | 0·407 | 0·656 |
| Decremento lineal | $L_d$ | 0·017* | 0·040* | 0·111* | 0·203 | 0·283 | 0·353 | 0·616 | 0·940 |
| $U_i \propto K+1-K_i$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·097 | 0·153 | 0·205 | 0·409 | 0·680 |
| Decremento hiperbólico | $L_d$ | 0·017* | 0·040* | 0·111* | 0·215 | 0·312 | 0·404 | 0·807 | 1·466 |
| $U_i \propto 1/k_i$ | $L_c$ | 0·006 | 0·014 | 0·041 | 0·103 | 0·171 | 0·235 | 0·528 | 0·011 |
| Decremento cuadrático | $L_d$ | 0·016* | 0·036* | 0·080* | 0·150 | 0·211 | 0·264 | 0·460 | 0·696 |
| $U_i \propto 1/k_i^2$ | $L_c$ | 0·006 | 0·014 | 0·040 | 0·099 | 0·155 | 0·207 | 0·407 | 0·656 |
| Incremento lineal | $L_d$ | 0·017* | 0·040* | 0·111* | 0·167 | 0·200 | 0·222 | 0·273 | 0·302 |
| $U_i \propto k_i$ | $L_c$ | 0·006 | 0·013 | 0·037 | 0·088 | 0·120 | 0·148 | 0·223 | 0·273 |

Dicotómica $1 + L = U(1 - U)(K - 1)^2/K$, también discreta con*

Discreta $1 + L_d = (\Sigma U_i k_i)(\Sigma U_i/k_i)$, con $K_i = i = 1, 2, 3, \ldots, K$

Continua $1 + L_c = \geqq Uk.dk. \geqq (U/k)dk$, with $1 \leqq k \leqq K$

Sólo dos valores, 1 y K, se utilizaron para $L_d$ for K = 1·3, 1·5 and 2

## TABLA 8.1.
## PROPOSITOS Y DISEÑOS PARA MUESTRAS PERIODICAS

| **Propósitos** | **Diseños** | **Esquema de Rotación** |
|---|---|---|
| A. Niveles actuales | A. Solapamientos parciales $0 < P < 1$ | abc - cde - efg |
| B. Acumulaciones | B. No solapamientos $P = 0$ | aaa - bbb - ccc |
| C. Cambios netos (medias) | C. Solapamientos $P = 1$ | aaa - aaa - aaa |
| D. Cambios globales (individuales) | D. Paneles | Los mismos elementos |
| E. Series cronológicas multipropósito | E. Combinaciones, SPD | |
| | F. Marcos maestros | |

## TABLA 8.2
## EFECTOS DE LOS DIFERENTES SOLAPAMIENTOS SOBRE LAS VARIANZAS DE LAS DIFERENCIAS DE DOS MEDIAS. SUPONGASE $S_x^2 = S_y^2 = S^2$ Y Srs O Deff = 1

| **Diseños** | **Tamaños de Muestra para casos especiales** | **Efectos sobre $S^2/h$ $- 2P_xP_yn_c/n_xn_y$** |
|---|---|---|
| A. Solapamiento parcial | $n = n_x = n_y$, $n_c = P_n$ | $2(1 - PR)$ |
| B. No solapamiento | $n = n_x = n_y$, $n_c = P = 0$ | 2 |
| C. Solapamiento completo | $n = n_x = n_y$, $n_c = P = 1$ | $2(1 - R)$ |
| D. Subconjunto | $n = n_x = n_y$, $n_c = P_n$ | $(1/P + 1 - 2R)$ |

# III. TRECE ASPECTOS BASICOS DEL MUESTREO DE ENCUESTAS

Estos trece aspectos que se exponen en este capítulo fueron tratados más ampliamente en el seminario al que corresponde esta publicación.

## 3.1. Diseño muestral y diseño de encuestas

### Diseño muestral

Métodos y procedimientos de selección.
Afijación a estratos.
Elección de unidades muestrales.
Afijación a conglomerados.
Estratificación.

### Diseño combinado

Estimación.
Análisis estadístico.
Errores muestrales.
No respuesta, no cobertura, imputación.
Población y elementos.

### Diseño de Encuestas

Variables de encuesta, Análisis substantivo.
Medidas, observaciones.
Errores de respuesta y sesgos.
Elección de dominios de análisis.
Presentación de datos y resultados.
Utilización de resultados.

