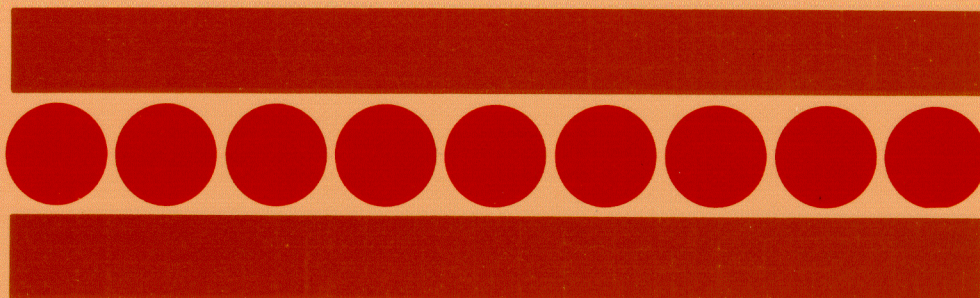


NAZIOARTEKO ESTATISTIKA
MINTEGIA EUSKADIN

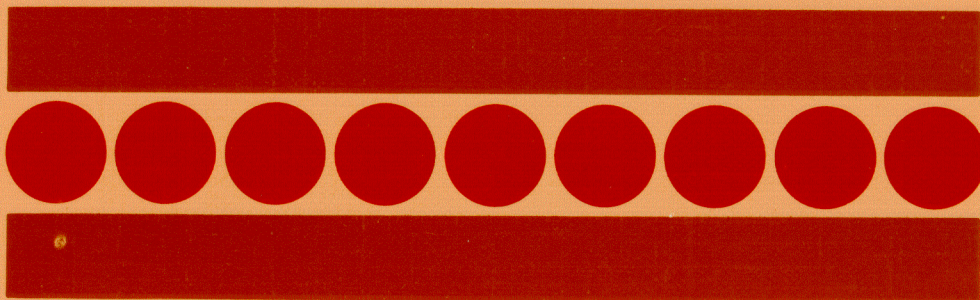
1988

SEMINARIO INTERNACIONAL
DE ESTADISTICA EN EUSKADI



SAMPLING AND NON-SAMPLING
ERRORS IN SURVEYS
(SEMINAR «QUALITY INDICATORS
IN SURVEYS AND CENSUS»)

A. MARTON



NAZIOARTEKO ESTATISTIKA
MINTEGIA EUSKADIN

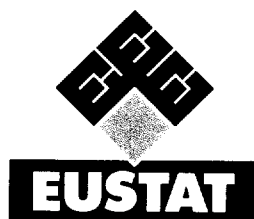
1988

SEMINARIO INTERNACIONAL
DE ESTADISTICA EN EUSKADI

**SAMPLING AND NON-SAMPLING
ERRORS IN SURVEYS
(SEMINAR «QUALITY INDICATORS
IN SURVEYS AND CENSUS»)**

A. MARTON

KOADERNOA 15 CUADERNO



Lanketa / Elaboración:

Euskal Estatistika-Erakundea /
Instituto Vasco de Estadística

Argitalpena / Edición:

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

© **Euskadiko K.A.ko Administrazioa**
Administración de la C.A. de Euskadi

Botaldia / Tirada:

1.000 ejemplares
I-1989

Fotokonposaketarako tratamendu informatikoa:

Tratamiento informático de fotocomposición:
Fotocomposición IPAR, S.C.L.
Particular de Zurbarán, 2-4 - 48007 BILBAO

Inprimaketa eta Koadernaketa:

Impresión y Encuadernación:

ITXAROPENA, S.A.
Araba kalea, 45 - Zarautz (Gipuzkoa)

Lege-gordailua / Depósito legal: S.S. 55/90

ISBN: 84-7542-127-10 Obra completa
ISBN: 84-7749-060-0

INTRODUCCION

Esta publicación tiene por objeto poner a disposición de los estudiosos interesados el texto del profesor Adam Marton sobre el curso «Quality indicators in surveys and census», que impartió en octubre de 1988 en el marco del VI Seminario Internacional de Estadística. El eje central de dicho curso gira en torno a la cuantificación de la influencia de los errores, tanto del muestreo como ajenos al mismo, en el trabajo estadístico; motivo por el que el texto entregado por el profesor Marton, y con ello la publicación, se titulan «Sampling and non sampling errors in surveys.» De especial interés encontramos la parte dedicada a los errores ajenos al muestreo, tanto en su tratamiento teórico como especialmente las experiencias llevadas a cabo por el profesor Marton en la Oficina Central de Estadística Húngara.

ADAM MARTON BIOGRAPHY

Country: Hungary

Born: 1934

Mathematician specialized in statistics.

His special interests are survey methodology sampling techniques, the problems of sampling and non-sampling errors, quality of data and small area estimation. He has published and lectured on these topics Hungary, at the U.N. and also in other countries. In 1968, he was awarded the Ph.D. by the Hungarian Academy of Sciences.

He has been working for the Hungarian Central Statistical Office since 1958.

Currently he is head of the unified system of household surveys section of the Hungarian CSO. He is responsible for the coordination and general methodological development of the household surveys carried out by the CSO.

He has been member and Hungarian representative of the Econometric Society since 1965, IASS member since 1984 and Regional Representative of the IASS since 1985.



CONTENTS

PREFACE	11
INTRODUCTION	11
NON-SAMPLING ERRORS IN SURVEYS	12
TARGET POPULATION AND COVERAGE	14
SAMPLING FRAMES	14
MISSING DATA, NON-RESPONSE	16
RESPONSE ERROR	17
CORRELATED RESPONSE VARIANCE	19
ESTIMATES OF RESPONSE ERRORS	21
SOME PRACTICAL PROBLEMS	23
SAMPLING ERROR: PORTABILITY	25
REFERENCES	27
APPENDIX	29
Introduction	31
The Sample	31
Organization	32
Survey Programme	32
Non-Response	32
Sampling errors	33
Response errors	38

PREFACE

In 1988 the 6th International Statistical Seminar was devoted to the topic of «Quality Indicators in Surveys and Censuses.» Though the accuracy of surveys is characterized first of all by the size of sampling and non-sampling errors, there are some additional indicators of accuracy too, such as e.g. timeliness or relevance of data.

As sampling error computations form a well-known chapter of statistics, we shall consider only the relation between sample design and sampling errors. Our attention will concentrate on non-sampling errors; the size of these depends on the subject of the survey as well as on its specific features, but —remarkably— it is actually independent of the type of survey, i.e. of the fact that the survey is a census or just a sample survey. The title of this paper indicates that those two kinds of errors will be considered which are of basic importance from the point of view of using statistical data.

This paper is based partly on methodological experience in surveys gained at the Hungarian Statistical Office. A brief outline of views and results on sampling and non-sampling errors will also be given; the questions will be discussed from the viewpoint of the official statistician, keeping mathematical considerations within reasonable limits. Details of the investigation will be left to the reader, whose work may be supported by the references. Considerations in polls and in sociological surveys are somewhat different from those in other types of research. Possible uses of survey data and the inferences based on them depend heavily on the methods used. Whatever the type of investigation, valid inferences may be expected only if information is processed by adequate methodology.

Introduction

Censuses and sample surveys are two most important methods of statistical data collection. In the case of censuses each unit of the population in consideration is observed. Quite different is the case of sample surveys where inferences to some population parameters are made on the basis of a relatively few units selected to the purpose from the population. Results of surveys are affected by numerous errors coming from different sources.

As in the case of sample surveys inferences to the whole population come from a small part of the total, statisticians are interested mainly in sampling errors when evaluating such surveys. This practice is based on the assumption that each observation, each piece of elementary information pertaining to some unit of the population is correct, and thus the lack of full coverage is the single source of error in survey data.

Some parts of data collected in household surveys can be regarded as correct; some data of economic statistics belong here, as well as certain personal data (as e.g. date of birth, sex, address). Such data can be checked or corrected partly by documents or registers, partly by re-interviews. A major part of household surveys is, however, connected with questions which are hard to answer.

The use of the term «error» in statistics is different from the everyday usage. The latter refers to something that has resulted from erroneous, irregular or careless behaviour, and that could be prevented by self-control and precise work. In statistics the term «error» denotes a much broader notion,

which can include in certain cases deviations that correspond to the everyday meaning of this word. Possible sources of errors in the course of collecting statistical data are inaccuracies of observation, memory, interpretation, measure instruments, data processing and analysis. The size of error is the deviation of the observed value from the true value.

Estimation of sampling errors is usually taken into account already in the sample design, and later it is part of evaluation of the results. A relatively new feature of investigations is represented by the endeavour to discover relations between sample design, sample subclasses and the size of sampling errors. This question will be dealt with at the end of the paper. By non-sampling errors we mean the entity of all errors coming from sources other than selecting the sample. Up to now such errors have not been thoroughly examined, though reliability of the results cannot be judged without them appropriately.

Non-sampling errors in surveys (1)

In recent literature on statistics particular attention has been paid to non-sampling errors, as their significance proved to be often greater than that of sampling errors.

The three main sources of non-sampling errors are the following:

- the target population is not fully covered by the sampling frame, some observations included in the design are unsuccessful. Some data are missing for non-response or for other reasons;
- some individual answers, measurements, observations are inaccurate;
- errors are made in the course of coding, editing, tabulating, etc.

Alternatively, sources of non-sampling errors can also be grouped by the successive survey phases as follows:

- design and preparations,
- data collection and
- data processing and analysis.

The different errors can also be classified as

- variable errors, and
- biases caused by some systematic effects.

This last distinction is very important from the view-point of reducing the errors, since variable errors usually decrease as the number of cases or replications increase, while completely different methods are needed for eliminating biases (auxiliary information and/or estimates). The most important distinction between variable error and bias is that biases can generally be reduced only by «doing something more» (e.g. designing a better questionnaire or improving supervision of field work), while variable errors are reduced by «doing more of something» (e.g. increasing sample size).

The size of error depends to some extent on human factors such as fitness and training of interviewers, but it is also affected by the economic and social environment. Many unforeseen things can happen in the course of data collection; nevertheless, it is assumed in principle that any survey can be replicated several times. It follows from this model that in every replication different answers will be obtained to the same questions. Random variation means that expectation of the observed values equals the actual value of the variable in consideration. Quite different is the case where the answers are systematically modified in the same way: then we have to do with bias, which is the difference between the averaged values (or expectation) of observations and the actual value.

Non-sampling errors are hard to estimate; in contrast with sampling errors, no exact estimation techniques are available here. The aim of investigating and recognizing non-sampling errors is mainly to reduce their effects in the different phases of designing and conducting surveys. This work is rather expensive and time-consuming, and this is the reason for restricting the investigation of non-sampling errors to periodic surveys in many countries. Examining non-sampling errors occurring in one-time surveys seemed unprofitable in most situations.

It is very hard or rather impossible to formulate statements of general validity. Defining proper solutions as well as producing estimates based on auxiliary information or re-interviews depends heavily on individual conditions and the scope of survey. In this paper problems arising in household surveys will be dealt with; such problems are in general extremely hard. The nature of errors in data collection of economic statistics enables better methods of control, and in certain cases some theoretical conclusions can also be applied.

Collecting data can proceed in different ways: by face-to-face interviews, by mailed questionnaires and by telephone interviews. In Hungary the tele-

(1) A comprehensive review of this field can be found in: National Household Survey Capability Programme. Non-Sampling Errors in Household Surveys: Sources, Assessment and Control. United Nations, New York, 1982.

phone network is not fit for the selection of a good sample of the population consisting of people having phones in their dwelling. Nor have mailed questionnaires (filled in by the respondents themselves) become really popular in this country, as experience has been rather unfavourable. In this paper we shall restrict ourselves to problems of face-to-face interviews. (Note that face-to-face interviews are sometimes used in countries with highly developed telephone network too, as nothing can substitute the advantage of immediate personal contact.)

Designing, conducting and evaluating household surveys are very difficult tasks, which necessitate considerable skill and experience. When designing different surveys we have usually to do with different sampling frames, questionnaires, methods to ask questions and to check them, furthermore, different ways of coding and data processing. Most surveys are «multi-purpose» investigations, in which different topics are examined on the same sample, either simultaneously or within short intervals. This implies that in the design phase certain issues receive greater emphasis, while others are treated as those of secondary importance. To have a proper design, which is necessarily some compromise between capacities (i.e. intellectual inputs and economic resources) and actual needs (such as small domains, accuracy) we need to know the nature and possible sources of errors.

Knowledge of the sources of errors is essential to have data of good quality, i.e. data which give a reliable, timely and detailed picture of the phenomena in consideration. Quality of statistical data is, however, a specific notion, which requires a definition of its own. As we shall see later, different criteria of quality may be in conflict with each other (e.g. timeliness-accuracy).

