Johnston

STATISTICAL PAPERS Series C No. 1, Rev.2

Recommendations for the Preparation of Sample Survey Reports (Provisional Issue)

UNITED NATIONS

DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICAL OFFICE OF THE UNITED NATIONS

STATISTICAL PAPERS Series C No. 1, Rev.2

Recommendations for the Preparation of Sample Survey Reports (Provisional Issue)

UNITED NATIONS
New York, 1964

NOTE

Symbols of United Nations documents are composed of capital letters combined with figures. Mention of such a symbol indicates a reference to a United Nations document.

ST/STAT/SER.C/1/Rev.2

UNITED NATIONS PUBLICATION

Sales No.: 64. XVII. 7

Price: $U.S. 0.35
(or equivalent in other currencies)

Inquiries should be directed to:

PUBLISHING SERVICE
UNITED NATIONS
NEW YORK, N.Y.

See also list of sales agents inside back cover

TABLE OF CONTENTS

FOREWORD

The first document on sampling terminology entitled The Preparation of Sampling Survey Reports was prepared by the Sub-Commission on Statistical Sampling of the Statistical Commission at its second session, 30 August to 11 September 1948. In its report to the Statistical Commission the Sub-Commission stated:

> "The Sub-Commission considered that if accounts of sampling investigations included the points enumerated, and if they dealt with the technical aspects of sampling processes in accordance with the recommended terminology, it would become increasingly possible to improve sampling practices in many important respects. The Sub-Commission believed that the wide circulation of reports, prepared generally in accordance with the suggestions contained in the memorandum, will foster international exchange of experience, and will suggest the use of sampling in various undertakings now being carried out by complete counts or by non-random partial surveys." (Document E/CN.3/52.)

The document was revised by the Sub-Commission at its third session, 12 to 23 September 1949.

The members of the Sub-Commission were: Mr. G. Darmois, Mr. W. E. Deming, Mr. P. C. Mahalanobis (Chairman), Mr. F. Yates and Mr. R. A. Fisher (Consultant).

The Statistical Commission, at its twelfth session held in 1962, considered that with the passage of time, development in many methodological aspects of sampling had rendered the Sub-Commission report somewhat out of date. It therefore recommended that the Secretary-General appoint a group of experts to undertake the preparation of a revised set of recommendations on sampling terminology, to be issued by the Secretary-General to supersede the document. In pursuance of this resolution a group of sampling experts was convened in Geneva from 1 to 12 July 1963.

The Group elected Mr. F. Yates as Chairman and Mr. P. C. Mahalanobis as Vice-Chairman. The other members of the Group were: Mr. T. Dalenius, Mr. N. Keyfitz, Mr. V. Monakhov, and Mr. P. Thionet. Mr. S. Zarkovic participated in the meetings as representative of the Food and Agriculture Organization of the United Nations. Mr. D. B. Lahiri of the Indian Statistical Institute was present as an observer. Mr. P. J. Loftus, Director of the Statistical Office, represented the Secretary-General and Mr. M. D. Palekar served as secretary of the Group.

The present document contains the recommendations of the Expert Group.

I. INTRODUCTION

These recommendations have been prepared in the hope that they will be of assistance to those engaged in preparing reports on sample surveys. They necessarily involve technical terminology, the use of which will foster clarity, comprehensiveness, and international comparability in reports that deal with aims, methods used, and accuracy attained. Information supplied on the lines suggested will enable those making use of the reports to utilize the results obtained to full advantage, to assess for themselves the reliability of these results, and to utilize the experience gained in the conduct of the survey in planning future surveys of similar nature. The recommendations are not intended to be final, but are to be regarded as an outline of the various points which require description and analysis in the application of sampling methods in various socio-economic fields. It may be mentioned that many of the recommendations are applicable also to censuses and other complete enumerations.

Three types of survey reports may be distinguished: (i) preliminary, (ii) general, and (iii) technical. Sections II to IV deal respectively with the nature and content of each type of report. Section V gives a brief description of the various types of sampling commonly employed, together with definitions of the technical terms used. Section VI gives a similar description of methods of selecting the sample. Section VII contains definitions and explanations of terms used in statistical analysis.

II. PRELIMINARY REPORT

A preliminary report is often required to make available data of current interest as rapidly as possible; such results may relate to selected important characters sometimes based on a sub-sample of the full sample. It should contain a brief statement concerning the survey methods and the limitations of the data. As a very minimum, information should be given concerning the size of the sample, the method of selecting the sample and discrepancies observed between external and internal data. Fuller details can be given in the general and technical reports on the survey.

III. GENERAL REPORT

This should contain a general description of the survey for the use of those who are primarily interested in the results rather than in the technical statistical aspects of the sample design, execution and analysis. The general description should include information on the following points:

1. Statement of purposes of the survey: A general indication should be given of the purposes of the survey and "margin of error" which would be permissible for the purposes of the survey, and the ways in which it is expected that the results will be utilized. In this connexion it is useful to distinguish the following types of surveys:

   (i) Integrated survey: In an integrated survey, data on several subjects (or items, or topics) are collected for the same set of sampling units for studying the relationship among items belonging to different subject fields. Such surveys are of special importance in studies on levels of living. Integrated surveys of consumption and productive enterprises are also of special importance in developing countries where the related activities are frequently undertaken in an integrated manner in the household.