## 3.2. Definición de población, elementos y unidades de encuestas

### Definición de población y elementos

Extensión y unidades de análisis.
Dominios = subpoblaciones de análisis.
Dominios de diseño y clases cruzadas.
Unidades observacionales.
Entrevistados, unidades de observaciones.
Unidades de muestreo.
Anidado jerárquico, identificación en muestreo multietápico A) PSU's − B, − C, −... − Elementos.
Poblaciones: muestral-marco-objetivo-inferencial.

Las muestras probabilísticas son la base de la población de encuesta obtenida, pero hay dos discrepancias entre ellas: los errores muestrales y las no respuestas de los items. Las dos difieren mucho entre variables y la cantidad de la no respuesta del item se muestra como muy diferente entre variables. Para ambas discrepancias las respuestas de la muestra sirven como base; los errores muestrales se calculan a partir de ellas; y se utilizan para «imputar» o pesar las no respuestas de los items.

Por consiguiente, las muestras probabilísticas se muestran como una base sólida y amplia de la población de encuesta, sobre la cual se construye la estructura de la inferencia acerca de ella. Para las discrepancias fuera de la población de encuesta se debe ir más allá de los datos de la muestra, con la ayuda de los datos implícitos o explícitos. El hueco de la población marco se debe a *no respuestas totales* de diversas clases (negativas, ausentes, etc.); el tamaño de las no respuestas puede estimarse a partir de los registros de la muestra (con esfuerzo y cuidado), pero el estimar sus efectos necesita modelos y datos auxiliares. El tamaño de *no cobertura* puede estimarse solamente con modelos o a partir de comprobaciones con fuentes exteriores, aunque esta parte también pertenece a la población objetivo. Esto puede también incluir una definida y deliberada *exclusión* de la cobertura.

Además, los datos muestrales se utilizan también para hacer inferencias más allá de las poblaciones objetivo y éstas son muchas, diversas y mal definidas. Las «Superpoblaciones» de la teoría del