J. Norwood wrote the following on quality of data (1). Quality is one of the most important questions arising in statistical activities, still it is difficult to define. This notion has different meanings for different people. In a possible approach quality may be tantamount to the question if some data fit —well or just hardly— in the place where they should be used. In this aspect non-sampling errors play a very important role.

According to T. Dalenius (2) those data which are very important are in general fairly accurate.

Those data which are rather unreliable are either not used or unimportant; in the latter case their possible use would not lead to serious troubles.

Needless to stress that certain information must be available in right time. Sample surveys are often conducted just for the reason that the very long time needed to process the data of censuses is not feasible for the timeliness of information. For the sake of rapid information censuses are usually accompanied by sample surveys; to this end 1-2 p.c. samples are selected from the census records.

On the basis of what has been said above the different sources of non-sampling errors can be divided into two groups. The first group relates to some approximation for the population considered (sampling frame, coverage, non-response, etc.), while the second refers to the so-called response error, i.e. to accuracy of individual answers. In other words, by response error we mean the deviation of the information recorded on the questionnaire from the true value —provided that the latter can somehow be defined. To the second group of errors belong e.g. the false assignment of some category, errors of coding, etc. Problems of non-response errors will be investigated in the following on the basis of this classification. Note that problems of analysis (inference from cross tables, factor analysis or log-linear models) will not be considered.

The Journal of Official Statistics published by the Statistical Office of Sweden devoted a whole issue —No. 4 in 1987— to problems of non-sampling errors. Noted experts of statistics and survey methodology dealt with the significance of this topic in that issue, and expressed their opinion on some questions of methodology. A brief summary of contributions from that specific number of J.O.S. may be helpful to give an adequate picture of the state of the art of non-sampling errors.

Lars Lyberg, the Editor-in-Chief of J.O.S. emphasized that errors should be taken into account, whenever statistical data are collected. While sampling errors in surveys are in general treated by adequate methods, non-sampling errors are often completely ignored, though they are in many cases more important than sampling errors. Non-sampling errors may come from different sources, and it is difficult to measure or control them. Their significance can best be reduced by solving the problems in the place where they arise.

Barbara Bailer finds that one part of non-sampling errors from different sources is treated properly, while other parts are completely neglected. This is partly explained by high expenditures. It is very important to know the nature and possible sources of errors at each stage of designing and conducting surveys. One of the most important

(1) What is Quality? Paper presented at the US Bureau of the Census Conference, Baltimore MD, 30 March 1987 (Manuscript.)

(2) Tore Dalenius: Relevant Official Statistics, Journal of Official Statistics, Vol. 1 (1985) No. 2, pp. 21-34.

tasks is to investigate the distorting effect of missing data.

Tore Dalenius compares accuracy of statistical data with their relevance. He claims that sometimes concessions should be made in accuracy for the sake of rapid publications, since striving for great accuracy may be rather time-consuming. (Preliminary data of censuses based on 1-2 p.c. samples may serve as good illustrations of this statement.) In Dalenius' view accuracy of survey data can best be improved by raising the level of methodology, field work and training of interviewers.

Robert Barnes claims that non-sampling errors will never be eliminated completely. In the United Kingdom efforts are made to ensure better cooperation of respondents as well as to improve the coding of answers.

Thomas Jabine has doubts concerning the ideas of people about non-sampling errors. It is very likely that many users of statistical data are not aware of the existence of non-sampling errors, though they may have heard something of sampling errors that can well be controlled in theory and practice. It is very important to have clear notions on total errors, particularly if attitudes are considered, and even the definition of errors may be uncertain.

Robert Tortora regards the problem of estimating non-sampling errors and bias as the most interesting and fascinating challenge for statisticians. Three kinds of measure would actually be needed: sampling- and non-sampling error, as well as bias. It is relatively easy to set up theoretical models, but putting these into practice is expensive and necessitates re-interviews, which may be unreliable.

Dennis Trewin uses some time series to point out importance of the problem. He stresses that evaluations are needed both in the design phase and in the phase of conducting surveys to explore accuracy of individual observations. He also points out that proper use of computational facilities improves considerably accuracy of the results.

Ben Kiregyera examines the problem from the point of view of developing countries, with special regard to Africa, since household surveys take place in more and more countries. He emphasizes the problems of sampling frames and coverage of populations in consideration. In addition, he deals with the role of interviewers, and —considering another topic— with problems of measurement in estimating the yield of agricultural production.

Target population and coverage

One of the first steps in designing statistical surveys is to determine the scope of research, i.e. to

define the population in consideration. There are a lot of obstacles to conducting censuses, a typical obstacle is e.g. that some units (persons, families, organizations pursuing production or trade, etc.) of the population are always on the road and cannot be found on their site. For this reason both deficits and surpluses can occur, as in the case of some units it will be difficult to decide if they belong to the population or not. The result will be undercoverage in the first case and overcoverage in the second.

The purpose of survey samples is to provide cheap and rapid information on some population in a simple way, by means of some hundred or some thousand observations. To this end a suitable sample is needed.

There are many ways of selecting a sample. In this paper we do not deal with methods of selecting samples and estimating parameters, which leads in the end to estimating sampling errors. Our interest concentrates on problems and sources of non-sampling errors, one group of which is represented by problems of sampling frames and coverage. There are many ambiguous cases where it is not easy to decide if some error is due to sampling or to other factors.

Probability samples are used in most instances⁽¹⁾. In such cases units of the sample are selected by some random procedure. This is tantamount to assigning some predetermined positive probability to each unit of the population in consideration, and units will be selected in the sample with those probabilities. In practice the selection method proceeds by means of a list containing all units of the population to be considered. If a suitable list representing the sampling frame is available, then a proper method for selecting the sample can easily be found.

Sampling frames

In practice it is not easy to define sampling frames. In the case of household surveys e.g. the target population is usually the whole population of the country or some part of it (women, youth, pensioners, etc.). In principle the existence of a proper, up-to-date list is a necessary condition of selecting samples with good coverage and good representation of the considered part of population. In prac-

(1) Problems of covering a given population by probability samples or by other methods are considered e.g. in Chapters 1-2 of «Statistical Design for Research» by Leslie Kish. John Wiley & Sons, New York, etc., 1987.

tice, however, such lists are rarely available; instead of them we are given a multipurpose sample which is regularly adjusted and ensures fairly good coverage of different strata of the population.

Samples in practice are rarely determined by simple random selection; for the sake of efficiency stratified and/or multistage samples are used. In the case of stratified and cluster sampling the frames must contain sufficient information also on selection probabilities, which are usually unequal in such cases.

Ultimate sampling units in some sampling methods may be different from observed units: the former can be e.g. dwellings, and the latter persons. In many cases some persons selected in the sample are not available at the time of interview, but their absence is not compensated by possible guests who happen to stay temporarily in the dwelling visited by the interviewer.

Because of inevitable deficiencies outlined above the sampling frame is different from the target population. It is common practice that from sample surveys inferences to populations other than the original target population are also drawn. Subsets of the latter can be considered in this respect, but the opposite case can also occur: from data of cities inferences can be drawn to the country as a whole.

There is an additional gap between the sample and the set of sample units having completed the questionnaire with success: the difference corresponds to refusals and non-responses.

Deficiencies of sampling frame and losses from the sample due to non-responses are two different sources of non-sampling errors.

Examining the problems of coverage is hard and rather time-consuming, and different kinds of deficiencies cannot always be identified. For these difficulties many designers prefer neglecting problems in connection with goodness of coverage.

Estimates of absolute frequencies such as population counts are particularly sensitive to deficiencies of coverage. The estimates can be entirely misleading e.g. in the case where differences between subgroups or units are needed, and the missing units of observation are not uniformly distributed among the domains of interest.

Special foresight is needed when on the basis of observations from a given population inference to another, similar population is to be made, e.g. when from data of some country those of neighbouring countries are to be estimated. (This question is

closely related to the so-called synthetic estimates and the condition of their application.) Such inferences can be justified only if some specific conditions —scope of study, behaviour of variables, validity of some models— are thoroughly examined.

Definite distinction must be made between deficiencies of coverage and the different forms of non-response (item nonresponse, refusal, not-at-home incapacity, etc). It is, however, not always easy to distinguish between things: if e.g. the observation for some member of a family is missing, though the person in question had to be interviewed, the reasons for this can be different: refusal, interviewer's error, ambiguity in the instructions. Without knowing the reason for the loss it is not possible to correct the error.

In the foregoing we have considered sample surveys, thus we have assumed that the target population and the sampling frame are well defined, and the sample ensures good representation of the target population. In controlled observations samples are selected in a different way: here the sampling frame does not cover the target population, and units are selected to the purpose, but not quite at random. (Polls are conducted mainly by this method.) In most cases target populations are not defined previously when conducting controlled observations; the strata or occupation groups of the population characterized by the answers can be identified only in the phase of analysis, on the basis of distribution of the observations (1).

Current surveys of population or households have usually sampling frames that are during longer periods. Such sampling frames need permanent updating in order that changes in time may not increase deficiencies of coverage.

In Hungary the Unified System of Household Surveys (abridged USHS) has a multi-purpose sample including about 16,000 families. Updating and rotation of this sample proceeds by updating the stock of addresses of enumeration districts and by registering new buildings, blocks of dwellings in settlements belonging to the sample. New districts —if any— should of course also be registered (see Appendix).

A multi-purpose sample can be used in actual surveys only if reasonable compromises are made in the different specific situations. Even the method of updating the sample assumes some compromise. Thus it is meaningful to raise the question if in the case of one-time «snapshot» studies or surveys

(1) See Leslie Kish, *op. cit.*, Chapter I.

repeated only in every fifth or tenth year defining a new sampling frame for each survey is not more advantageous than updating the old frame. This would be, however, very expensive; the common practice adopted therefore in most countries is to use multi-purpose samples (1).

There is no general method of updating sampling frames, although this question is discussed in detail in many handbooks. The methods and possibilities of adjusting sampling frames to changes in time are mainly determined by the specific features of individual countries such as reliability of available registers and the amount of resources. Needless to say that the aim of updating is to provide good representation of the population by the sample of each actual survey.

Missing data, non-response

The term «missing data» refers to losses of information due to unsuccessful observation of some units (families, persons, etc) of the population; the losses may be partial or total. There are different reasons for missing data.

If the sampling units are dwellings, the cases that nobody is at home or only some persons are absent (not everybody) are equally possible. Even if everybody stays at home, it can happen that they all refuse cooperation, in other cases some will be willing to cooperate, while others from the same dwelling not. Some persons may be unable to answer for different reasons such as age, influence of others, mental deficiency, etc. It is easily seen that non-response quantitatively depends both on the individual sampling units and the method of data collection. Literature on this subject in America and Western Europe devote considerable space to telephone interviews and mailed questionnaires, since the problem of non-response by these methods is quite different from that in the case of face-to-face interviews.

One of the important tasks of designing surveys is to specify for each question those persons who are competent to answer it. When e.g. a family is questioned, there may be questions concerning facts that are known to each member of the family. On the other hand, answers about plans, opinions and attitudes may be accepted only from persons actually concerned.

An important indicator of the success and reliability of a survey is the response rate which is defined as the proportion of all successful interviews to total sample size. When units of the sample have been selected by different probabilities, the response rate should also be determined as a weighted ratio.

For a proper treatment of non-response it is useful to know the reason for failure; as was pointed out above, this can be refusal, absence, wrong address, disability of some persons interviewed, etc.

Given the accuracy of estimates, the sample should meet the following conditions: proper size, suitable structure and good coverage of populations in consideration. Missing data lead to increase of sampling errors, but they also involve uncertainties which bring us to the alternative class of non-sampling errors. It seems that the latter can best be treated by the imputation of missing data, and this is why imputation techniques are so widely used in sample surveys.

There are two possibilities of eliminating or reducing non-response errors, namely,

- minimizing the number of refusals in the phase of data collecting,
- imputation of data missing for not-at-home incapacity or other reasons in the phase of data processing.

Answering the questions may involve burden, inconvenience or just come efforts for the respondents: this topic will be discussed in detail in connection with non-response errors. The success of data collecting depends heavily on the skill of interviewers, in particular on their ability to persuade people to cooperate; the length of questionnaire and the type of questions play an important role too in this respect. To sum up, the field work and the training of interviewers are decisive factors of the rate of refusals.

The size of losses due to other factors can be reduced by repeated calls, careful design and by efficient organization.

In many surveys some refusals can be expected even in the case of the most careful design. If refusals were distributed uniformly among different parts of the sample, their effect would be only tantamount to some decrease of the sample size, which in this case could be expressed unambiguously in terms of (increasing) sampling errors.

