ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
Hugo Steinhaus Jubilee Volume
X (1969)

T. DALENIUS (Stockholm)

TOWARDS A SURVEY MEASUREMENT SYSTEM

I. Introduction.

1. Survey methodological development — a review. In response to gradually increasing demands for better statistics at lower costs, survey statisticians in many parts of the world have in the last few decades devoted great efforts to the development of survey methodology, and especially to the development of methods for coping with various sources of *errors in surveys*. In this development, we may distinguish two main lines of development.

In the 1930's, the emphasis of these endeavours was on the *sampling errors*. Neyman's classical paper on the representative method, Neyman [18], meant the breakthrough in the realm of sample surveys of the revolution in statistics started by R. A. Fisher. In addition, it served as a powerful trigger to the subsequent evolution.

In the 1940's, this development came into full bloom. New measurable sampling designs evolved, and replaced various non-measurable ones. Among the important contributions of this decade, Hansen and Hurwitz [10] and Mahalanobis [17] deserve special mention.

At the same time, however, a strong interest arose in the problems caused by the *non-sampling errors*. This interest was especially pronounced in the United States and in India. A shift of relative emphasis of research and developmental efforts took gradually place in the 1940's and, especially, in the 1950's. It is reflected in the development and application of a variety of procedures for coping with *specific* sources of non-sampling errors, for example non-response and measurement errors. Mention should also be made of the growing interest in assessing the quality of various census surveys; the possibilities of carrying out various types of "quality checks" were indeed much enhanced by the access to efficient measurable sampling designs. Chevry [2] and Eckler and Pritzker [8] are two illustrations in kind.

Today, great efforts are devoted to the unification of the two lines of development outlined above. Thus far success has been greatest in the realm of descriptive surveys. The approach chosen makes extensive use of a set of "tools of survey quality control". In general outline, it is as follows.

The error of the survey result $\hat{\bar{y}}$ (based on a measurable design) is measured by the mean-square error $E(\hat{\bar{y}}_t-\bar{X})^2$, where $\hat{\bar{y}}_t$ is a stochastic variable defined within the framework of a certain "survey model", and $\bar{X}$ is the defined goal of the survey. The model provides a decomposition of the mean-square error into a set of error components reflecting a corresponding set of sources of (sampling and non-sampling) errors. Adherence to certain statistical standards calls for exercising such a control of these error components that the mean-square error is minimized for a given survey cost.

2. The purpose and organization of the paper. It is the prime purpose of this paper to present the systematics of the endeavours towards a unification of the two lines of the methodological development mentioned in par. 1.

To this end, we discuss in chapter II the notion of a descriptive survey. In chapter III, the main tools of quality control of surveys are exhibited. In chapter IV, an exposé is given of various applications of these tools. In chapter V, finally, we embark upon a discussion of the future of survey theory and methods.

For the details of the methodological development, the reader is referred to the references.

II. The notion of a descriptive survey.

3. The distinction between "experiments" and "surveys". The term "survey" is often used in the statistical literature in a rather loose way. In this paragraph we will therefore specify the meaning in which this term will be used in this paper.

Consider a real-life problem to the solution of which a statistical investigation of some kind is expected to contribute. This contribution may be conceived as the answer to some *explanatory* question such as "What is the effect on the yield of wheat from the use of fertilizers?", or "Does smoking cause lung cancer?". Or the contribution may be conceived as the answer to some *descriptive* question such as "What is the yield of wheat on farms using different amounts of fertilizers?", or "What is the prevalence of lung cancer among smokers and non-smokers, respectively?".

In the former case, an "experiment" would in principle be called for. It is characteristic of an experiment that it provides data for comparing the effects of experimental "treatments"; the statistician determines by design not only which factors to include in the experiment, but also the *levels* for these factors.

In the latter case, however, a "survey" would provide the answer. It is characteristic of a survey that it is based on *observational* data; the statistician collects data as they are generated by Nature.

For a comprehensive review of the distinction between "experiments" and "surveys" as put forward here, reference is given to Wold [23]. Some statisticians prefer to make a different distinction, one between "*absolute* experiments" and "*relative* experiments". For an elaboration of this distinction, reference is given to Anscombe [1](1).