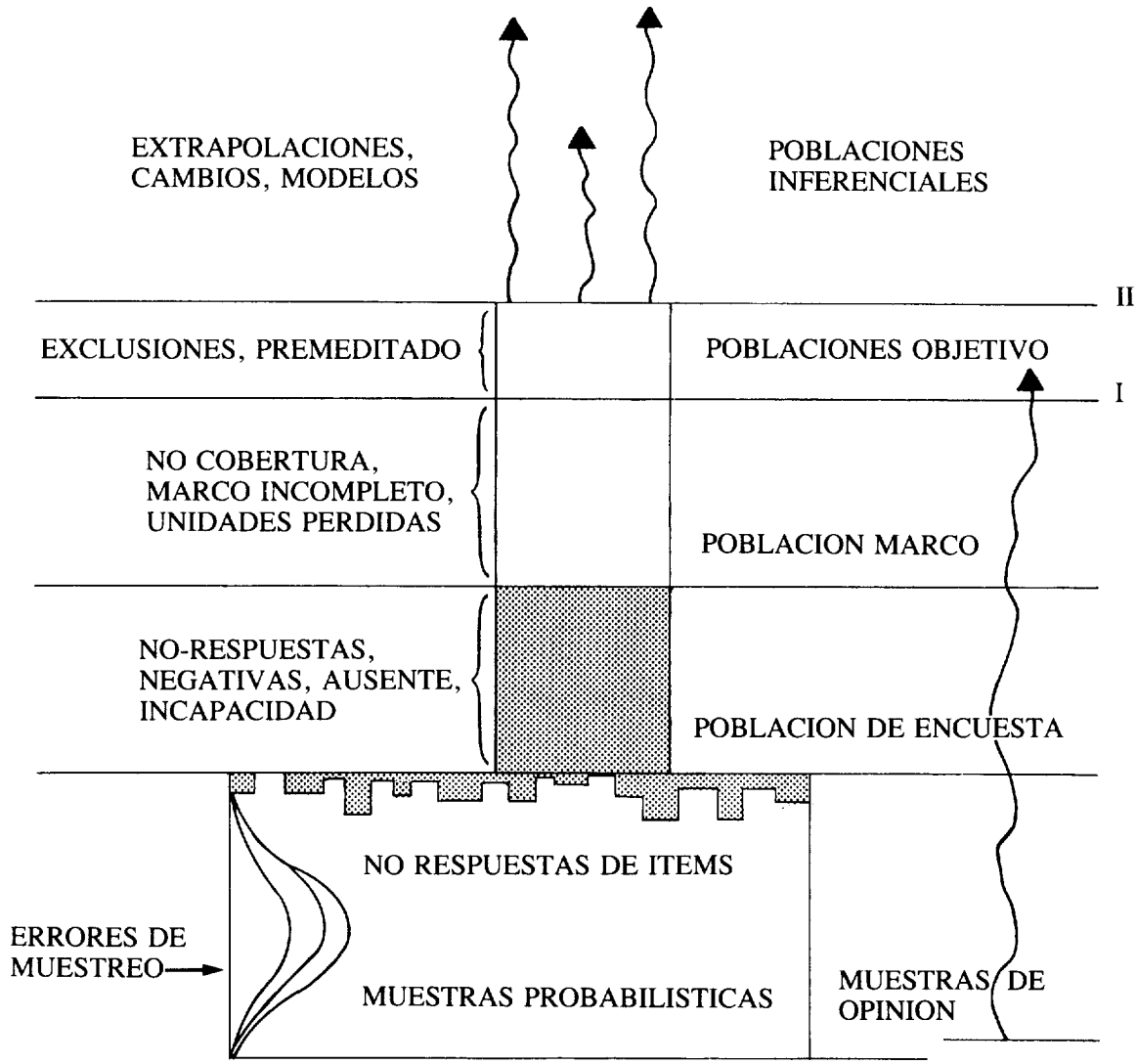



**FIG. 2.1.1**
**DISCREPANCIAS ENTRE CUATRO POBLACIONES (Reproducido de «Statistical Design for Research» por L. Kish, Wiley, 1987)**

muestreo no sólo están entre ellas, sino detrás de todas estas poblaciones inferenciales. Estas inferencias dependientes del modelo están demasiadas veces simplemente implicitas. Aún hay más vaguedad en la forma de describir el camino de las muestras de opinión directamente a la población objetivo; y dicha vaguedad se indica tanto mediante la delgada y ondulada línea de la población, como por las extrapolaciones a poblaciones inferenciales imprecisas.

### 3.3. **Selección de probabilidad**

Implicaciones.
Inferencia estadística.
Mensurabilidad.

Selección de probabilidad.
Selección mecánica, números aleatorios.
Listas o marcos.

**Alternativas al muestreo probabilístico**

Casual, conveniente, fortuita.
Selección dirigida.
Dependiente de la población.
Muestreo por cuotas.
Lugares restringidos, estudios de comunidades.
Replicaciones internas-ensayos extremos.
$P_i > 0$ conocida por todos los elementos $P_o$ confirmados por operación de selección.

### 3.4. Probabilidades iguales de selección - «Epsem»

Datos de ponderación propia a partir de selecciones Epsem p = f constante para todos los elementos de la población (*marco*).

f global lograda a través de tasas desiguales en etapas.

Selecciones PPS para conglomerados.

**Razones para Epsem**

La ponderación no es sencilla.
Errores de sistemas hombre-máquina.
Análisis secundario por investigadores distantes.
Estadísticos analíticos complejos.
Costes incrementados.
Incremento de las varianzas de pesos casuales.
Las relaciones públicas más fáciles con epsem.

### 3.5. Tasas de muestreo fijas (f), no tamaño de muestra (n)

Muchas veces se desconoce N y N' es variable.

Entonces n' = fN' siendo f fija y la n variable es mejor que f' = N'/n siendo n fijo y f' variable.

Ejemplos: n fijo tamaño de muestra total.

b fija para los conglomerados (es decir, bloques) con la variable $N_i$.

Inconvenientes de la f' variable.

Los pesos incrementan los costes, las varianzas, la complejidad.

Cuando b > N*i*.

Selección de campo de $b/N_i$ difícil.

Recuentos exactos de $N_i$ en campo.

### 3.6. Unir pesos a casos = elementos

Calcular el peso del elemento (caso) $w_j$ a partir de todas las fuentes, todas las etapas de selección, todos los ajustes.

Utilizar en todo momento:

$$\Sigma w_j y_j \; \Sigma w_j y_j^2 \; \Sigma w_j y_g x_j$$

También las razones

$$\frac{\Sigma w_j y_j}{\Sigma w_j} \quad \frac{\Sigma w_j y_j}{\Sigma w_j z_j} \quad \frac{\Sigma w_j y_j x_j \Sigma w_j}{\Sigma w_j y_j \; \Sigma w_j x_j}$$

el peso relativo $W_j = w_j/\Sigma w_j \quad \Sigma w_j = 1$

¿Replicación aleatoria para producir ponderación propia?

Replicaciones múltiples para reducir varianzas.

### 3.7. Fuentes y efectos de los pesos

**Fuentes**

a) Afijación desproporcionada - grandes diferencias estimaciones de dominios.
Afijación óptima a los estratos.

b) Desigualdades en marcos y procedimientos diferencias moderadas - ¿raras?

c) Ajustes para no respuesta y no cobertura -... ¿replicaciones para no respuesta de items?

d) Ajustes-ratios estadísticos, post-estratificación, tipificación.