   (ii) Multi-subject survey: When in a single survey operation several subjects, not necessarily very closely related, are simultaneously investigated for the sake of economy and convenience, the investigation may be called a multi-subject survey. The data on different subjects need not necessarily be obtained for the same set of sampling units, or even for the same type of units (e.g., households, fields, schools, etc.).

   (iii) Continuing surveys: The most usual example of these surveys occurs where a permanent sampling staff conducts a series of repetitive surveys which frequently include questions on the same topics in order to provide continuous series deemed of special importance to a country. Questions on the continued topics can frequently be supplemented by questions on other topics, depending upon the needs of the country.

   (iv) *Ad hoc* survey: This is a survey without any plan for repetition.

   (v) Multi-purpose survey: This term is sometimes used in connexion with sampling organizations which conduct surveys in various fields of interest to several departments or parties, keeping in view their diverse purposes. This permits economies in the technical and other resources and their more effective use, particularly in developing countries.

   (vi) Specialized or special-purpose survey: This may be defined as an investigation focussing on a single set of objectives which, because of their nature or complexity, requires specialized knowledge or the use of special equipment by a technical staff with training in the subject field of enquiry.

2. Description of the coverage: An exact description should be given of the geographic region or branch of the economy or social group or other categories of constituent parts of a population covered by the survey. In a survey of a human population, for example, it is necessary to specify whether such categories as hotel residents, institutions (e.g., boarding houses, sanatoria), persons without fixed abode, military personnel, were included and to indicate the order of magnitude of the categories omitted. The reporter should guard against any possible misapprehension regarding the coverage of the survey.

3. Collection of information: The nature of the information collected should be reported in considerable detail, including a statement of items of information collected but not reported upon. The inclusion of copies of the questionnaire or other schedules, and relevant parts of the instructions used in the survey (including special rules for coding and classifying) is of great value, and such documents should therefore be reproduced in the report if possible.

   The information may be collected by direct investigation, or by mail or telephone. Direct investigation may involve objective methods of observation or measurement. The method of collection should be reported, together with the nature of steps taken to ensure that the information is as complete as possible (e.g., methods of dealing with non-response). The extent and causes of non-response, etc. should be stated.

   It is of importance to describe the type and number of investigators, e.g., whether whole or part-time, permanent or temporary, with particulars of their training and qualifications.

4. Repetition: It is important to state whether the survey is an isolated one or is one of a series of similar surveys. Where the survey is repetitive and some of the sampling units reappear in the successive stages, this should be stated.

5. Numerical results: A general indication should be given of the methods followed in the derivation of the numerical results. Particulars should be given of methods of weighting and of any supplemen-

tary information utilized, for example to obtain ratio estimates. Any special methods of allowing for non-response should be described.

6. Date and duration: There are two periods of time which are important for any survey: (i) the period to which data refer, or reference period, (ii) period of collection, that is, the period taken for the field work.

   In order to minimize memory lapses, the length of the reference period is sometimes fixed but not the end points, as in "preceding week" questions (the collection period being comparatively long). In such cases, the reference period may be called moving in contrast with a fixed reference period when the end points are fixed. The reference period, if properly selected, enhances the use and value of the information collected in the survey. Sometimes different reference periods are used for different topics in the same survey. This is done with a view to eliciting more accurate information from the respondents, as in a family budget survey, where questions on food items may be asked for the preceding week or month, but information on clothing or furniture or some durable goods may be asked for the whole year. In health surveys, different reference periods are used for different items, e.g., illness during the preceding two weeks, hospitalization during the year. When the field work is conducted on a continuous and successive basis, the parts of the survey which are operationally separate are called rounds.

7. Accuracy: A general indication of the accuracy attained should be given and a distinction should be made between sampling errors and non-sampling errors.

8. Cost: An indication should be given of the cost of the survey, under such headings as preliminary work, field investigations, analysis, etc. Resources used in the conduct of the survey but not included in the costs should be stated.

9. Assessment: The extent to which the purposes of the survey were fulfilled should be assessed.

10. Responsibility: The names of the organizations sponsoring and conducting the survey should be stated.

11. References: References should be given to any available reports or papers relating to the survey.

IV. TECHNICAL REPORT

Technical reports should be issued for surveys of particular importance and those using new techniques and procedures of special interest. In addition to covering such fundamental points as the purposes of the survey, conditions to be fulfilled and resources available for the survey, the report should deal in detail with technical statistical aspects of the sampling design, execution and analysis; the operational and other special aspects should also be fully covered.

1. Specification of the frame: A detailed account of the specification of the frame should be given; this should define the geographic areas and categories of material included and the date and source of the frame. If the frame has been emended or constructed *ab initio* the method of emendation or construction should be described. Particulars should be given of any known or suspected deficiencies.

2. Design of the survey: The sampling design should be carefully specified, including details such as types of sampling unit, sampling fractions, particulars of stratification, etc. The procedure used in selecting sampling units should be described and if it is not by random selection the reporter should indicate the evidence on which he relies for adopting an alternative procedure.