4. The descriptive survey. In what follows, we will consider a survey as defined in par. 3; more specifically, we will consider a *descriptive* survey, as distinguished from an analytical survey(2).

The specifications of the survey as laid down in the survey design identify a well-defined population of N elements. Moreover, the specifications identify the methods by which observations (measurements)(3) are to be collected for (a sample of) these elements; these observations may concern some quantitative characteristic, for example volume of sales in 1968 for a population of retail stores, or be qualitative in nature, as would be the case, if the purpose of the survey is to count the elements of some population. The specifications identify, finally, the methods by which these observations are to be processed (the use of imputations in the case of non-response, editing procedures, tabulations, etc. and summarized, or — for a sample survey — the estimation procedure).

Corresponding to these specifications, there is a survey that would yield the defined goal, say $\overline{X}$, were it carried out in a faultless fashion.

The execution of the specifications yields, however, a statistic $\hat{\bar{y}}$. The deviation $\hat{\bar{y}}-\overline{X}$ reflects the impact of various (sampling and non-sampling) errors.

We conceive of a sequence of executions of the specifications, to be referred to as "trials". Each such trial yields a statistic $\hat{\bar{y}}_t$, $t = 1, 2, \ldots$ We will denote the average of these statistics over a large number of trials by $\underset{t}{E}\,\hat{\bar{y}}_t = \overline{Y}$. The difference $\overline{Y}-\overline{X}$ is defined to be the *bias* of the statistic $\hat{\bar{y}}_t$.

---

(1) For a discussion of the need for a unified theory of "random experiments", reference is given to Dalenius and Matérn [6].

(2) Cf. for example Cochran [3], chapter 1, and Yates [24], chapter 9.

(3) We will make no distinction here between "observations" and "measurements".

III. The main tools of quality control of surveys.

5. The content of chapter III. In this chapter, we will present a broad review of the main tools of quality control of surveys. More specifically, we will discuss the following tools:

(i) A survey model developed by the U. S. Bureau of the Census.
(ii) Certain "basic study schemes", related to this model.
(iii) Methods for statistical quality control.
(iv) Statistical standards.

6. U. S. Bureau of the Census survey model. The theory of sample surveys as presented in a variety of textbooks is to a large extent (but not exclusively) a sizeable collection of sampling models oriented towards the problem of measuring and controlling the sampling error.

We will use the term "survey model" to refer to a more general model, applicable to both sample and census surveys. It is characteristic of a survey model that it is oriented towards the problem of measuring and controlling the total error, that is the combined effect of two main sources of errors, the use of sampling, and the measurement procedure, respectively.

In the last few years, several survey models have been constructed. In this paper, we will limit our attention to a model developed at the U. S. Bureau of the Census. The model is presented in considerable detail in Hansen et al ([11] and [13]). We will therefore be content here with a summary presentation.

We will present the model by reference to (1) the survey situation where the survey is to be taken, (2) the postulates of the model, and (3) the observations.

(1) We consider the following survey situation. There is a population of N elements of some kind. This population is characterized by a certain parameter. An approximation to this parameter is to be provided by means of a sample or census survey.

The elements of the survey (that is, the elements of the sample in the case of a sample survey, and the elements of the population, in the case of a census survey) are observed in the course of the survey.

There are no *coverage* errors: all survey elements are somehow subjected to observation. It is important to realize that "observation" is used here in a broad sense; while most of the elements are being observed in the usual sense of this word, observations for some of the elements may be collected by means of "imputation" techniques.

There are, however, *content* (measurement) errors: some (or all) of the observations may be in error.

(2) The model assumes the following four postulates:

(i) An unequivocal rule exists for identifying the N elements of the population.

(ii) With each one of these elements, we associate an X-value:

$$X_1, \ldots, X_J, \ldots, X_N.$$

The parameter to be approximated is the mean per element:

$$\bar{X} = \frac{1}{N} \sum_J X_J.$$

(iii) A particular survey is taken under a general set G of conditions. To some extent we are able to control these conditions; to the extent we are not able to do so, this is reflected in chance variations in the observations.