In the course of data processing there are different ways of adjusting or imputing incomplete data; in any case, however, the following conditions should be taken into account:

(1) Leslie Kish: Multipurpose Sample Designs, Survey Methodology, Vol. 14, No. 1 (1988) pp. 19-32.

- the distribution of refusals among different strata and subgroups is not uniform, and
- information on missing or refusing persons is usually insignificant.

Reliability of the estimates can be improved by the use of poststratification, if proper characteristics and auxiliary information are available. In this case weights compensating for missing units can easier be calculated.

By imputation we mean replacing missing data by suitable estimates. If some conditions are met, this can be done by

- auxiliary information,
- cold-deck methods referring to the use of data from some similar but earlier survey, and
- hot-deck methods based on data already known from the current survey.

When using these methods it is essential to have some knowledge (number of members, or occupation, age, sex, social stratum, etc.) of units (families, persons) that provided incomplete information. This is the condition of substituting missing data by similar ones (1).

Many survey statisticians think that there are two extremes in connection with missing data: the best thing is to call back and get the missing information from the very person whom it belongs to, and the worst thing is to do nothing.

Response error

In the course of data collecting some pieces of elementary information may be wrong for several reasons. The difference between a given answer on the questionnaire and the corresponding «true» value is called response error. Data collection and data processing are sometimes difficult to distinguish, since e.g. the wrong code of a correct answer results in response error. In what follows it will be useful to declare the phase of data collection finished when completing the questionnaires, checking and editing are all over, and the dataset of the survey is available e.g. on a magnetic tape. In this sense response errors include the distorting effects of imputation too.

(1) For details see Graham Kalton: *Compensating for Missing Survey Data*, The University of Michigan, ISR, SRC 1983; Hakan Lindström: *Non-Response Errors in Sample Surveys*, Urval, No. 16, Statistics Sweden, Stockholm, 1983.

From theoretical point of view imputation is closely related to micro-simulation methods and statistical matching.

It seems trivial to assume that response errors are mistakes made by the respondents. This is, however, not acceptable: response errors depend —as we have seen above— heavily on the questionnaire, the way of asking questions, the interviewer and his/her training, the code instructions, efficiency of imputation, etc.

It is evident that accuracy of the answers depends on simplicity and clarity of the individual questions, i.e. on the style of questionnaire. Designing good questionnaires is rather hard; some people claim that it belongs to the category of arts, which is not always recognized. The methods of designing questionnaires will not be discussed in this paper (2), only a brief review of the different types of questions will be given, as this is very important from the point of view of quality of answers.

To begin with, we can distinguish between

- open and
- closed questions.

In the first case the respondent composes his/her answer in his/her own words, usually without any help of the interviewer. In the second case the possible answers are all given, and the respondent has to select of them the proper one, which fits his/her state or opinion. In many situations the possible answers are augmented with the additional category «Other(s)», which —if proper— should again be completed in the respondent's own words.

The questions can also be classified by their content, which may refer either

- to facts and events, or
- to opinions and attitudes.

In the first case there is always some objective true answer implying a unique value of response error. On the other hand, the experience of re-interviews shows that there is in general no unique answer to questions concerning opinions and attitudes, thus in this case response error can only be defined in terms of variability due to repeated questioning.

Correct answers can be expected from the respondent only if the following conditions are fulfilled:

- the respondent is willing to cooperate,
- he/she understands the question, and

(2) For a detailed treatment see R. Platek: *Some Important Issues in Questionnaire Development*, JOS Vol. 1, No. 2 (1985), pp. 119-130, and W. Sykes —J. Morton— Williams: *Evaluating Survey Questions*, JOS Vol. 3, No. 2 (1987), pp. 191-207.

—he/she is in possession of the information requested and is able to remember it.

From what has been said above it follows that besides the type of question the interviewer's responsibility is also heavy for building up a creative connection between himself/herself and the respondent (1).

The connection between interviewer and respondent (or the members of family called on) is very specific, in certain sense it is a byproduct of the interview. Though the respondents are informed on the purpose of survey as well as on its public benefit, it would not be wise to assume that they are interested in it. It is more realistic to assume that the drawbacks of being involved is more obvious for the respondents than the advantages; for them the interview may appear simply as waste of their time. Some questions may be definitely delicate, and efforts must be made to produce good answers.

The interview establishes a specific contact between two people unknown to each other; its purpose is to accomplish some common job. The interview is in general not too long; being impersonal by its nature, it is still governed by accepted norms of communications.

The interviewer's role is of key importance. The line of talk as well as its end is determined by him/her; in the case of obscure points he/she helps the respondent. It is logical to regard the interview as a tool by which verbal information can be obtained from the respondent; the way to this end is marked by the questionnaire.

Success of an interview is affected by numerous factors; in the following we shall consider those which depend on the type of questions.

Simplicity, logical structure and clear style are the properties which guarantee good, accurate answers to our questions. In certain cases, however, past events should be recalled in the respondent's memory. In such situations power of memory of the respondents play an important role; it is supported by different reminder notes such as e.g. the household diary. Recently research has been done on the basis of cognitive psychology which investigates the process of cognition (2). The most important results are contained in proceedings of the conference.

Experience shows that people do not like to admit things which make them be seen at disadvantage. They usually tend to blunt the edge of questions which are unpleasant or embarrassing for them; on the other hand, they try to give answers complying with expectations of society. Besides, they do not like to oppose the interviewer; they prefer to agree and readily accept answers proposed. Bias in the answer may be considerable if the question refers to some form of behaviour or activity which is gravely condemned by society or nearly breaks the barriers of law. In such situations tension is induced in the respondent, and not only the current answer will thereby be affected, but the rest of the interview too. (This fact is considered in conducting the interview, which —accordingly— may be soft, neutral or hard. This issue will not be discussed in this paper. Household or population surveys of statistical offices belong in general to the neutral category, which is characterized by polite, matter-of-fact and friendly atmosphere lacking both brutality of hearings and exaggerated friendliness of intimate relations.)

Remembering some past event is greatly influenced by its significance or complexity. It is obvious that everybody will accurately remember events of great importance such as e.g. buying a car or undergoing an operation for gall-stones. Note also that the success of the interview is affected by the level of intelligence and education too.

Now it is easily seen that response errors are greatly influenced by the interviewer too, the importance of his/her fitness and training is self-evident.

There is considerable interaction between interviewer and respondent: their role in the interview is equally important and is determined to a great extent by their intellectual power (3).

Considering the theoretical model of response error we see that this error depends only to some extent on the respondent, as the interviewer's work may also have considerable impact on the accuracy of information obtained.

It is worth while to consider the implications of what has been said on response error to some problems arising in the computation of price indexes.

Computation of a price index —no matter

(1) See e.g. S. Sudman-N.A. Bradburn: *Response Effect in Surveys*. Aldine Publishing Co., Chicago, 1974; S. Payne: *The Art of Asking Questions*. Princeton University Press, 1951.

(2) In the United States research is underway at large scale to discover the best methods of guiding the respondent towards the right answer. For details see *Questionnaire Design: Report on*

the 1987 BLS Advisory Conference. US Department of Labor Statistics, August 1987 (Manuscript.)

(3) This topic is dealt with in detail in: *The Influence of Interviewer and Respondent. Psychological and Behavioral Variables on the Reporting in Household Surveys*. National Center for Health Statistics, Ser. 2, No. 26, Washington D.C., 1968.

what type: producer, consumer or foreign trade—is always based on a sample. Laying the hard problem of sampling errors aside we turn our attention to the less complex problem of computing the price index provided that some products, goods and services are selected for this purpose and the corresponding weights are also given.

Given the sample we face two important problems, both having essential impact on our final result. First, we have to decide on the way of determining individual price indexes for products or goods called representative items in this context, in other words: on the method of calculating the pairs of prices whose ratio gives the individual price indexes. The second problem concerns the weights, i.e. the formula to be used.

Assortment, brand, quality, manner of packing, etc., of representative items selected for observation may be rather varied, hence even individual price indexes may heavily depend on the specification, assortment, etc., of products. Definition and specification of representative items is a well-known problem for experts of price statistics, as well as the fact that the «fuzzy» character of definitions, i.e. feasibility of changing certain elements has a perceptible impact on the change of price level. One part of the latter is pure price change, while the other part—which can hardly be determined—is due to quality change. Compromises must be made in order that price indexes may be computed and products may be compared: changes in the composition must not be treated very seriously, or else no price index can be given for a lot of products. This means that—as a consequence of the joint effect of the two factors described before—some response error occurs in the individual price indexes. This kind of response error is in general not systematic bias, it is rather of random character, particularly if the computation proceeds under real market conditions. The situation is different in economies with acute shortage: in that case a strong but hidden tendency of increasing prices prevails, which results in positive bias.

The second problem comes from weighting. It is well known that several formulae are available for computing price indexes and their time series, and the different formulae result in different values of price indexes for the same set of prices and volumes. The debate on index formulae has lasted for decades, and it is natural that there is no ideal formula acceptable by everybody under any circumstances. The variability among the different formulae can be regarded in certain sense as a kind of response error, since the problem of the unknown price index is well defined, the solution ought to be unique, and still we are not able to give a unique answer. (In practice some of the different formulae is adopted, and the unique price index obtained in this way is used.)

As we have seen above, the notion of non-

sampling errors reveals new aspects of some old problems concerning price indexes. Computational results are thereby not affected, but the new approach may be useful in many cases.

Correlated response variance (1)

Although the concepts underlying correlated response variance are straightforward, its mathematical treatment is relatively complex. The objective of this exposition is to give a general idea of what it is, how it can be measured, and its implications for the design of household surveys. Readers seeking a more rigorous mathematical treatment may consult Fellegi (1964, 1974); Hansen, Hurwitz and Bershad (1961); Kish (1962); and U.S. Bureau of the Census (1968).

The nature of correlated response variance associated with interviewers was described by Kish (1962) as follows:

«Each interviewer has an individual average 'interviewer bias' on the responses in his workload, and we consider the effect of a random sample of these biases on the variance of sample means. The effect is expressed as an *interviewer variance* which decreases in proportion to the numbers of interviewers.»

A similar statement could be made for components of correlated response variance associated with supervisors, editors, coders, key-punchers, etc.

An important property of the interviewer variance is that its contribution to the total mean square error is inversely proportional to the number of interviewers employed in the survey. For a fixed total sample workload, the larger the number of interviewers, the smaller will be the effect of interviewer variance on the results. Likewise, the effect of correlated errors introduced by coders would depend on the number of coders employed.

Other things being equal, this suggests that the ideal survey design would be to have a different interviewer for each respondent! In a sense, this is what happens when self-enumeration is used, as has been the case in recent years for censuses in several developed countries. In fact, experiments which demonstrated the effect of interviewer variance on census results had much to do with the trend toward

(1) See: National Household Survey Capability Programme: Non-sampling errors in household surveys: sources, assessment and control.

United Nations Department of Technical Co-operation for Development and Statistical Office. New York 1982. pp. 155-159.

greater use of self-enumeration (see, for example, Fellegi and Sunter, 1974).

However, self-enumeration is not a feasible alternative for most household surveys, and cost factors and other considerations place upper limits on the number of interviewers that can be used in a particular survey. Nevertheless, it is important to be aware that just as sampling errors may be quite large for estimates based on a small number of primary sampling units, the effects of interviewer variance can be substantial for estimates based on the work of a few interviewers. This is a major consideration in deciding whether a survey is adequately designed to produce subnational estimates, say for regions or provinces. If interviewers are not uniformly well trained and supervised, interviewer variance can overwhelm other sources of error in subnational estimates, even though its effects may be small at the national level.

The contributions to total error of correlated response variance may or may not be included in the usual estimates of sampling variance. Household surveys generally use multi-stage or clustered samples, hence, the sampling errors are also «correlated» in the sense that individuals coming from the same area or cluster tend to be more alike than do individuals in the population at large. Taking several individuals (or households) from the same cluster reduces the independent information contained in the sample; in other words, it increases the sampling variance. Practical methods of computing sampling variances are generally based on comparing the values obtained from different clusters in the sample: the larger the differences between the clusters, the higher the sampling error will be. (In addition, of course, the size of the sampling error depends inversely on the number of clusters in the sample.) In an analogous way, survey workers, say interviewers, impose their own «clustering» effect on the survey results insofar as errors in responses of individuals interviewed by the same interviewer tend to become somewhat similar due to the influence of that particular interviewer. If the survey arrangement is such that each interviewer is assigned to work in only one sample cluster, then the effect of interviewers will be completely confounded with the effect of clustering on the sampling variance, and usual methods of estimating sampling variance will automatically include the correlated interviewer variance. By contrast, if all interviewers worked as a team in each cluster, or if the interviewers' workloads were assigned at random, the usual estimation of sampling variance would *not* include the effect of additional variability due to the interviewers. The

effect may be partly included with other intermediate arrangements for assigning workloads to enumerators (see Verma, 1981).