3. Personnel and equipment: It is desirable to give an account of the organization of the personnel employed in collecting, processing and tabulating the primary data, together with information regarding their previous training and experience. Arrangements for training, inspection and supervision of the staff should be explained; as also should methods of checking the accuracy of the primary data at the point of collection. A brief mention of the equipment used is frequently of value to readers of the report.

4. Statistical analysis and computational procedure: The statistical methods followed in the compilation of the final summary tables from the primary data should be described. If any more elaborate processes of estimation than simple totals and means have been used, the methods followed should be explained, the relevant formulae being reproduced where necessary.

   It frequently happens that quantities of which estimates are required do not correspond exactly to those observed; in a crop-cutting survey, for example, the yields of the sample plots give estimates of the amount of grain, etc., in the standing crop, whereas the final yields will be affected by losses at various stages, such as harvesting, storing, transport, marketing, etc. In such cases adjustments may have to be made, the amount of which is estimated by subsidiary observations, or otherwise. An account should be given of the nature of these adjustments and the ways in which they were derived.

   The steps taken to ensure the elimination of gross errors from the primary data (by scrutiny, sample checks, etc.) and to ensure the accuracy of the subsequent calculations should be indicated in detail. Mention should be made of the methods of processing the data (punched cards, hand tabulation, etc.) including methods used for the control of errors.

   In recent years electronic computers have been applied to the making of estimates from sample surveys. In addition, they have been programmed to tasks of editing with consequent improvement of consistency of the primary data and possibly of its accuracy. Their use goes beyond the mere speeding up of existing methods and lowering cost. They have made possible some forms of estimation (e.g., the fitting of constants in analysis of variance with unequal numbers in the sub-classes) on a scale which would have been impracticable with hand methods. In some instances regression on previous surveys of a series with overlapping samples can substantially improve the precision of estimates. The use of these computers makes possible changes in the allocation of resources between the collection of the data and the processing.

   The amount of tabular matter included in the report, and the extent to which the results are discussed, will depend on the purposes of the report. If a critical statistical analysis of the results embodied in the final summary tables has been made, it is important that the methods followed should be fully described. Numerical examples are often of assistance in making the procedure clear. Mention should be made of further tabulations which have been prepared but are not included in the report, and also of critical statistical analyses which failed to yield results of interest and which are therefore not considered to be worth reporting in detail.

   The inclusion of additional numerical information which is not of immediate relevance to the report but which will enable subsequent workers to carry out critical statistical analyses which appear to them to be of interest should be carefully considered. If, for example, in addition to the class means of each main classification of the data, the sub-class numbers (but not the means) of the various two-way classifications are reported, a study of the effects of each of the main classifications freed from the effects of all other classifications can be made (provided the effects are additive) without further reference to the original information.

5. Accuracy of the survey:

(i) Precision as indicated by the random sampling errors deducible from the survey: Standard deviations of sampling units should be given in addition to such standard errors (of means, totals, etc.) as are of interest. The process of deducing these estimates of error should be made entirely clear. This process will depend intimately on the design of the sample survey. An analysis of the variances of the sampling units into such components as appears to be of interest for the planning of future surveys, is also of great value.

(ii) Degree of agreement observed between independent investigators covering the same material: Such comparison will be possible only when interpenetrating samples have been used, or when checks have been imposed on part of the survey. It is only by these means that the survey can provide an objective test of possible personal equations (differential bias among the investigators). Any such comparisons or checks should be fully reported.

(iii) Other non-sampling errors: The existence and possible effects of non-sampling errors on the accuracy of the results, and of incompleteness in the recorded information (e.g., non-response, lack of records, whether covering the whole of the survey or particular areas or categories of the material), should be fully discussed. Any special checks instituted to control and determine the magnitude of these errors should be described, and the results reported.

Another source of error is that due to incorrect determinations of the adjustments (referrred to in Section VII-4) arising from observation of quantities which do not correspond exactly to the quantities of which estimates are required.

6. Accuracy, completeness and adequacy of the frame: The accuracy of the frame can and should be checked and corrected automatically in the course of the enquiry, and such checks afford useful guidance for the future. Its completeness and adequacy cannot be judged by internal evidence alone. Thus, complete omission of a geographic region or the complete or partial omission of any particular class of the material intended to be covered cannot be discovered by the enquiry itself and auxiliary investigations have often to be made. These should be put on record, indicating the extent of inaccuracy which may be ascribable to such defects.

7. Comparisons with other sources of information: Every reasonable effort should be made to provide comparisons with other independent sources of information. Such comparisons should be reported along with the other results, and the significant differences should be discussed. The object of this is not to throw light on the sampling error, since a well designed survey provides adequate internal estimates of such errors, but rather to gain knowledge of biases, and other non-random errors.

Disagreement between results of a sample survey and other independent sources may of course sometimes be due, in whole or in part, to differences in concepts and definitions or to errors in the information from other sources.

8. Costing analysis: The sampling method can often supply the required information with greater speed and at lower cost than a complete enumeration. For this reason, information on the costs involved in sample surveys are of particular value for the development of sample surveys within a country and are also of help to other countries.