(iv) The process of observing an element is repeatable: we may at will carry out a sequence of "trials". Moreover, it generates a random variable, whose value at one trial is uncorrelated with its value at any other trial.

(3) In a sequence of trials, we collect a sequence of observations, to be denoted by:

$$y_{j1G}, y_{j2G}, \ldots, y_{jtG}, \ldots$$

for the jth element, $j = 1, \ldots, n$, in the case of a sample survey, and by:

$$y_{J1G}, y_{J2G}, \ldots, y_{JtG}, \ldots$$

for the Jth element, $J = 1 \ldots N$, in the case of a census survey.

In what follows, we will assume that we are dealing with a sample survey; we will also assume that the sample elements are selected by means of simple random sampling.

We define, for the jth element, the expected value over trials:

$$\underset{t}{E}\, y_{jt} = Y_j$$

where for convenience the index G has been dropped. For the sample of n elements observed at trial no. t, we have

$$\hat{\bar{y}}_t = \frac{1}{n} \sum_j y_{jt}$$

and

$$\hat{\bar{Y}}_t = \frac{1}{n} \sum_j Y_j.$$

As an approximation to $\bar{X}$, the statistic $\hat{\bar{y}}_t$ has an error which may be measured by the mean-square error:

$$MSE(\hat{\bar{y}}_t) = E(\hat{\bar{y}}_t - \bar{X})^2$$

where the expectation is taken over all trials and samples.

If we denote the "response deviation" by:

$$d_{jt} = y_{jt} - Y_j$$

and write (using E in the same sense as above):

$$Ed_{jt}^2 = \sigma_d^2,$$

$$Ed_{jt}d_{kt} = (n-1)\varrho_d\sigma_d^2;$$

the mean-square error may be decomposed as follows:

$$MSE(\hat{\bar{y}}_t) = \frac{1}{n}\sigma_d^2 + \frac{n-1}{n}\varrho_d\sigma_d^2 + E(\hat{\bar{Y}}_t - \bar{Y})^2 + 2E(\hat{\bar{y}}_t - \hat{\bar{Y}}_t)(\hat{\bar{Y}}_t - \hat{\bar{Y}}) + (\bar{Y} - \bar{X})^2.$$

For simplicity, we write

$$MSE(\bar{y}_t) = C_{11} + C_{12} + C_2 + C_3 + C_4,$$

where the components have the following interpretation: C_{11} — the simple response variance, C_{12} — the correlated response variance, C_2 — the sampling variance, C_3 — an interaction term, C_4 — the square of the bias.

For a census survey, C_2 and C_3 are zero, leaving:

$$MSE(\hat{\bar{y}}_t) = C_{11} + C_{12} + C_4$$

as the total error.

7. Basic study schemes. The practical usefulness of the survey model presented in par. 6 will, among other things, depend upon the realism of the model. We will not dwell upon this aspect here; we state, however, that the model has successfully stood the tests of applications.

Granted the satisfactory realism of the model, it remains to develop designs for estimating the error components. We limit our discussion here to a brief review of such designs, to be referred to as "basic study schemes", for estimating (1) the simple response variance, (2) the correlated response variance, and (3) the square of the bias.

(1) For the estimation of the simple response variance, *the replication method* may be used.

As a simple illustration, consider a sample survey comprising n elements selected by simple random sampling. Each element is observed twice; the observations are labelled $y_{jt}, j = 1, \ldots, n, t = 1, 2$. We define the gross difference as:

$$g = \frac{1}{n} \sum_j (y_{j1} - y_{j2})^2.$$

It can be shown that (under certain conditions):

$$Eg = 2\sigma_d^2 - 2Ed_{j1}d_{j2}$$

where σ_d^2 is defined in par. 6. For certain types of surveys, the covariance term equals zero (or may safely be assumed to be small); then $g/2$ may be used to estimate σ_d^2 which appears in the component C_{11} of $MSE(\hat{\bar{y}}_t)$.