To make *separate* estimates of interviewer variance or other types of correlated response variance, it is necessary to introduce some degree of randomization or interpenetration of workloads assigned to the particular category of survey personnel involved. For example, in a recent study of interviewer variance in a crime victimization study conducted in 8 cities in the United States, the design was as follows (Bailey, Moore and Bailar, 1978): I) the sample workload in each city was divided into 18 crew-leader assignments, each containing 8 interviewer assignments from a total of 144. Each interviewer assignment consisted of about 80 households II) within each crew-leader area, 4 pairs of interviewer assignments were formed; III) interviewers were assigned to interviewer assignment areas close to their homes; IV) 36 of the 72 assignment pairs were selected at random. Within each of these 36 pairs, housing units in each assignment were assigned in an alternating systematic pattern to the two interviewers. In other words, the workloads of the two interviewers assigned to a particular pair of assignments were *interpenetrated*.

This design made possible the estimation of the interviewer variance, following a method proposed by Fellegi (1974), using data from both sets of paired interviewer assignments — those with interpenetration and those without it.

Randomization of assignments generally increases the cost and complexity of survey operations. In practice there are many constraints which limit the degree to which complete randomization can be effected, though in multistage surveys it is frequently possible to select two «interpenetrating samples» within each area and assign them at random to two different interviewers who have been assigned to work in that area. Fortunately, useful results can be obtained even when complete randomization is not possible (Kish, 1962.)

So far, there have been only a few studies to obtain separate estimates of interviewer variance and other components of correlated response variance in household surveys. More studies are needed; without adequate information about this important component of error effective application of the principles of total survey design is not possible.

Specific *measures of correlated response variance* are presented in connexion with the examples which follow

Estimates of response errors

In theory nice models can be built for response errors and expressions for their variance can also be developed. For practical treatment of response errors a methodological survey based on postevaluations and re-interviews is needed. Conducting a single postevaluation does not provide estimates for response errors immediately, still it yields some indicators measuring the fitness of answers.

With reference to what was said above, the following facts should be taken into account when post-evaluations are conducted to investigate response errors:

- in the case of certain questions people usually strive for responses giving a favourable account of themselves;
- answers to questions on opinions, plans and attitudes may contain considerable uncertainty originated from (mis) interpreting the questions, lack of a definite standpoint, factors depending on the atmosphere of interview;
- answers to delicate questions are always uncertain;
- reminiscence of past events always contains some bias.

There is a wide range of income and consumption data (consumption of alcoholic beverages, income from principal occupation, attendance of cinemas and theatres, etc.) where comparison with extraneous information is possible at aggregate level. Thus we have a possibility of estimating response error of aggregates. Unfortunately, there is in general no way of making inference from the bias of some aggregate to that at lower level, e.g. at the level of units. Suppose e.g. that we have found that on the average observed consumption of some goods is 10 p.c. less than the actual value; the inference that every respondent admitted some consumption (purchasing) exactly 10 p.c. less than what he/she had actually purchased would obviously be false.

In Hungarian censuses of population as well as those of livestock are usually evaluated by means of re-interviews(1). This happens by repeating some part of the survey in order that some information on accuracy of answer (to questions concerning age, occupation, school qualification, etc.) may be obtained. Experience of these re-interviews shows

that even in the case of such «core data» response errors may amount to several percentages.

Biases coming from re-interviews may (and ought to) be reduced to minimum by employing different interviewers at the first and the second (the third, fourth, etc.) calls. It is very useful if repeated calls are made by the best interviewers or by subject-matter statisticians, as particular skill is needed to persuade people to consent to a second (third, fourth, etc.) interview. Repeated surveys are suitable also for examining the interviewer's effect on quality of data. Such repeated surveys are to be designed very carefully, and only a few cases are reported on in the literature.

A theoretical model for estimating response errors can be outlined as follows (2).

Suppose we are given a variable having a unique true value. Individuals in the sample are repeatedly observed, and the succeeding observations are independent. The response obtained from individual i at interview (or observation) α is denoted by $y_{i\alpha}$ which can be written as follows:

$$y_{i\alpha} = \mu_i + e_{i\alpha}$$

where μ_i and $e_{i\alpha}$ stand for true value and response error, respectively.

The distribution of $e_{i\alpha}$ can be estimated on the basis of repeated observations. Let us denote expectation and variance of $e_{i\alpha}$ by β_i and σ_i^2 , respectively. $e_{i\alpha}$ represents random error if $\beta_i = 0$ and systematic bias otherwise. the bias β_i itself can be regarded as a variable unless it equals the bias β pertaining to the whole sample.

The random component of measurement error is given as $d_{i\alpha} = e_{i\alpha} - \beta_i$; expectation of this component is equal to 0. Setting

$$\mu_i' = \mu_i + \beta_i = E(y_{i\alpha}|i)$$

where μ_i' is the mean of observations on individual i we have

$$d_{i\alpha} = e_{i\alpha} - \beta_i = y_{i\alpha} - \mu_i'$$

For a given sample the random errors $d_{i\alpha}$ are in general not uncorrelated (a good example can be afforded by clustered samples). The correlation

(1) The 1970 Census of Population. Vol. 31: Survey Methodology and the Method of Data Processing (in Hungarian). Central Statistical Office, Budapest, 1977. Similar controls were used in connection with the 1980 Census of Population in Hungary too.

(2) We borrow here some ideas and notations from W.G. Cochran: Sampling Techniques, John Wiley, New York, etc., 1977, Chapter 13. See also: Hansen, M.H., Hurwitz, W.N., and Bershad, M.: Measurement Errors in Censuses and Surveys. Bulletin International Statistical Institute, Vol. 38 (1961) No. 2, pp. 259-374; Fellegi, I: Response pp. 1016-1041.

coefficient ρ can be defined by the following formula:

$$\rho \sigma_d^2 = E(d_{ix} d_{jx})$$

total variance of response can be written now in the following form:

$$V(\bar{d}_x) = \frac{\sigma_d^2}{n} [1 + (n-1)\rho]$$

where σ_d^2 denotes the population average of the variances of measurement errors. For a sequence of observations the mean square error (MSE) including also sampling errors can be expressed as

$$MSE(\bar{y}_x) = \frac{1}{n} \{S_{\mu}^2 + \sigma_d^2 [1 + (n-1)\rho]\} + \beta^2$$

Here the sampling variance S_{μ}^2/n and the term $\sigma_d^2(1-\rho)/n$ decrease as the sample size n increases, but the terms $\rho \sigma_d^2$ and β^2 depend on n in a rather different way. In the case of large samples sampling errors will be negligible as compared to response errors which will represent the dominant part of MSE. (This expression does not reflect the uncertainty originated from deficiencies of coverage and from non-response.)

If we have at least two observations for some variable(s) on some part of the sample, the above formulae can be used and response errors can be estimated (1). In the case of two observations total variance of response can be estimated by the following formula:

$$V(d) = \frac{E(y_{.1} - y_{.2})^2}{2}$$

By some simple modification of the above model the interviewer's effect or other impacts on response errors can also be studied.

Samples of postevaluation surveys are usually subsamples of the original surveys (2); these should be of proper size and should ensure good representation from the point of view of characteristics in consideration.

The proportion of response variance to total variance is called the index of inconsistency:

(1) Within the frames of USHS one part of the Survey on Prestige of Occupations in 1988 was repeated on a subsample of 560 units (persons). This was selected with simple random method from the sample of the original survey. The second call was made in all cases by a «fresh» interviewer. Evaluation is underway.

(2) Postevaluations are rather expensive, and increase the burdens of respondents. The size of the corresponding samples is therefore restricted to some hundred units.

$$I = \frac{\sigma_d^2}{\sigma_d^2 + S_{\mu}^2}$$

(The denominator equals total variance if random errors in the sample are uncorrelated.)

In the case of categorical data consistence (or in other words: simple reliability) can also be examined by means of a square matrix whose rows and columns contain feasible categories of answers obtained as a result of the two calls. The information p_{ij} in cell (i,j) indicates (absolute or relative) frequency of the event that somebody whose answer pertained to category i at the first call specified category j with his/her second answer. The number of consistent cases in the individual categories is indicated by p_{ii} ($i = 1, 2, \dots$) (3).

It is evident that no response error can be registered if the same answer is obtained to each question at both calls. At the other extreme, if in each case different answers are obtained to the same question, then uncertainty is maximal, and it is very likely that no useful information can be expected by the method of measurement adopted.

Distribution of answers

Re-interview		1 2 3 ... n				
First interview						
A	1	P_{ij}				$p_{1.}$
n	2					$p_{2.}$
s	3					$p_{3.}$
w	.					.
e	.					.
r	.					.
s	n					$p_{n.}$
		$p_{.1} p_{.2} p_{.3} \dots p_{.n}$				

Using standard notations, the dot (.) refers to taking the sum over all values of the index it replaces:

$$\sum_i p_{ij} = p_{.j}$$

If the p_{ij} 's represent proportions, the raw index of consistence (A) is given as follows:

$$A = \sum_i p_{ii}$$

(1) For details see C.A. O'Muircheartaigh: The Magnitude and Pattern of Response Variance in the Lesotho Fertility Survey. ISI Scientific Reports, No. 70, december 1984.

In this case the measure of deviations is

$$D = 1 - A.$$

The index A has two deficiencies:

- consistence or inconsistency may be the consequence of random factors too;
- the index A gives no information about the distance between the two observations. (This problem does not arise in 2×2 contingency tables corresponding to the dichotomy of answers «Yes» and «No».)

Inference to consistence due to random effects may be drawn from the two marginal distributions. It is useful to define measure of consistency (k) in the following way:

$$k = \frac{\text{deviations observed}}{\text{deviations expected}} = 1 - \frac{1 - p_o}{1 - p_e} = \frac{p_o - p_e}{1 - p_e}$$

where

$$p_o = \sum_i p_{ii} \quad \text{and} \quad p_e = \sum_i p_{i.} p_{.i}$$

The index k can also be written in the alternative form:

$$k = \frac{\sum (p_{ii} - p_{i.} p_{.i})}{1 - \sum p_{i.} p_{.i}}$$

This measure of consistence is particularly useful in the case of skew distributions, though the presence of some category dominating all others may be misleading even with this index.

Interpretation of deviations in the cross table depend on the nature of variables in consideration. If some categorical data are examined, then every piece of information contained in off-diagonal cells may be regarded as wrong. On the other hand, off-diagonal information can well be interpreted in the case of measurable variables such as e.g. income. In such situations the measure of consistence can be computed e.g. under the condition that deviations up to 1-2 categories are feasible.

It is worth while to pay some attention to the case of dichotomy of categorical variables. Consider the following 2×2 cross table.

First response	Second response		
	1	0	
1	a	b	a + b
0	c	d	c + d
	a + c	b + d	m

Suppose that $\sigma_{d_1}^2 = \sigma_{d_2}^2$ and that d_{i1} and d_{i2} are uncorrelated. Then we have

$$\sigma_d^2 = \frac{\sum (y_{i1} - y_{i2})^2}{2m} = \frac{b + c}{2m}$$

and the index of inconsistency is given as

$$I = \frac{\sigma_d^2}{\hat{P}\hat{Q}} = \frac{m(b + c)}{(a + b)(c + d) + (a + c)(b + d)}$$

If $g = \frac{b + c}{m}$ is the proportion of deviations (expressed in percentages), then $g/2$ is a good measure of simple response variance.

Some practical problems

As we have seen previously, the impact of some factors on the increase of non-sampling errors can in principle be estimated, while in other cases the size of the factors themselves can be determined by some method, but not the size of their effect on errors of data. In particular, the effects of deficiencies of coverage as well as those of refusals ought to be investigated by simulation experiments or by some analytic methods, by which missing data can somehow be replaced.

If a sample is selected from a well defined population, and we may assume that

- the target population is completely covered,
- each unit of the sample can be observed and each observation is accurate, and, finally,
- a well defined estimation technique is used,

then an estimate \hat{y} is obtained whose deviation from the corresponding population parameter can be estimated by the standard error (which is S/\sqrt{n} in the case of estimated means).

Due to non-sampling errors, in practice some estimate \tilde{y} will be obtained instead of \hat{y} . As was shown in the previous chapters, total variability of \tilde{y} can be determined in terms of the two variances S^2 and σ^2 .

The non-sampling variance σ^2 depends partly on some deficiencies (coverage, non-response, etc.), partly on inaccuracy of individual observations (response errors). Considering individual observations, one part of non-sampling errors — a considerable part in general — can be estimated, the other part remains unknown.