It is therefore recommended that fairly detailed information should be given on costs of sample surveys. Costing information should be given under such headings as planning (showing separately the cost of pilot studies), field work, supervision, processing, analysis and overhead costs. In addition, labour costs in man-weeks of different grades of staff and also time required for interviewing, travel and transport costs should be given. The collection of such information is often worthwhile, since it may suggest methods of economizing in the planning of future surveys. Moreover, the preparation of an efficient design involves a knowledge of the various components of costs as well as of the components of variance. It should be emphasized that the concept of cost in this respect should be regarded broadly in the sense of economic cost and should therefore take account of indirect costs which may not have been charged administratively to the survey. Wherever possible, the costing data should distinguish the time and resources devoted to the various operations involved in the survey.

9. Efficiency: The results of a survey often provide information which enables investigations to be made on the efficiency of the sampling designs, in relation to other sampling designs which might have been used in the survey. The results of any such investigations should be reported. To be fully relevant the relative costs of the different sampling methods must be taken into account when assessing the relative efficiency of different designs and intensities of sampling.

Such an investigation can be extended to consideration of the relation between the cost of carrying out surveys of different levels of accuracy and the losses resulting from errors in the estimates provided. This provides a basis for determining whether the survey was fully adequate for its purpose, or whether future surveys should be planned to give results of higher or lower accuracy.

10. Observations of technicians: The critical observations of technicians in regard to the survey, or any part of it, should be given. These observations will help others to improve their operations.

V. DESIGN OF THE SURVEY

1. Elementary units: These are the smallest physical units which are capable of possessing one or more characteristics which are to be observed or measured in the survey. Elementary units of different types may therefore occur in the same survey, and one type may comprise another. Thus, number of children is a characteristic of a family, age is a characteristic of an individual; and families comprise individuals. Individual elementary units or groups of elementary units used for recording are sometimes called recording units.

2. Sampling units (or sample units): These are the units which form the basis of the process of sampling from the aggregate of elementary units intended to be covered in the survey. The sampling unit may be (i) the same as the elementary unit, or (ii) a group of elementary units. A group may consist of contiguous elementary units or of units arranged in an assigned configuration. A systematic pattern of elementary units may, for instance, constitute a sampling unit.

   It is important that the sampling units be so defined that it is possible to identify them clearly, correctly and unambiguously. For example, in enquiries concerning the level of living, the sampling unit may be a family of which a possible definition would be a group of persons living together, related by blood or marriage, and having a common family budget.

   It is also conceptually convenient that they be so defined that the totality of sampling units contains every elementary unit once, and once only. If this condition is departed from (e.g., by the use of circular areas in area sampling), attention must be paid to the question of whether bias is likely to arise owing to boundary conditions or other factors.

   It is important to specify the quantitative and qualitative characteristics of the sampling units which are to be recorded in the enquiry, as these form the basis of all subsequent analysis, and in the cases of multi-stage and multi-phase sampling, may also have been used in determining the sampling procedure.

   It may often be possible to associate a measure of size with the sampling unit. Sampling units may be of the same or different size. They may also contain the same, or approximately the same, number of elementary units, or they may contain widely different numbers. In multi-stage sampling, the sampling units for each stage are sub-divisions of the units for the previous stage.

3. Physical field (or field of interest): The physical field or field of interest comprises a finite number of elementary units of one or more types (or in special cases a finite continuum which is divided into arbitrary parts which constitute elementary units for the purpose of survey) and which are located within a given boundary demarcating the field intended to be covered. Thus, the boundaries of the physical field may be the same as those of a country or geographic area, and the elementary units may be human beings, families, schools, villages, farms, fields, enterprises, etc., "lying within the specified boundaries".

4. Frame: The frame consists of previously available descriptions of the objects or material related to the physical field in the form of maps, lists, directories, etc., from which sampling units may be constructed and a set of sampling units selected; and also information on communications, transport, etc., which may be of value in improving the design for the choice of sampling units, and in the formation of strata, etc. Supplementary information available for the field covered by the frame may also be of value in improving the accuracy of estimates obtained from the survey. Frames that are initially available often require emendation, particularly in the later stages of multi-stage sampling, before they may be considered adequate, and even then may be incomplete; at times a frame may require to be constructed *ab initio*. In repeated surveys information collected in the earlier surveys may often serve to improve the frame for later surveys.

5. Field of enquiry: The entire field of enquiry is co-extensive with the frame used as its basis. If the frame is incomplete the field of enquiry will not coincide exactly with the field of interest.

6. Population (or universe or aggregate):1/ The term population is used in several senses which must be clearly distinguished. In addition to its common usage, e.g., the human population of a country, it is used in statistical terminology to denote an aggregate of elementary or sampling units, finite in number, or an aggregate of numerical quantities or qualities appertaining to these units, such as the heights or sexes of the individuals of a human population, not all of which may be known, but all of which are conceptually determinable, or to denote numerical quantities which are generated by the process of random sampling from a defined distribution, and which are conceptually infinite in number, e.g., random sample values from a normal distribution.