(2) For the estimation of the correlated response variance, *the interpenetration method* may be used. This method has a long standing in the realm of efforts to cope with the non-sampling errors; it was used in India by Professor P. C. Mahalanobis in the 1930's.([4])

The interpenetration method calls for observing distinct samples. In the case of two samples, the observations are $y_{j1}, j = 1, \ldots, n$, and $y_{k2}, k = 1, \ldots, n$. We define:

$$C = \frac{1}{2}\left[\frac{1}{n}\sum_j y_{j1} - \frac{1}{n}\sum_k y_{k2}\right]^2$$

and

$$D = \frac{\sum_j (y_{j1} - \bar{y}_1)^2 + \sum_k (y_{k2} - \bar{y}_2)^2}{2n(n-1)},$$

where $\bar{y}_1$ and $\bar{y}_2$ are the means of the two samples, respectively. Now it can be shown that (subject to certain assumptions):

$$E(C-D) \doteq \varrho_d \sigma_d^2$$

that is, $C-D$ may be used in the estimation of the component C_{12} of $MSE(\hat{\bar{y}}_t)$.

(3) The estimation of the bias term in the expression for the mean-square error offers considerable conceptual and methodological difficulties. One approach calls for the following use of the replication method: in trial no. 1, observations are made by the "regular" measurement procedure; in trial No. 2, a "superior" measurement procedure is used. The difference:

$$b = \bar{y}_1 - \bar{y}_2$$

may then be taken as kind of an estimate of the bias term.

([4]) This also holds true for the replication method.

8. Statistical quality control. Statistical quality control originated in the 1930's in industrial mass production. Today, it is widely used in that area: in addition it has found important applications in other areas, and notably in the area of large scale statistics production.

We will point here to three aspects (1)-(3) of special interest in the present context.

(1) In most applications, the purpose of statistical quality control is one or both of the following ones:

(i) To guide action on the output of a certain operation and reveal errors made in a previous operation — "product control".

(ii) To guide action on the operation itself — "process control".

(2) We may usually identify the following three stages of quality control operations:

(i) The specification of what is wanted, in terms of "quality".

(ii) The production on the basis of this specification.

(iii) The inspection to find, if the production takes place in accordance with the specification, or not.

(3) Finally, statistical quality control is, to a large extent, based on sampling. Most schemes in actual use belong to one of the following three categories:

(i) Acceptance sampling.

(ii) Continuous sampling.

(iii) Shewhart control.

For a fuller discussion of statistical quality control, we refer the reader to the special literature of the subject.

9. Statistical standards. "Standards" is used here as a concise term denoting all kinds of rules, principles, etc. which are applied to achieve a certain quality.

In what follows, we will briefly present two categories of standards: (1) fundamental standards, and (2) standards for the design and the operations. The discussion will be implicitly and explicitly tied to the previously given expression:

$$MSE(\hat{\bar{y}}_t) = C_{11}+C_{12}+C_2+C_3+C_4.$$

(1) We will point to two fundamental standards.

First, the survey should be based on a *measurable* design; this is implied by the use of the $MSE(\hat{\bar{y}}_t)$ as the measure of the total error of the survey.

Second, the survey should exploit *reproducible* methods. This standard implies that the same results are arrived at (within the range of sampling and response variances), were the survey repeated on the basis of the same design, by the same or some other statistician.

(2) For the control of the components of the $MSE(\hat{\bar{y}}_t)$, standards for the design and standards for the operations are used.

The planning of a survey calls, to a large extent, for excercising a choice between alternative methods. A reasonable standard for this choice is to choose that method which contributes the least to $MSE(\hat{\bar{y}}_t)$ for a given expense.

For the control of the performance of a chosen method, we need in addition performance standards, in principle one such standard for each separate operation.

For a comprehensive discussion of standards used for purposes of quality control of surveys, reference is given to Hansen et al [14].

IV. Applications.

10. The problem of quality control of surveys. In par. 4, we introduced the notion of a descriptive survey. Corresponding to a set of specifications, there is a defined goal $\bar{X}$. Actual performance yields, however, the statistic $\hat{\bar{y}}$ with the mean-square error $MSE(\hat{\bar{y}}_t)$.