The model of response error outlined above is based on possible inaccuracy of individual responses. The behaviour of e_{ix} forms the mathema-

tical basis of proper treatment of variables and individual observations. Nevertheless, two situations should definitely be distinguished in practice.

In the case of categorical data (e.g. male-female) it is not meaningful to speak about the deviation of observed value from the actual one; we can only say that some answer is good or bad. In other cases the measure of accuracy of an answer obtained may better be defined. In other words the deviation e_{ix} can be interpreted and regarded as a measure of response error.

«Goodness» of categorical data can be examined by contingency tables. In this way some probability can be assigned to any actual answer which—in view of general conditions of the survey—can be regarded as accurate with this probability. In other cases an interval covering the actual value y_i with some given probability can be determined.

Interpreting response errors pertaining to estimates \bar{y} (which are usually proportions in the case of categorical data) can always proceed in the same way, no matter what types of variables are considered.

In contrast with categorical data, measurable variables have the property that some deviation between some answer and the corresponding actual value is not necessarily a mistake making the information meaningless. It is e.g. quite natural if somebody cannot remember the exact figure of his/her income. It can happen, however, that the deviation between answer and actual figure is too large, and then it is proper to check if there is some misunderstanding, or some deliberate distortion or some gross error in the usual sense. Though our model makes no distinction between small and large errors, sometimes it is worth while to check if the answer is acceptable from the point of view of the question.

The main task of editing data is to detect contradictions, i.e. codes that are infeasible by logical considerations, furthermore, to correct such codes or data if possible. Evaluating the responses from the point of view of our model we may conclude that extreme answers, outliers ought to be studied separately. (In the case of comparing occupations by prestige there are e.g. a lot of answers which would contradict any kind of reasoning.)

There are methods of measurement and investigation that would accept a wide range of personal judgements, still it is useful to query the feasibility of some answers. The following attitude is well known: people who ought to answer «I don't know», «I can't catch it» or «I prefer to refuse» would rather say something else which is often just the opposite of what they actually think.

Different experts have different views on this question. Some believe that no answer may be ignored, others think that extreme answers should be neglected.

The connection between response error and price statistics has been discussed in a previous chapter. Recall that improper composition of products and goods is a possible source of serious errors. Without going into details we note that checking extreme price changes has always led to improper compositions, and omitting too low and too high individual prices has always produced good results.

The above conclusion applies also to surveys of social statistics: extraordinary responses ought to be examined separately here too, and if they prove to be absurd, they should be ignored.

Finally some features of the USHS will briefly be summarized.

The census enumeration districts are steadily modified by new construction and demolishing old buildings; this is the basis of updating the sampling frame. As the size of the USHS sample is fixed, expansion factors (i.e. the factors or weights used in expansion—in other words: inflation—estimators) in the strata of the sample are computed as the proportions of the updated numbers of dwellings in the population to the numbers of dwellings observed in the sample. This procedure ensures automatic adjustment for non-responses in each stratum; it can be regarded as a specific «hot-deck» imputation which uses stratum mean in each stratum of the sample. The rate of refusals in USHS surveys is in general not significant, it is usually well under 10 %. The rate of non-response due to other factors is greater and shows an increasing tendency; it amounts to nearly 10 %. An important factor in this respect is that there are many empty dwellings.

In the phase of processing and checking questionnaires missing data are replaced by some proper imputation technique, whereby problems of partial non-response are eliminated. This means that in the USHS completed and checked questionnaires are regarded as accurate.

Aggregates from surveys are usually compared with extraneous information and macro data; this comparison leads sometimes to proper adjustment of survey data.

Censuses in Hungary are always accompanied by postevaluation studies; this is, however, not typical of the USHS. The postevaluation of the survey «Prestige of Occupations in 1988» is an exception. The sample of postevaluation consisted of 560 per-

sons, the questionnaire was an abridged version of the original. As was mentioned earlier, analysis of the results is underway.

The surveys are conducted by a constant network of interviewers. The latter are well trained and equipped with considerable knowledge of their field; most of them are women. One-third or a half of the sample is changed in a year, thus every family in the sample should be called on several times; topics of succeeding interviews are not necessarily the same. As a result of repeated calls certain contact develops between families and interviewers, particularly in smaller settlements. This contact is definitely advantageous, it has a favourable impact on the respondents' willingness to cooperate with the interviewers.

Sampling error: portability

A sample covers only part of the universe under inquiry. The sampling error shows the accuracy of the \bar{y} supposing that it is unbiased and there is no other (non-sampling) error. The sampling error is a function of the sampling design: sample size, type of stratification, allocation of the sample to the strata, clustering, estimation procedure.

The sampling variance of an estimate is a measure of how the estimate would be expected to vary over repeated sample selections. Fortunately it can be estimated from the results of a single sample. It decreases with the increase of sample size.

A given sample design with a given sample size is good only for certain aggregates, cross-classes, subgroups, territorial break-downs.

Bias versus unbiasedness: Sometimes it is better to have a biased estimate with small variance than an unbiased with great variance (E.g. small areas estimates: synthetic, SPREE, etc.)

Mean square error reflects the combined effect of sampling variance and bias.

$$ME = \sigma^2 + B^2$$

Coefficient of variation:

$$CV = \frac{\sigma}{\bar{y}}$$

The sampling variance depends on (approaching it from the opposite direction) the aim of the survey and also the resources available.

A given sample is good for one thing, but not for another.

The goal determines the size, design etc. But most surveys and samples are multipurpose.

«Conceptually, the magnitudes of the various components of the mean square error are functions of the design of the survey. Thus, we may modify the sampling variance by introducing stratification, changing the size or nature of the sampling units at the various stages of sampling, or by other well-known devices. Similarly, we may modify the response variance components and the coding variance components by changing the form or content of the questionnaire, the methods of recruiting, training or supervising the interviewers, the procedures for editing or coding responses, or other aspects of the survey. Each change alters the mean square error of the results and also the cost of the surveys. We have the ultimate goal of seeking those changes which minimize the mean square error attainable with given resources or, alternatively, which minimize the resources required to attain a specified mean square error.» (Jabin and Tepping, 1973.)

Sample surveys and statistical publications: Many variables and technical difficulties in calculating and publishing sampling errors. The best solution would be: to have a few figures and based on them the cv could be calculated or at least well estimated: this is called portability.

In case of simple random sampling the sampling error of a given variable depends on the sample size n . But this is not true in case of complex samples.

The problem of portability, the role of sampling design was investigated deeply in WFS (1).

Design effect or design efficiency factor measures the efficiency of the design as compared with what it would have been if the sample was selected entirely at random.

The main factor effecting the design efficiency is the clustering. If there is a correlation inside the cluster the amount of information from a cluster of n_k individuals is not n_k times the amount of information derived from one individual. In practice one may expect a positive correlation and this will reduce the efficiency of sampling.

«Computation of sampling errors for surveys are generally multipurpose in two ways: First, they

(1) L. Kish, R.M. Groves and K.P. Krotki: Sampling Errors for Fertility Surveys. WFS Occasional Papers, No. 17, January 1976.

V. Verma, C. Scott and O'Muircheartaigh: Sample Design and Sampling Errors for the World Fertility Survey. J.R. Statist. Soc. A (1980), 143, Part. 4, pp. 431-473.

concern many statistics for many variables; second, they can serve several different needs. Computing sampling errors involves more than variances and standard errors. Design effects (deff^2), ratios of homogeneity (roh) and coefficients of variation are useful and used; from these, averages and other functions can be computed also, and they often are.

Design effects, the ratios $\text{deff}^2 = \text{actual/srs var}$, have several important uses. 1) They may be averaged for greater stability, when the computed variances are subject to great variation because they are based on few primary units, or «degrees of freedom». 2) They can and should be used to check for gross errors in variance computations. Gross errors are the most common and easiest to spot for deviations from the base of 1. 3) Their main purpose is in models and conjectures for other statistics from the same survey. But for this purpose the function $\text{roh} = \text{deff}^2/(\bar{b} - 1)$ is preferable, especially for cross-classes. 4) They may be «borrowed» to serve in conjectures about sampling errors for other surveys. 5) They may be used for designing other surveys. (L. Kish, *Statistical Design for Research* 1987, p. 202.)

Consider the sum of a number of variables x_i to x_n , all for some of which may be correlated. Its variance is given by

$$\text{var}(\text{Sum}) = \sum_{i=1}^n \text{var } x_i + \sum_{i \neq j} \text{cov}(x_i, x_j)$$

If the variances of x_i are all equal then

$$\frac{\text{var}(\text{Sum})}{\text{var } x_i} = n + n(n-1)\rho$$

where ρ is the average correlation among the x 's. Thus

$$\frac{\text{var}(\text{Sum})}{n \sum \text{var } x_i} = 1 + (n-1)\rho$$

The expression on the left is the deff of this particular cluster and hence

$$\rho = \frac{\text{deff} - 1}{n - 1}$$

Where the whole design is under consideration, comprising various clusters, one may take an average of the cluster size and write

$$\text{roh} = \frac{\text{deff} - 1}{\bar{b} - 1}$$

where \bar{b} is the average size. The deff is calculated from the data by formulae set out in the text and hence roh is computed. If there is no intraclass correlation roh is zero and the deff is unity. Theoretically roh can attain unity, in which case the deff will be high (an inefficient design, as is otherwise obvious from the fact that all the members within a cluster give identical answers to the variable concerned). In practice values of roh between zero and 0.2 can be expected.

Sometimes the roh and deff can be estimated beforehand. The roh is better, because it does not depend on the cluster size. That is the reason why it is called «portable».

It is an open question whether the deff s obtained from individual variables can be amalgamated into a single index to give some idea of the deff as a whole. Generally there is no deff or variance of the design only of a variable!

The following indirect method (1) of imputation can be used from a computed standard error (ste_0) to an unknown one (ste_i):

$$\text{where } \text{deff} = \sqrt{\text{deff}}$$

We impute across roh values because of their relative stability across diverse subclasses for each variable from a sample, and also for similar variables across samples. We transform values of ste into deff , and these into roh ; then, after imputing roh for a new statistic, we transform into the new deff , and finally to the needed ste . The direct imputation from ste_0 to ste_i is seldom justified. The path from deff_0 to deff_i is usually difficult also, due to large differences in sample sizes, hence cluster sizes, for diverse subclasses. (Randomly distributed variables have roh values near zero, whereas highly clustered items are found near 0.10 or 0.20, and perfectly segregated variables can theoretically approach 1.0.)

When using complex samples the basic question is: what is the trade-off between the loss of efficiency and the reduction of costs.

(1) To demonstrate, how portability works, see Appendix ch. 6.