Pesos casuales incrementan las varianzas,
¿las b, c, d anteriores? ¿no a)?

Ejemplos

1) Unidades de viviendas de edificios con 1 a 62:2.
2) b = 3 de segmentos con 1 a 200, grandes incrementos.
3) Muestras iguales para 19 provincias: incremento de 2.
4) Pesos h/4 para h llamadas no respuesta, incrementar 1, 3;
5) Los adultos en las familias censales en Estados Unidos (teléfono), aumentan 1.0.

Desviaciones moderadas tienen efectos pequeños en las varianzas.
Desviaciones grandes tienen efectos de moderados a grandes.
Los incrementos tienden a ser heredados en las clases cruzadas.

La afijación multipropósito puede aproximarse a epsem.

Incremento = $(\Sigma w_i k_i)(\Sigma w_i/k_i) = (\Sigma w_i k_i)^2/(\Sigma w_i)^2 = 1 + C_r^2$

**TABLA 1**
**RAZON DE LA VARIANZA DE LA MEDIA DEBIDO A DESVIACIONES DE LA FIJACIONOPTIMA, PARA DESVIACIONES RELATIVAS ($k_i$) EN EL RECORRIDO DE 1 A K. PARA UNA DISTRIBUCION RECTANGULAR DE LA FRECUENCIA DE LA POBLACION $U_i$ EN LA POBLACION. WITH $k_i$ = 1, 2, 3, ... K PARA DISCRETA, Y 1 < $k_i$ < K PARA CONTINUA**

| K | 1.3 | 1.5 | 2 | 3 | 4 | 5 | 10 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Discreta | 1.017 | 1.042 | 1.125 | 1.22 | 1.30 | 1.37 | 1.61 | 1.89 | 2.30 | 2.62 |
| Continua | 1.006 | 1.014 | 1.040 | 1.10 | 1.16 | 1.21 | 1.41 | 1.66 | 2.14 | 2.35 |

## 3.8. **Problemas del marco**

La habilidad más importante del arte (efecto) de las encuestas muestrales.

Queremos listar = elemento L - E o L - U U = unidad muestral.

### Imperfecciones

L - O Blancos y elementos extraños, no-miembros, no-miembros de población, o de subclase, no respuesta.

L - E - L Listados duplicados, replicados, múltiples, marcos múltiples dobles.

E - L - E Pequeños conglomerados de elementos.

O - L Elementos que faltan, marco incompleto, no cobertura.

(O - L) - (L - O) = subcobertura neta.

¿Evitar problemas?
1) Corregir listas.
2) Redefinir la población.
3) Ignorar los pequeños problemas.
Procedimientos específicos para mantener f para las E. También errores corrientes de cada una.

## 3.9. **Diseño de los tamaños muestrales**

Los costes son restricciones reales en la práctica.

Las varianzas son demasiadas y diversas.

Demasiados propósitos y mal definidos.

Los costes y los métodos de recogida de campo son decisivos.

Decisión real:
$C = C_o + cn \Rightarrow n = (C - C_o)/c \Rightarrow Var(\bar{y}_g) = S_g D_g^2/n$

$D_g^2$ = efecto de diseño

Teórico: $n = S^2/var(\bar{y})$

Multipropósitos.
Diferentes estadísticos: media, mediana, coeficiente de correlación.
Diferentes variables.
Diferentes sujetos.
Diferentes dominios.

Los **dominios** en la población están reflejados por las subclases en la muestra.

Dominios de diseño: provincias, estados, urbano, rural.

Clases cruzadas: sexo, edad, educación, ocupación, etc.

Dominios mayores: 5 ó 10, bien representados por la mayoría de las muestras.

Dominios menores: de 20 a 100 pobremente representados por la mayoría de la muestra.

Mini dominios: de 200 a 300 no representados por la mayoría de las muestras.

## 3.10. **Desviaciones de Srs**

I.I.D.: Idéntica e independientemente distribuidas.
«dadas n variables aleatorias I.I.D.», todo falso.
Independencia una poderosa herramienta.
Hipótesis algunas veces aproximada por condiciones.

Rara vez seleccionada realmente.
Falta de independencia en el muestreo de encuestas.

Los estadísticos **descriptivos** dependen sólo de $P_j$ de la muestra probabilistica.
Incluir $\hat{Y}$, $\bar{y}$, $s^2$, r, b, etc.
Robustez para las selecciones complejas $E(\bar{y}) = \bar{Y}$, $E(s^2) = S^2$, etc.