Corresponding to an alternative set of specifications, there would be an alternative defined goal $\bar{X}'$. Actual performance would in this case yield the statistic $\hat{\bar{u}}_t$ with the mean-square error $MSE(\hat{\bar{u}}_t)$.

This points to two alternative ways of controlling the level of the mean-square error of a statistic:

(i) By choice of the specifications.

(ii) By quality control of the survey operations.

These ways will be illuminated in par. 11-16.

11. Choosing the specifications. We will give here three examples (1)-(3) of this approach.

(1) For the collection of the observations of a population census, one may make a choice between two alternative procedures: to use a *direct* enumeration, for example by means of special census enumerators, or to use an *indirect* enumeration, for example by eliciting the observations from existing administrative records. Moreover, as illustrated by the 1950 and 1960 censuses of population and housing in the United States, there is still — within the realm of the direct procedure — considerable opportunities for choice.

(2) In order to collect data on household expenditures, a choice may be made between the record-keeping approach and the personal interview approach.

(3) In household surveys, where the observations are collected by interviews, a choice may be made between interviewing a designated member of the household, for example the head, and interviewing any responsible member.

12. Quality control of the survey operations. We will use the term "quality control" to denote any action having as its purpose to excercise an influence on the level of the mean-square error. The question is now how such an influence may be brought about.

In par. 6, we showed that:

$$MSE(\hat{y}_t) = C_{11}+C_{12}+C_2+C_3+C_4.$$

The methods for coping with the sampling errors, discussed in par. 1, are directly applicable to the control of C_2 and C_3. In what follows, we will limit the discussion to the components of the mean-square error of a *census* survey, that is to:

$$MSE(\hat{y}_t) = C_{11}+C_{12}+C_4.$$

Thus, the specifications of the survey, and therefore the defined goal $\bar{X}$, are taken as a datum.

For the purpose of the following presentation, it is instructive to identify three main survey operations(5):

(i) The pre-field operations.

(ii) The field operations.

(iii) The post-field operations.

13. Quality control of the pre-field operations. The following is an example of the application of quality control of a pre-field operation.

In the 1960 censuses of population and housing in the United States, extensive use was made of FOSDIC(6). This medium for input to the computers imposed specific requirements on the quality of the printing of schedules. To achieve this quality, a special quality control program was implemented as discussed in Hansen et al [9]. This reference should also be consulted for additional examples.

14. Quality control of the field operations. When we are confronted with evidence of unsatisfactory field operations, it is natural to apply such general devices as intensified training and supervision of the field--workers. As emphasized in Hansen and Steinberg [15], comparatively few attempts have been made to evaluate this kind of quality control.

Irrespective of the amount of training and supervision used, we will be faced with the necessity of controlling the quality of the field operations. We will distinguish between control of coverage errors and content errors, respectively.

(1) There is today a sizeable body of devices for coping with *coverage errors*. We will give two examples here; for simplicity, they refer to the case of a census survey.

(5) This terminology reflects the case of surveys using interviewers, enumerators etc. to collect the observations; the following discussion is, however, quite general.

(6) FOSDIC stands for "Film Optical Sensing Device for Input to Computers".

One device exploits an *independent* sample from the population. For each element in this sample, it is determined whether it was covered or not in the census survey. This provides a basis for the action called for. The device just described is especially powerful, if it is possible to measure the coverage errors of groups having a high risk of unsatisfactory coverage. For a real-life application, reference is given to U. S. Bureau of the Census [20].

An alternative device is the predecessor-successor check, as described in Hansen et al [13].

In this context, reference should in addition be given to various devices for coping with the special kind of coverage problem known as the non-response problem. For a review of some devices available, reference is given to Dalenius [4].

(2) With respect to *content errors*, we will present two useful devices.

One such device is replication: the procedure of observing an element is repeated one or several times. Replication has been extensively used as the basis for a quality control of the field operations in crop surveys in India, as discussed in Mahalanobis [17].