REFERENCES

- ANDERSON, R., J. KASPERK, M.R. FRANKEL, et al.: Total Survey Error. San Francisco: Jossey-Bass, 1979.
- ASAI, A.: Problems of Survey techniques in Japanese official statistics. Invited paper 46th Session of the ISI, 8-16 September Tokyo, 1987.
- ATKINSON, J.: A handbook for interviewers. Her Majesty's Stationery Office, London, 1979.
- BAILAR, B.: The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*. Vol. 70. pp. 23-30. 1975.
- Rotation sampling biases and their effects on estimates of changes. 43d session of the International Statistical Institute, Manila. 1979.
- BAILAR, S.A.: Information needs, surveys, and measurement errors. Presented at the International Symposium on Panel surveys. 1986.
- DEL BARRIO, L.A.: Vital statistics evaluation in Spain. Seminar on Statistical Methodology. U.N. CES, Geneva, 1-2 February 1988.
- BRADBURN, N.M. and SEYMUR AUDMAN: Response effects in surveys. Aldine Publishing Company, Chicago. 1974.
- BRADBURN, N.M. and SEYMOUR AUSMAN: Improving Interview Method and Questionnaire Design. San Francisco: Jossey Bass. 1980.
- BETHLEH M, J.G.: The data editing research project of the Netherlands Central Bureau of the Census. U.S. Department of Commerce, Washington D.C. 1987.
- CAMBERS, R.: Outlier robust finite population estimates. *Journ. Amer. Stat. Ass.* 81. pp. 1063-1069. 1986.
- CAPMAN, D.W.: A survey of non-response imputation procedures. Proceedings, Social Statistics Section, American Statistical Association, pp. 245-251. 1976.
- COCHRAN, W.G.: Sampling techniques. John Wiley and Sons New York 1977.
- DALENIUS, T.: Sampling in Sweden. Stockholm: Almqvist and Wiksell. 1957.
- Ends and means of total survey design. University of Stockholm. 1974.
- DALENIUS, T.: Bibliography on non-sampling errors in surveys. *International Statistical Review* 45: 71-89 (A to G) 187-197 (H to Q), 303-317 (R to Z). 1977.
- FAULKENBERRY, G.D. and R.D. TORTORA: A case study of nonsampling error. National Agricultural Statistics Service Report, U.S. Department of Agriculture, Washington, D.C. 1983.
- DALENIUS, T.: Elements of survey sampling. SAREC, Stockholm 1986.
- FELLEGI, I.P.: Response variance and its estimation. *Journal of the American Statistical Association*, Vol. 59: pp. 1016-1014. 1964.
- The evaluation of the accuracy of survey results: some Canadian experiences. *International Statistical Review*, Vol. 41: pp. 1-14. 1973.
- An improved method of estimating the correlated response variance. *Journal of the American Statistical Association*, Vol. 69: pp. 496-501. 1974.
- FELLEGI, I.P., and D. HOLT: A systematic approach to automatic edit and imputation. *Journ. Amer. Stat. Ass.* 71. pp. 17-35. 1976.
- FELLEGI, I.P., and A.B. SUNTER: Balance between different sources of survey errors - some Canadian experiences. *Sankhya: the Indian Journal of Statistics*, 36. (Series C): pp. 119-142. 1974.
- GONZALES, M.E., J.L. OGUS, G. SHARPIRO, and R.J. TEPPIG: Standards for discussion and presentation of errors in survey and census data. *Journ Amer. Stat. Ass.* 70. 351, Part II. 1975.
- HANSEN, M.H. and W.N. HURWITZ: The problem of non-response in sample surveys. *Journal of the American Statistical Association*, Vol. 41: pp. 517-529. 1946.
- HANSEN, M.H., W.N. HURWITZ and M.A. BERSHAD: Measurement errors in censuses and surveys. *Bulletin of the International Statistical Institute*, 38(2): pp. 359-374. 1961.
- HANSEN, M.H., W.N. HURWITZ, and T.B. JABINE: The use of imperfect lists for probability sampling at the Bureau of the Census. 34th session of the International Statistical Institute, Ottawa. 1963.
- JABINE, T., M. STRAF, J. TANUR, and R. TOURANGEAU: Cognitive aspects of survey methodology: Building a bridge between disciplines. National Academy Press, Washington. 1984.

- KALSBECK, W. and J. LESSLER: Total survey design: effect of non-response bias and procedures for controlling measurement errors. In *Health Survey Research Methods*. Second Biennial Conference. Department of Health, Education and Welfare Publication No. (PHS) 79-3207. 1978.
- KALTON, F.: Compensating for missing survey data. Research Report Series; Survey Research Center Institute for Social Research, The University of Michigan. 1983.
- KEMSLEY, W.F.F.: Interviewer variability in expenditure surveys. *Journal of the Royal Statistical Society*, Vol. 128, Series A, pp. 118-139. 1965.
- KIREGYERA, B.: Sources of non-sampling errors in African censuses and surveys. *The Statistician*. 31. No. 2. 1982.
- KISH, L. Studies of the Interviewer variance for attitudinal characteristics. *Journal of the American Statistical Association*, Vol. 57: pp. 92-115. 1962.
- Survey Sampling, New York: Wiley. 1965.
- Statistical Design for Research, Wiley, 1987.
- KISH, L., R.M. GROVES and K. KROTKI: Sampling errors for fertility surveys. WFS Occasional Paper No. 17. The Hague: International Statistical Institute. 1976.
- KISH, L. and I. HESS: On non-coverage of sample dwellings. *Journal of the American Statistical Association*, Vol. 54: pp. 509-524. 1958.
- A «replacement» procedure for reducing the bias of nonresponse. *American Statistician*, Vol. 13: No. 4: pp. 17-19. 1959.
- KORDOS, L.: Methods of controlling the quality in household budget surveys in Poland. Seminar on Statistical Methodology 1988.
- LYBERG, L. and E. RAPAPORT: The advance of statistical theory and methodology and their uses in official statistics. Invited paper 46th Session of the ISI, 8-16 September Tokyo, 1987.
- LINDSTRÖM, H.: Non-response errors in sample surveys. *Urväl* No. 16. Sweden 1983.
- MARQUIS, K.H., M.S. MARQUIS, and J.M. POLICH: Response bias and reliability in sensitive topic surveys. *ASA* 1986. Vol. 81. No. 394. pp. 381-389.
- O'MUIRCHEARTAIGH, C.A. and A.M. MARCKWARDT: An assessment of the reliability of the WFS data. World Fertility Survey Conference, London. 1980.
- O'MUIRCHEARTAIGH, C.A., A.M. MARCKWARDT, and V. VERMA: Response Reliability of WFS Data, *Bulletin of the International Statistical Institute*, 47. No. 3. 1979.
- O'MUIRCHEARTAIGH, C.A.: The magnitude and pattern of response variance in the Lesotho fertility survey. WFS Scientific Reports No. 70. 1984.
- MUNCK, I. and B. WÄRNERYD: Two approaches to reliability measurement in surveys: Reinterviews and confirmatory factor analysis with SIMPLIS/LISREL. Seminar on Statistical methodology U.N. CES Geneva, 1-4 February 1988.
- NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME: Non-sampling errors in household surveys: sources, assessment and control. United Nations Department of Technical Co-operation for Development and Statistical Office. New York 1982.
- NATIONAL HOUSEHOLD SURVEY CAPABILITY PROGRAMME: Sampling frames and sample designs for integrated household survey programmes. United Nations Department of Technical Co-operation for Development and Statistical Office. New York, 1986.
- NETER, J. and J. WAKSBERG: Response errors in collection of expenditures data by household interviewers: an experimental study. Bureau of the Census Technical Paper. No. 11. 1965.
- NORWOOD, J.L.: What is quality? Research Conference of the U.S. Bureau of the Census Baltimore, Maryland March 30, 1987.
- PAYNE, S.: *The Art of Asking Questions*. Princeton University Press. 1951.
- PLATEK, R.: Some factors affecting non-response. *Survey Methodology*, Vol. 3. 1977.
- Questionnaire design: Report on the 1987 BLS Advisory Conference. U.S. Department of Labor Bureau of Labor Statistics, August 1987.
- RUST, K.: Practical problems in sampling error estimation. Invited paper 46th Session of the ISI, 8-16. September Tokyo, 1987.
- SÄRNDAL, C.E.: How survey methodologists communicate. *Journal of Official Statistics*, Vol. 1. No. 1. 1985. pp. 49-63.
- STEVENS, J.A. and B.A. BAILAR: The relationship between various interviewer characteristics and the collection of income data. *Proceedings, Social Statistics Section, American Statistical Association*, pp. 785-790. 1976.
- TUCKER, C.: Identifying measurement error in consumer expenditure reports. *ASA*, 1985. pp. 476-481.
- Toward the development of an ordered set of data quality categories. Meeting of the ASA in San Francisco 1987.
- TURNER, G., F.H. WOLTMAN, R. FAY and B. CARLSON: Sample survey design in developing countries-three illustrations of methodology. *Bulletin of the International Statistical Institute*, 47. 1979.
- VERMA, V.: Sample designs for the World Fertility Survey. *Bulletin of the International Statistical Institute*, Vol. 46. No. 3. 1977.
- Training, the key to quality data. *POPULI*, Vol. 6, No. 1. 1979.
- Sampling for National Fertility Surveys. World Fertility Survey Conference, London. 1980.
- Assessment of errors in household surveys. *Bulletin of the International Statistical Institute*, Vol. 49. 1981.
- VERMA V., C. SCOTT, and C. O'MUIRCHEARTAIGH: Sample design and sampling errors for the World Fertility Survey. *Journal of the Royal Statistical Society, Series A*. Vol. 143, Pt. 4. pp. 431-473. 1980.
- WOODRUFF, R.: The use of rotating samples in the Census Bureau's monthly surveys. *Journ. Amer. Stat. Ass.* 58. pp. 454-467. 1963.

APPENDIX

UNIFIED SYSTEM OF HOUSEHOLD SURVEYS IN HUNGARY

1. Introduction

The unified System of Household Surveys (USHS) was established in 1976 to meet this increasing demand with a reconciled programme covering a time horizon of ten years. It is a nationwide institution. Its sample covers the whole country, the different types of settlements, and in certain cases it allows for computing estimates even for some smaller regions.

The household budget survey is the oldest and best known household survey. At present it is carried out biannually (i.e. in odd years) on the subsample H, covering 12.000 families. The socio-statistical surveys are carried out in the even years on subsample M, covering about 16.000 households. In this frame the survey of Basic Socio-demographic Data (BSD) provides the fundamental demographic and socioeconomic information called cora data.

2. The Sample

The sample or master sample is selected in two or three stages: the primary sampling units (PSU's) are enumeration districts of the 1980 census in larger settlements (typically in cities) and the settlements themselves in other cases. The ultimate sampling units are dwellings in all instances. The requirements imposed on the sample are the following:

- the samples selected from the sampling frame have to represent the major population groups on the national level and—in some cases— on the level of larger regional units or of countries, as well;
- the master sample as sampling frame has to

ensure a sufficient number of dwellings (addresses) for household surveys in the decade 1983-1992, taking the necessary rotations (i.e. replacements of dwellings) into account, which is inevitable to avoid exposing families to response burden for an excessively long period;

- economizing the available resources should be an important aspect of determining the sampling design (this leads inevitably to built-in disproportions together with the requirement on representativity).

The selection of enumeration districts constituting the sampling frame was carried out in one or two stages and using two-way stratification in the following manner:

- all settlements with 25 thousand or more inhabitants were included, i.e. were self-representing. Smaller settlements were stratified by size (8 strata) and socio economic character (5 strata) and were selected with probability proportional to size (measured by the number of dwellings in the 1980 census);
- the enumeration districts were also stratified by character in a similar manner as the settlements. Within the selected non self-representing settlements an equal number of census districts were selected also with probability proportional to size;
- in order to have more settlements in the master sample and to be able to represent certain smaller population groups better by the samples, the sampling fraction in Budapest and in the big cities was lower than in other settlements and in one of the strata by

character of settlements and districts, respectively, was higher than in the rest of strata;

- in any settlement having at least two enumeration districts both subsamples H and M are represented by the same number of districts none of which may pertain simultaneously to H and M;
- for any actual survey an equal number of dwellings is selected from all sample districts with equal probability: thus e.g. 12 dwellings from the districts of subsample M for the BSD survey and 9 from the H subsample districts for the household budget survey.

Since the selection of enumeration districts is carried out with unequal probabilities, sample estimates are determined by inflation, i.e. by means of a suitable weighting procedure. There are 270 different strata in the sample, and, consequently, 270 different sampling fractions and 270 different weighting factors (sample weights). The weighting factor for a given stratum is the ratio of the updated total number of inhabited dwellings and the number of dwellings in the stratum observed in a survey.

In 1983 the number of dwellings observed in the subsample M was 15756 of which 2256 (14 %) were in Budapest. In 1985, the household budget survey (subsample H) covered 11844 households, of which 1692 represented the capital. Besides Budapest, there are 72 towns and 316 villages in the sample (this concerns H as well as M) which are represented by a total of 2630 enumeration districts.

Between 1983 and 1987 three household budget surveys were carried out. The rotation mentioned above means that subsample H operated as a rolling sample; two-thirds of the respondents participated in the survey in two consecutive periods, while one-third of them provided data over three consecutive survey-years (1).

Because of technical reasons there was no rotation in subsample M in 1984. For 1986 a completely new sample was selected from the dwellings of the enumeration districts of M. In 1983-84 there were nearly 8000 families who participated in all five surveys. The possibility of analysing the different surveys in a complex way — i.e. by the use of matching techniques — has not been completely exploited yet.

(1) This can be regarded as a special three-wave panel.

3. Organization

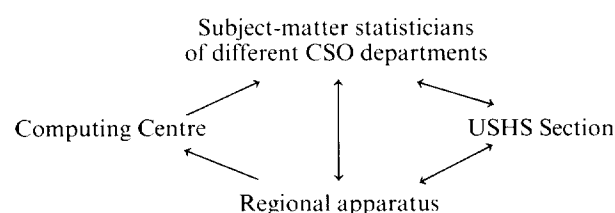
The USHS, as a tool for data collection, is available for the CSO and its institutions. It is managed by a staff of different levels of the organization:

- subject-matter statisticians of the different departments are responsible for the specific problems;
- the central USHS section is responsible for general methodological development, co-ordination, and reliability issues connected with the surveys (e.g. sampling and response errors);
- organization of field work, management and supervision of interviewers is the task of the regional offices of the CSO;
- data processing is carried out in the Computer Centre of the CSO;
- analysis and publication of the results are performed by the subject-matter departments responsible for the survey.