1/ English sampling terminology has taken terms like "population" from common usage and given them a technical meaning. The preservation of this identity with common usage is not necessary in other languages. Suitable words, if necessary with different meanings, should, therefore, be used in translations to make the technical meaning clear in each case in each language.

In sample surveys the term population is restricted to the following uses:

(i) When there is one type of elementary unit, as equivalent to the field of interest or enquiry. (The term universe is sometimes also used). When there are two or more types of elementary units, these would be regarded as constituting two or more populations.

(ii) Occasionally to denote the totality of sampling units of a given type constructed from the elementary units.

(iii) To denote the totality of numerical quantities of a particular type relating to individual elementary units or sampling units of a given type. These numerical quantities may be subject to errors of measurement; in this case there is associated with each unit a conceptually infinite population of measurements, of which the actual measurement constitutes a sample of one value (or several values if the measurement is replicated).

7. Domain of study (or sub-population): Any sub-division of the field of enquiry about which the survey is used to supply information may be termed a domain of study. The sub-divisions may coincide with those used for stratification, or they may differ. It is desirable to indicate the smallest domains of study about which the enquiry may be expected to provide information of adequate accuracy. It should be noted that stratification will not greatly reduce the sampling errors of domains of study that cut across strata.

8. The sample: The aggregate of the sampling units selected constitutes the sample. A random or probability sample supplies both objective estimates (of means, totals, etc.) and, if certain conditions are satisfied, information for estimating their precision. The word sample may be used either for the aggregate of sampling units chosen from a single stratum or at a single stage, or in a wider sense, for the whole of the sampling units chosen in a complete enquiry.

9. Simple random sample (or unrestricted random sample, or fully random sample): A simple random sample is one in which the sampling units are independently selected with equal probability, the sampling being "without replacement" and without any restrictions. This is equivalent, for a sample of n units from a population of N units, to the selection with equal probability of one out of all possible combinations of n out of N units. The term random sample (without qualification) is sometimes used to denote a simple random sample.

10. Sampling with probability proportional to size: In certain cases it may be advantageous, or convenient, to select the sampling units from the whole population or from within strata with probabilities proportional to some known quantitative characteristic of the units, such as size, or indeed some function of this quantity, e.g., the square root; if more than one unit is to be selected from the population (or from individual strata) exact probabilities proportional to size can only be simply attained if the sampling is "with replacement", units selected twice or more being counted twice or more.[2/]

The methods of calculating the population estimates appropriate to samples selected with probability proportional to size differ from those for samples selected with uniform probability.

11. Stratification: The totality of sampling units included in the frame may be divided into groups or strata, each stratum being sampled separately and independently, so that a specified number of sampling units or a specific fraction of all such units is obtained from each stratum. Such strata may be geographic sub-divisions, branches of the economy, or other divisions dependent on some quantitative or qualitative variate appertaining to the sampling units, etc.

12. Uniform and variable sampling fractions: The numbers of sampling units specified may be such that from each stratum the same fraction of unit is selected, in which case the term stratification with uniform sampling fraction or proportionate sampling, is used, or the numbers specified may be such that different fractions of the different strata are selected, in which case the term stratification with variable sampling fraction is used. In this sense, the sample is usually a specified fraction of the stratum it represents.

13. Multiple stratification: In certain cases the totality of sampling units may be divided simultaneously according to two or more classifications, each of which depends on one quantitative or qualitative variate. Each cell determined by the two or more way classification, itself potentially constitutes a stratum from the totality of sampling units. If each of the cells is sampled separately, as in an ordinary stratified sample, the term multiple stratification may be used without qualification. Thus, one may stratify farms according to size and according to geographic regions. If the farms in each region are classified into size groups before taking the sample, then the region-size-group combinations form the individual sub-strata. If the available information is not adequate for this to be done, so that the numbers of sampling units in the main strata only can be pre-determined, this has been termed multiple stratification without control of sub-strata, but is better termed marginal stratification. In these circumstances, however, the selection of a sample presents both theoretical and practical difficulties, and a calculation of the sampling error is also troublesome.

14. Balanced sample: If the average value of some character to be used for purposes of control is known for units that make up the whole population, it is possible, provided the value of the character is

---

2/ A method of overcoming the difficulties that arise in sampling without replacement is described by J. N. K. Rao, H. O. Hartley, and W. G. Cochran in "On a Simple Procedure of Unequal Probability Sampling Without Replacement." Journal of the Royal Statistical Society, Series B, Vol. 24, No. 2 (1962), pp. 482-491.

known for each sampling unit, to select a sample in such a manner that the average value of this character for the units selected approximates, within an assigned margin, to the average for the population. Provided the numbers of units selected from the individual strata are adequate, a stratified sample may be separately balanced for each stratum.

15. Multi-stage sampling: In multi-stage (two-stage, three-stage, etc.) sampling, the frame is first divided into sampling units, referred to as first-stage units (or primary sampling units). A sample of such units is selected. Next, each such unit selected is further divided into second-stage units (or secondary sampling units). A sample of such units is then selected from each first-stage unit selected. This procedure may be extended to include three, four, etc., stages.