Another device is (unitary or aggregative) validation, that is, a comparison between an observed and a true (or possibly prefered) value. As a device for measuring control errors, validation is in principle unequalled. Its practical usefulness is, of course, restricted by the need for access to a true (prefered) value. For an interesting example of the use of validation, reference is given to Horn [16].

15. Quality control of the post-field operations. The post-field operations of a survey embrace such operations as coding, card punching, editing, and tabulation. In view of the nature of these operations — they are, in the realm of clerical work, counterparts to industrial mass production operations — it is not surprising that statistical quality control has proved applicable and useful. We will give some illustrative examples.

(1) The problems associated with quality control of coding may be illuminated by reference to two plans for such control: dependent and independent verification, respectively.

It is characteristic of dependent verification that the verifiers review the work of the production coders and thereby determine whether the codes assigned are correct or not. In specific instances — Hansen et al [9] provides evidence — it has been demonstrated that this kind of plan may provide rather inadequate control.

Independent verification calls for having the verifiers assigning codes independently of the production coders and thereafter comparing the result. This plan has proved considerably more powerful as a basis for quality control, as illustrated in Hansen et al [9].

(2) Statistical quality control was early applied to the card punching operation; a classic reference is Deming and Geoffrey [7].

In spite of the increased use of such techniques as "mark sensing" and FOSDIC, card punching is still an important task in surveys. By the same token, statistical quality control still finds important applications, as illustrated by U. S. Bureau of the Census [21].

(3) Editing is a third post-field operation, where quality control may profitably be applied.

While most of the problems represented by inaccurate or missing data in initial reporting are present irrespective of the data processing facilities available, the access to electronic computers has made it possible to carry out certain editing rules — earlier applied (if at all) by clerks — in a more satisfactory way: with higher precision, more rapidly, and at lower cost.

For an elementary review of some techniques used, reference is given to Dalenius [5]. For a comprehensive review, reference is given to Pritzker et al [19].

(4) Mention should finally be made of the possibilities of quality control of the preliminary results of a survey.

16. Quality control through evaluation programs. In par. 1, reference was given to two early instances of quality checks of census surveys. Today, comprehensive evaluation (and research) programs are undisputed elements of plans for large-scale census and sample surveys in many countries around the world.

One important objective of such a program is to provide comprehensive measures of the quality of the current survey (for example measures of the coverage error and the content error) corresponding to a *retrospective* quality control aspect.

Another important objective is to guide the statisticians in their future choice between alternative procedures, as well as in the development of new, improved procedures, corresponding to a *prospective* quality control aspect.

For a detailed analysis of the role of evaluation and research programs as a tool of quality control of surveys, reference is given to U. S. Bureau of the Census [20].

V. What is ahead?

17. The present situation. In the previous chapters, we have presented a broad review of an important aspect of current methodological development: the endeavours towards a unification of the two main lines of methodological development identified in par. 1.

Considerable progress has indeed been made towards this unification.

More progress is, however, needed to improve the existing tools, and to construct new tools. Especially, survey models have to be developed for descriptive surveys where the response error mechanism is different from the one operating in the model presented in par. 6, for example to cope with the phenomenon of "telescoping"; cf U. S. Bureau of Census [22]. Moreover, survey models have to be developed for surveys having as their purposes to estimate some other parameter than $\bar{X}$, for example for surveys to estimate change over time. There are reasons to believe that the efforts devoted to research concerning these aspects will provide useful results in the near future.

18. Towards a Survey Measurement System. In the last few years, the attack on the methodological front has been considerably broadened. A possible outcome of this attack is a comprehensive Survey Measurement System.

Following Hansen et al [14], we view a survey from the following three perspectives:

(i) The requirements.
(ii) The specifications.
(iii) The survey operations.

The requirements are imposed by the real-life problem, to the solution of which the survey is expected to contribute. Corresponding to these requirements, there is an ideal survey which — in principle if not in practice — would yield the ideal goal, a set of statistics, which we denote by $\bar{Z}$.