In 1985 the regional offices of the CSO employed about 430 interviewers. It is due to the new sample that 86 % of the interviewers are primarily employed. 72 % of them work part time due to the nature of the regional spreading of the sample.

81 % of the interviewers work only at one settlement, 15 percent at two settlements and 4 percent at three settlements. The majority of the interviewers are responsible for 6-8 enumeration districts. The USHS sections at the regional offices employ altogether 110 persons, the majority having secondary school or higher qualifications.

The diagram illustrates the USHS organization in a somewhat simplified way:



4. Survey Programme: see Table 1

5. Non-Response

The issue of non-response is treated in different manner in the case of subsamples H and M. The

general policy in household budget by one who is willing to answer. On the other hand, in the case of subsample M the problem of non-response is solved by inflating the sample units, i.e. by modifying the sample weights in a suitable manner.

The respondents in the household budget survey have to keep a diary on their daily incomes and expenditures for two months and this special burden explains the fact that the rate of refusals is considerably higher than in other surveys (see Table 2).

Table 1
The Program of the USHS

1983	Sample
Income survey	J
Household budget survey	H ^a
Health expenditures	H ^a
Labour force survey (BSD)	M
Social mobility survey (aged over 14)	J
Prestige survey (aged 14-70)	1/2 J
1984	
Opinions on living conditions	H ^a
Opinions on health services	M
Labour force survey (BSD)	M
Living conditions of young people (aged 15-34)	1 % enlarged 2 %
As part of the microcensus:	
a) Real estate ownership, plans for moving house, flat conversion	
b) Living conditions of pensioners and persons of pensionable age	
c) Longitudinal mortality survey	
1985	
Household budget survey	H
Cultural and recreational expenditures	H
1986	
Basic socio-demographic data (BSD)	M
Living conditions of women (aged over 15)	M
Health status survey	M
Expenditures yielding non-observable incomes	M
Time budget-way of life (aged 15-79)	1/3 M
Teachers' time budget-way of life	1300 teachers
Students' time budget-way of life	3000 students
1987	
Household budget survey	H
Agricultural production of households	H
1988	
BSD	M
Income survey	J
Prestige survey (aged over 14)	M
Medical facilities and morbidity	M

1989	Sample
Household budget survey	H
Health and cultural expenditures	H
1990	
BSD	M
1991	
Household budget survey	H
Recreational expenditures	H
Household facilities	H
1992	
BSD	M

Notations: H^a 8000 households, 6 families from each H district.
H 12,000 households, 9 families from each H district.
M 16,000 dwellings, 12 from each M district
J 16,000 households; half of the sample M and 8,000 families selected from the H districts.

In the past five years the rate of non-response in subsample M was not considerable. Nor did the number of refusals increase perceptibly in 1986, though the burden of the respondents was far from being negligible. In that year all surveys were conducted on subsample M, and this implied that, due to the size of the family and/or to randomization, some families were visited 12-16 times during the survey period.

The rate of non-response is varying from Budapest, to the other cities and villages. The underlying reason for this is probably the different way of life and personal relations in large cities and in smaller settlements, which makes the interviewers' task somewhat difficult.

The number of refusals depends on the subject of the survey as well. People generally like to talk about their problems concerning medical facilities or issues which are personally important to them. The rate of refusals was higher in the case of the prestige survey than usual, since the subject was not clear to most people, and they did not see the aim of the survey.

6. Sampling errors

The questionnaires are in general comprehensive, they contain several hundred variables (questions, data locations, etc.); thus e.g. the household budget survey in 1985 contained 298 questions, and the questionnaire of the survey on «Conditions of

Table 2
Non-Response Rate in the USHS Surveys

	Actual sample size	Rate of refusals (%)			Other non-response (%)		
		Budapest	Provinces	Total	Budapest	Provinces	Total
<i>1983</i>							
Household budget	7,986 ^a	50.4	14.7	19.8	44.8	12.4	17.0
Labour force	15,756 ^b	4.3	1.0	1.5	7.9	7.7	7.7
Income	15,780 ^b	5.8	1.6	2.2	7.9	7.1	7.2
Social mobility	34,571 ^c	6.3	2.3	2.8	5.8	3.6	3.9
Prestige	8,413 ^c	11.4	10.4	10.5	7.0	4.9	5.1
<i>1984</i>							
Labour force	15,756 ^b	7.9	1.8	2.7	—	—	—
Living conditions of the young	26,682 ^c	5.5	1.7	2.1	10.8	6.9	7.4
Opinions on health serv.	20,950 ^c	1.5	0.4	0.5	5.3	2.9	3.2
<i>1985</i>							
Household budget	11,844 ^a	39.4	13.3	17.0	43.1	14.3	18.4
<i>1986</i>							
BSD January	15,756 ^b	5.0	1.7	2.2	4.5	1.2	1.7
Time budget 1st period	11,241 ^c	3.2	1.6	1.8	6.1	3.9	4.2
Time budget 2nd period	10,696 ^c	1.9	1.4	1.4	8.6	2.2	3.0
Time budget 3d period	10,366 ^c	1.6	2.4	2.3	6.0	3.6	3.9
Time budget 4th period	9,626 ^c	1.0	1.8	1.6	5.0	2.4	2.8
Conditions of women	16,085 ^c	4.6	2.1	2.4	1.9	1.1	1.2
Health status	39,750 ^c	1.7	2.2	2.2	4.5	3.4	3.6
Unregistered income	13,728 ^a	—	0.5	0.4	—	1.2	1.1

a) households, b) dwellings, c) persons.

women aged 15 or more and not attending regular school or university» (1986) contained about 250.

All surveys contain the core data which include the most important socio-demographic characteristics (sex, age, marital status, educational attainment, occupation, employment, dwelling conditions). This enables the statistician to analyze the results of the surveys in different contexts by means of suitable cross-tables.

In the emphasis on detailed information increases, reliability of the data obtained from surveys will, of course, decrease. In such situations the question of qualitative comparison becomes relevant: people would like to know if some indicator is higher in some domain (geographic region) than in another, etc.

To compute sampling error of data estimated from the USHS sample such a device is needed that takes the design of this sample — which is rather complex — into account. Of the software facilities available at the Computing Centre of the CSO only

the package program CLUSTERS developed for the purpose of the World Fertility Survey is compatible with the USHS sample design. Unfortunately, CLUSTERS proved to be somewhat slow when applied to USHS problems, moreover, in certain cases its use was definitely difficult. Considering these problems, a new program for computing USHS sampling errors was developed at the Hungarian CSO. This uses the same method as CLUSTERS, but owing to some technical improvements, works considerably faster and is much easier to use than its predecessor (1). It should be noted that the new program determines less derived statistics than CLUSTERS, e.g. it does not compute the DEFT and ROH values for the variables in consideration.

(1) See «New CLUSTERS»: a simple means to compute USHS sampling errors (in Hungarian). Egységes Lakossági Adatfelvételi Rendszer Módszertani Tanulmányok, No. 3. (Manuscript) Budapest, 1986.

Development of statistical software in the last 2-3 years resulted in an increased number of sampling error computations in connection with USHS surveys. Earlier sampling error computations were rather rare, and were always characterized by drastic simplifications that were tantamount to ignoring the actual sampling design. As such computations became frequent, the users had to face the problem of disseminating the results in a universally acceptable and useful form.

It is obvious that some kind of *portability* may be useful here, and, as a matter of fact, it has been in use in some sense. If e.g. sampling errors were estimated for some variables of household budget survey in 1980, the computations were not repeated in 1982, since it was logical to assume that there was no essential change in reliability of the estimates in a period of 2-4 years, considering that no structural change took place and the same variables were estimated on samples of (practically) the same size.

Sample design has a considerable impact on the value or sampling errors. Different features of some sampling plan influence sampling error in different ways, e.g. stratification tends in general to reduce sampling error while clustering has just the opposite effect. The statistics *deft* and *roh* derived from sampling error are widely known and may be useful to measure the performance or efficiency of some sample plan (1).

In the course of investigating the results of some fertility surveys prior to the World Fertility Survey sampling error of the estimates was thoroughly examined in many relations and an attempt was made to find some portability by identifying the components of sampling error (2). The problem was further investigated in the paper by Verma, Scott and O'Muircheartaigh (3) where the method of investigating *deft* and *roh* was also improved.

Sampling error computations could be made much easier if there were some reliable methods by which confidence limits of the variables could be

extrapolated provided that sampling error is determined *for a suitably chosen subset of the variables*. The same applies to the considerations concerning the plan of future surveys repeated at regular intervals.

In what follows sampling error computations performed in connection with USHS surveys will be reviewed, and the existence of certain relations among derived statistics will (*ROH* and *DEFT*) be investigated and checked.

Sampling error computations were performed among others in connection with the USHS surveys.

- Living conditions of youth and
- Opinions on health conditions

(both conducted in 1984); the objective here was to find a proper strategy for small area estimation methods. In addition, simulation experiments based on sampling error computations were carried out to examine the performance of USHS samples as a function of sample size.

In many cases the values of *ROH* and *DEFT* were difficult to interpret, as

- *ROH* became negative in a lot of cases,
- and in the majority of cases the average PSU size was less than 6, when numerical instability hampers any meaningful conclusion.

Considering these difficulties we had to define a subset of our sampling error computations such that *ROH* is nonnegative for all variables and for all subclasses in consideration. It was necessary to have at least two subclasses in any case, since the *ROH*'s were to be averaged over subclasses. It was not easy to find a proper subset with this property; finally it included portions of computations related to the surveys «Living conditions of youth» and «Opinions on health conditions» as well as to the simulation experiment mentioned above.

The results of our computations are presented in Table 3-5. Data in these tables pertain to provinces total (i.e. country total minus the capital) and to three counties (Baranya, Szabolcs and Vas) in Table 5 provinces total is replaced by a region formed from five counties. The tables were compiled by the following principles. Column 1 contains *ROH* values belonging to some geographical domain (e.g. a country) as a whole, while column 2 displays *EOH*'s averaged over subclasses. Column 3 contains the ratios formed from the corresponding entries of col's two and one; by Kish, Groves and Krotki this ratio ought to vary around 1.2. Columns.

(1) See Kish, L.: Survey Sampling, New York and London, John Wiley and Sons, 1965.

(2) L. Kish, R.M. Groves and K.P. Krotki: Sampling Errors for Fertility Surveys. WFS Occasional Papers, No. 17, January 1976.

(3) V. Verma, C. Scott and C. O'Muircheartaigh: Sample Designs and Sampling Errors for the World Fertility Survey. J.R. Statist. Soc. A (1980), 143, Part. 4, pp. 431-473.

Table 3
Some ROH Values from the Error Computations for the Survey «Living Conditions of Youth» (1984)

Variables	Provinces total			Baranya			Szabolcs			Vas		
	1. ROH	2. Subclass means	3. (2)/(1)	4. ROH	5. Subclass means	6. (5)/(4)	7. ROH	8. Subclass means	9. (8)/(7)	10. ROH	11. Subclass means	12. (11)/(12)
1.	0,130	0,189	1,46	0,064	0,088	1,37	0,106	0,083	0,79	0,107	0,086	1,91
2.	0,151	0,229	1,51	0,129	0,224	1,73	0,085	0,102	1,2	0,070	0,068	1,98
3.	0,105	0,132	1,26	0,125	0,096	0,77	0,033	0,092	2,43	0,009	0,006	0,61
4.	0,210	0,398	1,90	0,071	0,128	1,80	0,272	0,541	1,99	0,101	0,055	0,55
5.	0,158	0,210	1,33	0,201	0,320	1,59	0,167	0,183	1,10	0,342	0,374	1,09
6.	0,181	0,227	1,25	0,184	0,257	1,39	0,123	0,191	1,55	0,128	0,104	0,81
7.	0,180	0,218	1,21	0,327	0,307	0,34	0,189	0,230	1,22	0,108	0,139	1,39
8.	0,210	0,247	1,18	0,085	0,070	0,82	0,177	0,157	0,89	0,086	0,068	0,79
Mean of rows 1-8	0,166	0,231	1,39	0,148	0,186	1,26	0,145	1,98	1,37	0,119	0,112	0,95

Table 4
Some ROH Values from the Error Computations for the Survey «Opinions on Health Conditions» (1984)

Variables	Provinces total			Baranya			Szabolcs			Vas		
	1. ROH	2. Subclass means	3. (2)/(1)	4. ROH	5. Subclass means	6. (5)/(4)	7. ROH	8. Subclass means	9. (8)/(7)	10. ROH	11. Subclass means	12. (11)/(12)
1.	0,077	0,087	1,12	0,011	0,051	4,64	0,066	0,079	1,19	0,089	0,090	1,01
2.	0,099	0,104	1,05	0,027	0,024	0,89	0,385	0,395	1,02	0,060	0,052	0,86
3.	0,023	0,043	1,86	0,039	0,033	0,85	0,011	0,085	7,72	0,042	1,036	0,85
4.	0,056	0,077	1,37	—	—	—	0,066	0,076	1,15	0,064	0,056	0,87
5.	0,041	0,047	1,14	0,039	0,051	1,31	0,030	0,013	0,43	0,039	0,067	1,71
Mean of rows 1-5	0,059	0,071	1,20	0,029	0,040	1,38	0,111	0,129	1,16	0,058	0,060	1,03