In multi-stage sampling a complete frame of first-stage units is required: for each first-stage unit selected, a frame of second-stage units is required, etc. One of the advantages of multi-stage sampling is that second-stage frames are required only for the first-stage units selected, and so on. When the frame of the last stage is constructed in the field the last-stage units are sometimes termed listing units.

For example, a country may be considered as divided into a number of districts: each district into a number of villages: each village into a number of farms. In multi-stage sampling a number of districts is selected in the first stage: within each such selected district a number of villages is selected in the second stage, and from each selected village a number of farms is selected in the third stage for enquiry. In the case of a crop-cutting investigation, the work may be carried further by the selection of fields from each selected farm and by plots within a field.

The devices of stratification and a variable sampling fraction may be used at any stage. If a variable sampling fraction at the first stage is used, the sampling fractions at the later stages may be determined so as to give a uniform over-all sampling fraction. If the selection is with probability proportional to size at the first stage (see Section V-10), the sampling fractions at the later stages may be similarly determined to give equal probability of selection to all elementary units.

16. Controlled sampling: This term is defined to mean any process of selection in which, while maintaining the assigned probability for each unit, the probabilities of selection for some or all preferred combination of n out of N units are larger than in stratified random sampling (and correspondingly the probabilities of selection for at least some non-preferred combinations are smaller than in stratified random sampling).

Controlled sampling is mainly practicable and useful when the sample is made up of a few large first-stage units. It then enables additional control to be introduced beyond what is possible by stratification alone, this being limited by the fact that the number of strata cannot exceed the number of units in the sample. No reliable estimate of the first-stage sampling error is possible for a controlled sample.[3/]

17. Multi-phase sampling: It is sometimes convenient and economical to collect certain items of information on the whole of the units of a sample, and other items of information on only some of these units, these latter units being so chosen as to constitute a sub-sample of the units of the original sample. This is termed two-phase sampling. Information collected at the second or sub-sampling phase may be collected at a later time, and in this event, information obtained on all relevant units of the first-phase sample may be utilized, if this appears advantageous, in the selection of the second-phase sample. Further phases may be added as required.

An important application of multi-phase sampling is the use of the information obtained at the first phase as supplementary information to provide more accurate estimates (e.g., by the method of regression or ratios), of the means, totals, etc., of variates on which data are obtained only in the second phase.

Information obtained in a complete census may be used in this manner to improve the estimates obtained from a sample, in which case the complete census and sample are analogous to the first and second phase sample respectively.

Multi-phase sampling may be combined with multi-stage sampling. In a scheme for the estimation of the areas and yields of an agricultural crop, for example, a two-stage sample of villages and farms may be taken for the estimation of areas, and a sub-sample of these farms may be taken for the estimation of yields. This is two-phase at the second stage.

An extension of multi-phase sampling is the use of information obtained from first-phase units for the construction of new sampling units made up of groups of the first-phase units. These are then used as a basis for one-or-more-stage sampling.

18. Interpenetrating (networks of) samples: Whatever be the basic design of the survey (with or without stratification: single- or multi-stage; single- or multi-phase), it is possible if desired to arrange the sampling units in sets of two or more interpenetrating networks of units, each of which constitutes a sub-sample, and to collect and process the information for each such sub-sample in an independent manner so that each sub-sample would supply an independent estimate of the population values of the variates under study. In some cases the interpenetrating sub-samples may be regarded as analo-

3/ Details are given by Roe Goodman and Leslie Kish in "Controlled Selection: A Technique in Probability Sampling." Journal of the American Statistical Association, Vol. 45, No. 249 (March, 1950), pp. 350-372.

gues of plots in the theory of design of experiments, and an analysis of variance can be carried out in the usual way. Another term for this procedure is replicated sampling.

It is possible to keep some of the sampling units the same in two (or more) interpenetrating networks of samples so that information for such sampling units is independently collected twice (or more than two times). It is then possible to make detailed comparisons between the two (or more) sets of observations. Such sampling units which are observed more than once are called duplicated (triplicated, quadruplicated, etc.) sampling units.

Interpenetrating samples can be used to secure estimates of sampling errors and information on non-sampling errors such as differences arising from differential interviewer bias, different methods of eliciting information, etc.

The interpenetrating network therefore provides a means of control (i.e., appraisal) of the quality of the information. For example, in a family budget enquiry, in each domain of study the sampling units may be chosen in the form of two or more independent sets each of which covers the whole domain, and the information for every such set may be collected by a different investigator. In this way, an independent estimate for the whole domain would be obtained based on the information collected by each investigator, and a comparison of such independent estimates (for example by an analysis of variance) would show whether there were significant differences between different investigators which may often indicate appropriate action for adjustments of the results of the survey or for improving the planning of similar surveys in the future.

The contrasts between the estimates provided by the sub-samples of an interpenetrating network can be used to provide estimates of the sampling error for these estimates. These estimates of error include any contributions due to coding and computational errors and those due to differential bias between observers (and also differential bias in coding if separate coders are allocated to the different sub-samples). In a stratified sample independent contrasts will be available from each stratum, and these will often provide a basis for reasonably reliable estimates of the error of the whole sample. Similar estimates of error (but excluding observer bias) can in many cases be obtained from a sample which was not planned as an interpenetrating network by sub-dividing the sample into sub-samples in some appropriate random manner.