Estadísticos **inferenciales**, errores muestrales, intervalos de confianza.
Dependen de los métodos de selección.
Son sensibles a las desviaciones de srs.
Dependen de: las variables
los métodos de estimación (métodos estadísticos de selección)

$$\text{efectos de diseño} = \text{deft}^2 = \frac{\text{varianza real}}{\text{varianza srs}}$$

### 3.11. Muestreo de elementos estratificados

Selección de srs dentro de los estratos = subpoblaciones.
Muestreo de elementos aleatorios proporcionales (pres)

$$f_h = n_h/N_h = f = n/N \;/\; n_h/n = N_h/N$$

Las mismas tasas de muestreo = la muestra es una «miniatura» de la población.
Fácil de hacer con una selección sistemática a través de estratos.
Corriente, fácil, comprensible.

$\text{deft}^2 < 1$, pero pequeño
Para las clases cruzadas $\bar{M}_c$ $\text{deft}^2 \Rightarrow 1$ con $\bar{M}_c$
Para las diferencias $(\bar{y}_c - \bar{y}_b)$ $\text{deft}^2 \simeq 1$
y estadísticos analíticos.

Afijación desproporcionada (óptima)

$$f_h = \frac{n_h}{N_h} \propto \frac{S_h}{C_h}$$

### 3.12. Muestreo por conglomerados

Las unidades muestrales son conglomerados de elementos.
Los conglomerados están generalmente estratificados para una selección eficiente.
Seleccionar conglomerados compactos es lo más sencillo cuando los tamaños de los conglomerados son pequeños e iguales.

Utilizar el submuestreo cuando los conglomerados son grandes y desiguales.
Muestreo bietápico: submuestreo de los PSU seleccionados.
Muestreo multietápico.
Selecciones en etapas de subdivisiones jerárquicas.
PPS con MOS para crear tamaños de submuestra aproximadamente iguales a partir de conglomerados muy desiguales, para un primario f fijo.

$\text{deft}^2 > 1$, muchas veces $\text{deft}^2 >> 1$
$\text{deft}^2 = 1 + \text{roh}\,(\bar{b} - 1)$ donde $\bar{b}$ es el tamaño de la submuestra
$\text{deft}^2 = 1 + \text{roh}_s(M_s\, b - 1)$ decrece con $\bar{M}_s$ para clases cruzadas
$\text{deft}^2$ incluso está más próximo a 1 para $(\bar{y}_s - \bar{y}_t)$
$\text{deft}^2$ es también más bajo para los estadísticos analíticos r y B.

pero existe $> 1$ para todos los errores muestrales.

**Muestreo por áreas** es un procedimiento para la *identificación* de elementos con las unidades de muestreo
construir marcos para la selección sobre bases de muestreo jerárquicas.
Familias censales, granjas, etc., pueden identificarse con segmentos de área.

### 3.13. Cálculo y presentación de los errores de muestreo

Mensurabilidad es casi tan importante como la selección probabilística.
**Errores de muestreo** incluyen varianzas y funciones
efectos de diseño = var real/var srs
$\text{roh} = (\text{deft}^2 - 1)/(\bar{b} - 1) \;/\; CV^2(x)$ para estabilidad.

La **combinación** y la generalización son necesarias porque:

1) falta de precisión, no suficientes PSU, especialmente en dominios de diseño.
2) demasiados estadísticos en encuestas multipropósito.
3) utilización para otras encuestas y diseños.

La **mensurabilidad** de la varianza *necesita solamente*
conglomerados primarios.

identificación en todos los casos de los PSU y de los estratos.

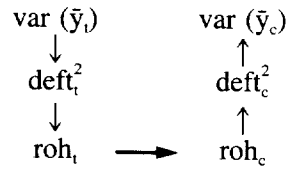
**Estrategia** para los cálculos.

**Calcular** $\text{var}(\bar{y}_t)$ para *muchas* variables
$\text{deft}^2(\bar{y}_t) = \text{var}(\bar{y}_t)/\text{var srs}$ para *muchas*
¿$\bar{y}_t$ promedio para clases de variables?

Calcular o imputar $\text{var}(\bar{y}_c)$ y $\text{deft}^2(\bar{y}_c)$ para clases cruzadas.
Calcular o impurtar $\text{var}(\bar{y}_d)$ y $\text{deft}^2(\bar{y}_d)$ para clases de diseño.

Computar o imputar var y deft para estadísticos analíticos.

Ruta para imputaciones




Presentaciones en tablas de errores de muestreo para errores típicos de clases.