Table 5
Some ROH Values from the Error Computations of a Simulation Experiment based on the Microcensus (1984)

Variables	Provinces total			Baranya			Szabolcs			Vas		
	1. ROH	2. Subclass means	3. (2)/(1)	4. ROH	5. Subclass means	6. (5)/(4)	7. ROH	8. Subclass means	9. (8)/(7)	10. ROH	11. Subclass means	12. (11)/(12)
1.	0,028	0,046	1,63	—	—	—	0,032	0,097	3,03	0,039	0,164	4,21
2.	0,037	0,062	1,66	0,028	0,036	1,27	0,058	0,076	1,30	0,046	0,070	1,52
3.	0,129	0,138	1,07	0,008	0,057	7,13	0,356	0,367	1,03	0,011	0,022	2,00
4.	0,086	0,033	0,96	0,034	0,021	0,59	0,121	0,116	0,95	0,087	0,077	1,89
Mean of rows 1-4	0,070	0,082	1,20	0,011	0,038	3,42	0,142	0,164	1,16	0,046	0,083	1,80

4, 5 and 6 play the same role as columns 1, 2 and 3, respectively, and the same holds for columns 7, 8 and 9, etc. In each table the last row contains the simple arithmetic mean of the corresponding entries in the other rows; however, the third, sixth, ninth and twelfth entries of the last row make exceptions, being the ratios of the two preceding values. In connection with Tables 1-3 it is useful to note the following.

<i>Table 3:</i>	Sample size:
Provinces total	23372
Baranya	1225
Szabolcs	1864
Vas	804
Subclasses: physical workers not in agriculture, physical workers (in all active earners).	

<i>Table 4:</i>	Sample size:
Provinces total	17643
Baranya	942
Szabolcs	1282
Vas	587
Subclasses: males, females	

<i>Table 5:</i>	Sample size:
Five counties (as a region)	5086
Baranya	965
Szabolcs	1177
Vas	625
Subclasses: males, females	

In evaluating the results two questions will be dealt with:

- do our computations confirm the conjecture that the ratio of averaged ROH's over subclasses to the ROH of the sample as a whole is about 1.2?

and

- if the answer to the first question is «yes», does the computation scheme for portability work satisfactorily?

Considering Tables 1-3 one may conclude that the answer is «yes» to the first question. Some remarks are in order here. First, the ratio of averaged ROH to ROH will get closer to 1.2 in the case of large samples than in the case of small ones; the lambda-values seem more stable in the last row of the tables than in other rows. Similarly, lambda-values pertaining to larger geographic units (provinces total or a region of five counties) are better than lambda-values of counties. In this sense $\lambda = 1.2$ for Tables 4-5 and $\lambda = 1.39$ for Table 3. Second, data of Table 2 are available for all of the 19 counties; the corresponding lambda-values—in the usual order of counties—are as follows:

1,38, 0,98, 0,97, 0,97, 1,06, 1,18, 2,10, 0,96, 3,12, 1,17, 0,95, 1,29, 2,73, 1,16, 2,74, 1,32, 1,03, 1,39, 2,27.

To check the possibility of portability by means of our results, the corresponding procedure should be outlined here. The problem is

to find an estimate of the standard error (SE_r) of some variable for a given subclass, and assume that we are given

- \bar{B} , the average of the size of PSU's,
- ROH for the sample as a whole,
- M_r , the proportion of the given subclass to the population as a whole,
- SR_r , an estimate of the standard error that would have been obtained if we had an equivalent simple random sample.

The computational scheme is as follows:

$$DEFT_r = \sqrt{1 + 1.2 \text{ ROH } (M_r \bar{B} - 1)}$$

and

$$SR_r = DEFT_r \times SR_r$$

The performance of this method of imputation can be judged by comparing the SE_r values obtained by the above formulas with the standard errors computed by CLUSTERS. Data for this comparison can be found in Table 4 where the values of \bar{B} , ROH, M_r and SR_r as well as SE_r in the last column were computed by CLUSTERS. The values of SE_r in the last but one column are those obtained by the imputation procedure.

Table 6 contains 21 comparisons, and on the basis of the differences between the entries of the last two columns one may conclude that the imputation procedure suggested by L. Kish performed surprisingly well. Of the 21 cases there were only three such that the difference might be significant: 0.027, 0.006 and 0.023 was obtained in place of 0.036, 0.009 and 0.029, respectively, and there was no difference in the order of magnitude at all. Thus in principle the imputation procedure may be justified.

From technical point of view portability as described above may be difficult in some instances. In our calculations the values \bar{B} , ROH, M_r and SR_r were taken from lists produced by the CLUSTERS program. What can be said of a real situation? One may assume that \bar{B} and ROH for the sample as a whole are known, but this may not be true for the weight M_r of the subclass within the population as well as for the standard error SR_r corresponding to the simple random sample of equivalent size. If one would insist on computing these data by CLUSTERS, then SE_r in question were also obtained, without any imputation. This problem actually occurs when one has to estimate means or

Table 6
Comparison of Computed and Imputed Values of SE_r

Variable	\bar{B}	ROH	M_r	SR_r	Imputed value	Computed value
1.	35,2	0,067	0,417	0,004	0,006	0,009
2.	25,3	0,107	0,640	0,012	0,021	0,023
3.	15,2	0,085	0,044	0,008	0,014	0,011
4.	15,5	0,126	0,598	0,011	0,017	0,017
5.	15,5	0,071	0,598	0,016	0,021	0,024
6.	30,3	0,167	0,750	0,010	0,023	0,021
7.	20,8	0,128	0,793	0,021	0,039	0,036
8.	26,7	0,042	0,509	0,011	0,012	0,015
9.	19,2	0,011	0,526	0,021	0,022	0,026
10.	32,0	0,185	0,529	0,012	0,033	0,033
11.	26,7	0,064	0,491	0,012	0,017	0,016
12.	19,2	0,039	0,474	0,012	0,014	0,013
13.	26,7	0,042	0,509	0,011	0,017	0,015
14.	16,9	0,008	0,441	0,014	0,014	0,016
15.	25,0	0,039	0,435	0,022	0,027	0,036
16.	16,9	0,016	0,559	0,021	0,019	0,021
17.	25,0	0,058	0,43	0,022	0,029	0,011
18.	25,0	0,011	0,565	0,012	0,013	0,013
19.	16,9	0,034	0,441	0,024	0,027	0,026
20.	25,0	0,032	0,569	0,019	0,023	0,029
21.	25,0	0,046	0,435	0,030	0,037	0,039

ratios; to compute SR_r for SRS on the basis on 500 observations by means of a desk calculator just for the sake of portability would certainly not be meaningful. The situation is different when the problem is estimating proportions.

Suppose now that one has to estimate some proportion and a fairly good estimate P is available. The standard error SR_r that would have been obtained under simple random sampling can be estimated as $\sqrt{P(1-P)/N}$. It is natural to ask if the conclusions from Table 4 were to be modified in the case where the entries in the column SR_r were replaced by the corresponding values of $\sqrt{P(1-P)/N}$. Carrying out the corresponding calculations we have found that with the precision of three decimals exactly the same values of SR_r are obtained as previously by CLUSTERS. This means that in the case of estimated proportions the imputation scheme can be used also in practice; the knowledge of M_r can in general be assumed as the result of the standard processing.

7. Response errors

In 1983 there were two USHS surveys which contained some information on personal

incomes(1). One was the so called Labour Force Survey, where only a rough estimate of the personal income was asked. (Seven categories, see table 7). In the second one there were detailed questions as far as incomes were concerned. So this second one can be regarded as more precise than the first one. 9406 persons gave answer to both questions. You have to consider, that the size of income is a sensitive question, so there is certain underreporting.

The results are in the table 7.

$$A = \sum p_{ii} = 0,33, \quad D = 1 - A = 0,67$$

$$K = 1 - \frac{1 - \sum p_{ii}}{1 - \sum p_{i.} p_{.i}} = \frac{0,141}{0,811} = 0,173$$

Under reporting:

$$\sum_{i \geq j} p_{ij} = \frac{5287}{9406} = 0,56$$

Over reporting:

$$\sum_{i \leq j} p_{ij} = \frac{1016}{9406} = 0,11$$

(1) Keleti, A., Marton, A.: Comparative Analysis of Income Data Obtained from Sample Surveys of the Population. Statisztikai Szemle, Vol. 64(1986) No. 11. p. 1119-1125. (in Hungarian).

Table 7

The comparison of the 1982 monthly average earnings data of the 1983 labour force survey and the 1983 income survey on the two surveys' joint sample

INCOME SURVEY (information obtained from the employer)	LABOUR FORCE SURVEY (information obtained from the respondent)							
	Monthly average earnings (Forint)							
Monthly average earnings (Forint)	Below 2500	2510- 3500	3501- 4500	4501- 4500	5501- 5500	6501- 6500	Over 7500	Total 7500
	Hungary							
Below 2500	263	141	48	26	3	2	1	484
2501-3500	354	889	271	85	11	2	2	5,1 1614
3501-4500	120	1018	879	263	31	9	1	17,1 2321
4501-5500	25	267	1010	699	74	11	0	25,0 2086
5501-6500	8	51	345	689	184	16	10	22,2 1303
6501-7500	1	18	96	346	214	44	9	13,8 728
Over 7500	3	14	57	223	252	176	145	7,7 870
Total	774 8,2	2398 25,5	2706 28,8	2331 24,8	769 8,2	260 2,8	168 1,8	9406 100,0

The relatively high response error is due to the fact, that the two surveys used different methods (questions) and also the character of the questions. Evaluating the results you have to consider this uncertainty.

A Social Prestige Survey took place in the second half of 1988 previous one in 1983. In order to study the response error of different questions, there were 560 reinterviews 3 weeks after the first interview. The second interviewer was different from the first one. The questionnaire was shorter than the original one, but contained the fundamental questions. The evaluation of the interviews are not yet completed. Some preliminary results:

30 professions (physician, priest, policeman, postman, carpenter, etc.) were tested. Those who answered were asked to range from 1 to 30 the cards with the names of the 30 professions.

The results were quite surprising: practically all professions have to get all the possible scores from 1 to 30. But the average score of the individual professions were very stable.

This means that the individual scores have a fairly great variable error, but the average of this error (560 cases) is nearly zero.

Professions	Scores	
	1st interview	Reinterview
1.	4,13	4,06
2.	5,40	5,44
3.	6,66	6,74
4.	6,84	6,17
5.	7,79	8,15
6.	9,54	9,53
7.	10,77	10,33
8.	12,15	12,09
9.	12,72	12,83
10.	13,55	13,47
11.	14,03	14,00
12.	14,32	13,55
13.	15,18	15,32
14.	15,47	15,77
15.	15,91	15,69
16.	16,09	15,99
17.	16,16	15,91
18.	16,67	16,90
19.	18,12	18,16
20.	18,18	18,17
21.	18,28	18,64
22.	19,18	18,96
23.	19,49	19,68
24.	20,08	20,09
25.	20,76	21,04
26.	20,86	20,77
27.	21,24	21,66
28.	21,76	21,90
29.	25,59	25,39
30.	28,02	28,27

If you take the economic activity of the individual persons, you will find 77 % of answers were exact. If you take one category difference, the agree-

ment 90 %. This means that economic activity is a fairly unstable idea for a lot of people and possibly the interviewer-effect had also some impact.

Economic activity

First interview	Second interview									Total
	0	1	2	3	4	5	7	8	9	
0	38	1	0	1	0	2	0	0	1	43 7.65
1	0	129	3	0	2	2	0	4	1	141 25.09
2	0	14	112	18	0	3	0	4	4	155 27.58
3	0	3	11	34	1	3	0	0	2	54 9.61
4	0	0	0	0	3	0	0	0	0	3 0.53
5	1	0	5	2	0	19	0	0	1	28 4.98
7	0	1	0	0	0	0	6	1	1	9 1.60
8	0	5	2	0	0	0	2	49	13	71 12.63
9	0	1	5	0	0	0	0	11	41	58 10.32
Total	39 6.94	154 27.40	138 24.56	55 9.79	6 1.07	29 5.16	8 1.42	69 12.28	64 11.39	562 100.00

A = 0,767; D = 0,233; k = 0,715.