19. Probability sampling: The selection of sampling units (at a given stage) is said to be carried out by probability sampling, if each sampling unit has a known, non-zero, probability of being selected. The term embraces many very different types of sampling, e.g., random sampling of all kinds, systematic sampling with random starting point, and controlled sampling, and is therefore unsuitable for use in reports of surveys without specification of the actual type, except when the purpose is to emphasize that methods like purposive sampling are not being used.

20. Cluster sampling: This is a general term which includes single-stage sampling with sampling units formed of groups or clusters of elementary units (usually of compact form to facilitate the collection of information), and two-or-more-stage sampling, in which the units at one stage are clusters of elementary units, these clusters being sub-sampled at the next stage. The report should specify clearly the sampling procedure used since the term embraces many different procedures.

21. Sub-sampling: This term has been used in several senses: (i) to denote the selection of, for example, second-stage units from a selected first-stage unit; (ii) to denote the selection, for example, at the second phase of two-phase sampling; and (iii) to denote the selection of the individual samples that constitute an interpenetrating network of samples.

The term is a loose one which should not be used except where its meaning is clear from the context.

22. Area sampling: This term refers to the use of areas as sampling units. The areas may be administrative units, or naturally bounded areas (such as "segments", "blocks", etc.), or especially demarcated areas (such as circular or rectangular plots used in crop estimation surveys). The areas observed may, therefore, be selected by reference to a map (which constitutes the frame) or determined in the course of the field operation. The areas may be directly observed as in land use surveys, or they may be used to define and observe the group of units (e.g., households) associated with the respective areas.

23. Line sampling: A method of sampling a geographic area using a line to select sampling units at any stage. Parallel lines are taken, either equally (systematically) spaced, or randomly located within blocks. The lines may be directly observed as in forest surveys, or intercepts may be measured to give proportions of the area falling in different categories.

24. Point sampling: A method of sampling a geographic area, using, at any stage, points located at random or systematically, to select the sampling units. The points, or small areas demarcated with reference to them, may themselves constitute sampling units, as in forest surveys, or they may be used to select sampling units, e.g., farms, with previously determined boundaries, a unit being selected if a point falls within it, in which case probability of selection is proportional to the area of the unit. Simple counts of the points falling in different categories will give estimates of the proportions of the area falling in these categories.

25. Selection of the sample by place of work: This term is used in the USSR in sample surveys of the level of living of families of any given social group. First, enterprises, institutions, and organiza-

tions are selected; next, within the selected enterprises, institutions and organizations, workers are selected and finally, for the families of the selected workers, record is kept of such items as incomes, expenditures, food consumption, acquisition of goods, housing, education, and other characteristics.

26. Composite sampling schemes: There are occasions on which different methods of sampling are required for different parts of the material. In sampling a human population, for instance, some form of sampling of areas (in which the sampling unit is a small area) may be most suitable for the rural parts of the country whereas some form of sampling based on lists of households may be best in the towns.

27. Successive (or repeated) surveys: Often a sample survey of the same kind is repeated at suitable intervals to constitute a series of successive surveys, or a continuing survey with successive rounds. In the USSR successive surveys repeated over a long period on a uniform plan are called systematic surveys. In successive surveys it is possible to choose entirely independent sampling units in each survey, or to keep a certain number (or a certain proportion) of sampling units the same in two or more surveys, e.g., one third of the units may be replaced in each survey, each unit then being retained for three surveys. This latter procedure is particularly useful for the study of changes when there is a high correlation, for individual sampling units, between successive periods. However, it should be recognized that retention of the same units for a long series of surveys presents practical and theoretical difficulties since the frame is changing; repeated surveys on the same units may also lead to atypical changes in these units. The continuing operation of a survey agency on successive surveys of the same type usually also leads to a gradual improvement in the quality of the information.

28. Longitudinal surveys: This term is used, especially in sociological and medical investigations, to denote a survey in which a sample of individuals (or other units) is observed at intervals over a period of time, so as to study their development as individuals. It may be contrasted with a cross-sectional survey, in which a sample of individuals in various stages of development, e.g., children of different ages, is observed at one time. A cross-sectional survey only furnishes information on changes in average values with age, etc. The sudden spurt of growth that occurs in a child at puberty, for example, will be masked in such a survey, since puberty occurs at different ages in different children.

29. Pilot and exploratory surveys: In undertaking large scale surveys, particularly of unexplored material, it is usually advisable to conduct pilot and exploratory surveys to test and improve field procedure, and schedules, and to train field workers; also to obtain information which will enable the sampling design to be planned more efficiently, and to obtain an estimate of the cost. For example, the results of a pilot survey may be used to estimate the first- and second-stage components of variance relevant to some two-stage sampling process which is envisaged for the main survey and also the relevant components of cost, from which it is possible to determine the optimum intensity of sampling at each of the two stages.

A distinction ought to be drawn between pilot surveys and exploratory surveys. Pilot surveys must necessarily be on a random or probability sample basis, whereas exploratory surveys need not necessarily be so.