The specifications are, as discussed in par. 4, laid down in the survey design. The goal thus defined is denoted by $\bar{X}$. In practice $\bar{X}$ will differ from $\bar{Z}$, as specifications which are completely congruent with the requirements may not be operationally feasible or efficient.

The survey operations yield the statistics $\hat{\bar{y}}$, with the expected value $\bar{Y}$. Thus, $\hat{\bar{y}}-\bar{Z}$ is the error of the survey relative to the ideal goal.

We now generalize the survey model presented in par. 6. Thus, we use $E(\hat{\bar{y}}_t-\bar{Z})^2$ as the measure of error. We develop:

$$E(\hat{\bar{y}}_t-\bar{Z})^2 = E(\hat{\bar{y}}_t-\bar{X})^2+(\bar{X}-\bar{Z})^2+2(\bar{Y}-\bar{X})(\bar{X}-\bar{Z}).$$

The first term is the mean-square error of $\hat{\bar{y}}_t$ relative to the defined goal $\bar{X}$. This term is discussed in par. 6.

The remaining terms reflect the degree to which the survey provides *relevant* statistics: $(\bar{X}-\bar{Z})^2$ is the square of the bias of the survey specifications, while $2(\bar{Y}-\bar{X})(\bar{X}-\bar{Z})$ is an interaction term.

The use of this generalized survey model calls for answers to such questions as:

(i) What are the relevant statistics?

(ii) What are useful standards of relevance?

(iii) Which devices(7) are available for control of the mean-square error $E(\hat{\bar{y}}_t - \bar{Z})^2$?

Today, much research and developmental work is devoted to these and related questions.

19. Acknowledgements. The author wants to express his gratitude to Mr. Morris H. Hansen, Assistant Director for Research and Development, U. S. Bureau of the Census, and several members of his staff, for stimulating discussions of the topics discussed in this paper. The author is, however, sole responsible for the presentation.

This paper has been prepared within the frame-work of a research project supported by the Bank of Sweden Tercentenary Fund.

References cited in the text

[1] F. J. Anscombe, *The validity of comparative experiments*, J. Roy. Statist. Soc. A 61 (1948), pp. 181-200.

[2] G. Chevry, *Control of a general census by means of an area sampling method*, J. Amer. Statist. Assoc. 44 (1949), pp. 373-379.

[3] W. G. Cochran, *Sampling techniques*, 2nd ed., New York 1963.

[4] T. Dalenius, *Sampling in Sweden*, Stockholm 1957.

[5] T. Dalenius, *Automatic estimation of missing values in censuses and sample surveys*, Statistisk Tidskrift 11 (1962), pp. 395-400.

[6] T. Dalenius and B. Matérn, *Is there a need for a unified theory of random experiments?* Metrika 82 (1964), pp. 235-247.

[7] W. E. Deming and L. Geoffrey, *On sample inspection in the processing of census returns*, J. Amer. Statist. Assoc. 36 (1941), pp. 351-360.

[8] A. R. Eckler and L. Pritzker, *Measuring the accuracy of enumerative surveys*, Bull. Internat. Statist. Inst. 33 (1951), pp. 7-24.

[9] M. H. Hansen, H. H. Fasteau, J. J. Ingram and G. Minton, *Quality control in the 1960 censuses*, Proc. of 1962 Middle Atlantic Conference, American Society for Quality Control, Milwaukee, Wisconsin, 1962.

[10] M. H. Hansen and W. N. Hurwitz, *On the theory of sampling from finite populations*, Ann. Math. Statist. 14 (1943), pp. 333-362.

[11] M. H. Hansen, W. N. Hurwitz and M. A. Bershad, *Measurement errors in censuses and surveys*, Bull. Internat. Statist. Inst. 38 (1961), pp. 359-374.

[12] M. H. Hansen, W. N. Hurwitz and T. B. Jabine, *The use of imperfect lists for probability sampling at the U. S. Bureau of the Census*, Bull. Internat. Statist. Inst. 40 (1964), pp. 497-517.

[13] M. H. Hansen, W. N. Hurwitz and L. Pritzker, *The estimation and interpretation of gross differences and the simple response variance*. In: C. R. Rao (editor), *Contributions to statistics presented to Prof. P. C. Mahalanobis on the occasion of his 70th birthday*, Oxford 1964.

[14] M. H. Hansen, W. N. Hurwitz and L. Pritzker, *Standardization of procedures for the evaluation of data: measurement errors and statistical standards in*

(7) The approach to the control of the mean-square error discussed in par. 11 is perhaps meaningful only in the present context.

the Bureau of the Census, Paper presented at the 36th session of the International Statistical Institute, Sydney 1967.

[15] M. H. Hansen and J. Steinberg, *Control of errors in surveys*, Biometrics 12 (1956), pp. 462-474.

[16] W. Horn, *Reliability survey, A survey on the reliability of response to an interview survey*, Het PTT-bedrijf 10 (1960), pp. 105-156.

[17] P. C. Mahalanobis, *On large-scale surveys*, Philos. Trans. Roy. Soc. London B 231 (1944), pp. 329-451.

[18] J. Neyman, *On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection*, J. Roy. Statist. Soc. 97 (1934), pp. 558-606.

[19] L. Pritzker, J. Ogus and M. H. Hansen, *Computer editing methods — some applications and results*, Bull. Internat. Statist. Inst. 41 (1966) pp. 442-466.

[20] U. S. Bureau of the Census, *Evaluation and research program of the U. S. censuses of population and housing, 1960: Background, procedures, and forms*, Series E R 60, no. 1, Washington, D. C., 1963.

[21] U. S. Bureau of the Census, *Censuses of population and housing in Puerto Rico, 1960: processing the data*, Washington, D. C., 1963.

[22] U. S. Bureau of the Census, *Response errors in collection of expenditures data by household interviews: an experimental study*, Technical Paper no. 11, Washington, D. C., 1965.

[23] H. Wold, *Non-experimental statistical analysis from the general viewpoint of scientific method*, Paper presented at the 36th session of the International Statistical Institute, Sydney 1967.

[24] F. Yates, *Sampling methods for censuses and surveys*, 3rd ed., London 1960.

Some additional references

W. E. Deming, *On errors in surveys*, Amer. Sociol. Review 9 (1944), pp. 359-369.

A. R. Eckler and W. N. Hurwitz, *Responce variance and biases in censuses and surveys*, Bull. Internat. Statist. Inst. 36 (1958), pp. 12-35.

I. P. Fellegi, *Response variance and its estimation*, J. Amer. Statist. Assoc. 59 (1964), pp. 1016-1041.

M. H. Hansen and W. N. Hurwitz, *The problem of nonresponce in sample surveys*, J. Amer. Statist. Assoc. 41 (1946), pp. 517-529.

R. H. Hanson and E. S. Marks, *Influence of the interviewer on the accuracy of survey results*, J. Amer. Statist. Assoc. 53 (1958), pp. 635-655.

D. B. Lahiri, *Recent developments in the use of techniques for assessment of errors in nation-wide surveys in India*, Bull. Internat. Statist. Inst. 36 (1958), pp. 71-93.

P. C. Mahalanobis, *Recent experiments in statistical sampling in the Indian Statistical Institute*, J. Roy. Statist. Soc., 109 (1946), pp. 325-370.

G. L. Palmer, *Factors in the variability of response in enumerative studies*, J. Amer. Statist. Assoc. 38 (1943), pp. 143-152.

S. A. Rice, *Contagious bias in the interview*, Amer. J. Sociol. 35 (1929), pp. 420-423.

U. S. Bureau of the Census, *The accuracy of census statistics with and without sampling*, Technical Paper no. 2, Washington, D. C., 1960.

U. S. Bureau of the Census, *The current population survey reinterview program: some notes and discussion*, Technical Paper no. 6, Washington, D. C., 1963.

S. S. Zarkovic, *Quality of statistical data*, Vol. II of *Sampling methods and censuses*, Food and Agriculture Organization of the United Nations, Rome 1963.

INSTITUTE OF STATISTICS
UNIVERSITY OF STOCKHOLM

Received on 26. 10. 1967