VI. METHODS OF SELECTING THE SAMPLE

1. Random sampling: A process is properly described as random if each sampling unit is independently selected with determinate probability. One expeditious way of effecting a random selection is by the use of random sampling numbers; equally, with more labour, this may be done by any of the apparatus used in games of chance.

   The exact method of selection will depend on the type of sampling adopted. Thus, in a simple random sample, the required number of units is selected "without replacement" (i.e., no unit is selected twice) from all units in the population, each unit being independently selected with equal probability; this may be done by "drawing lots" or any equivalent process. In a stratified sample a selection of the requisite number of units is made from each stratum independently, the procedure for each stratum being the same as is used for the whole population in selecting a simple random sample. If selection with probability proportional to size is required the actual selection can be made by forming a running total of the sizes of the separate units: if the total of all units is T one can then (using a table of random numbers), select numbers at random between 1 and T; a number greater than the running total for unit s-1, and not greater than the running total for unit s, indicates that unit s is to be selected.

2. Selection of balanced and controlled samples and samples with marginal stratification: A balanced sample may be selected by using a process of replacement. A random or stratified random sample of the required size is first selected, the order of selection being recorded, and the average value of the known quantitative character is calculated for this sample. If this average is, say, greater than the average value for the population, each unit of the original sample which has a value greater than the average value for the population is taken in turn and replaced by a newly selected unit which has a value less than the average for the population; if the new selection gives a unit with value greater than average the selection is repeated until one less than average is obtained. The process is continued until an adequate degree of balance is attained.

   A sample with marginal stratification is selected by a somewhat similar but more complicated process using rejection instead of replacement.[4/] Selection of a controlled sample is also complicated.[5/]

3. Systematic sampling: A wholly or partly deterministic procedure for the selection of a sample, in which the selection of any particular sampling unit in the frame depends solely upon its relative place in the frame. The exact nature of the procedure will depend upon the kind of frame available. Four examples will illustrate this point:

   (i) Systematic sampling from a list: A systematic sample may be selected by taking every k'th entry. If the first entry is selected at random, the selection is termed systematic sampling with random starting point.

   (ii) Systematic sampling from an ordered sequence: If a list of sampling units in the frame is reordered, before the selection of the sample according to some important variate, (e.g., in a list of workers according to the incomes of the individual workers) a systematic sample from the ordered list may be expected to be more accurate than one of the same size selected from the original list without ordering. Similar considerations apply when a sequence is naturally ordered, e.g., in sampling a line stretching up a hill-side in a forest.

   (iii) Systematic sampling with probability proportional to size: Instead of selecting numbers at random between 1 and T (Section VI-1) every k'th number may be taken, with, if desired, a random starting point as in (i). The list of sampling units may be reordered as in (ii), if this is considered appropriate, before computing the running totals.

   (iv) Systematic sampling from areas: A systematic sample may be selected by placing a regular grid of points over the area, and selecting for the sample those units (farms, etc.) on which these points fall.

   Systematic selection is often used when the person responsible for the planning of the survey is satisfied that it is in practice equivalent to a random selection in the respects required. In such cases he accepts personal responsibility for the judgment on which his plan is based.

4. Purposive selection: In a purposively selected sample one or more control variates are chosen, for which the values are known for all sampling units of the population. A sample is then selected which matches the population in some defined manner with regard to these control variates.

4/ Details are given by F. Yates in Sampling Methods for Censuses and Surveys. 2d ed. revised and enlarged. New York, Hafner Publishing Company, 1953.

5/ Details are given by M. H. Hansen, W. N. Hurwitz, and W. G. Madow in Sample Survey Methods and Theory. Vol. I: Methods and Applications. New York, John Wiley and Sons, Inc., 1953.

Thus with a single control variate the sample may be selected so that:

(i) The average of the control variate in the sample agrees approximately with the average for the population; if the control variate refers to some type of elementary unit contained in the sampling units, e.g., average income of individuals, a weighted average (with weights equal to the numbers of elementary units in the sampling units) may be taken;

(ii) The distribution of the control variate for the units of the sample agrees with that for all the units of the population;

(iii) Conditions similar to (i) or (ii) are satisfied for each stratum separately.

It is important to realize that the matching, however elaborate, of the sample and the population with respect to one or more control variates provides no guarantee that the sample is representative for the variates under study, nor is the sample equivalent to a random sample. Nor is there any way of estimating the accuracy attained. Only by introducing appropriate random components into the selection process can a representative sample of ascertainable accuracy be obtained. Thus, condition (i) can be satisfied by using a balanced sample, condition (ii) by using a stratified sample with strata formed by grouping according to values of the control variate. For these reasons purposive selection has ceased to be used when objective results of ascertainable accuracy are required.

5. Quota sampling: A type of purposive selection, mostly used for opinion surveys and market research, is known as quota sampling. Each interviewer is given a definite quota of people to interview, e.g., in different social classes, or age groups. An interviewer may, for example, be instructed to interview 100 people, 50 men and 50 women, 20 each of the men and women to be between 21 and 34 years of age, the remaining 30 to be 35 years of age or older. Except for these restrictions, he is free to select any person for interview.

Quota sampling suffers from the same defects as other forms of purposive selection.

6. Mechanical sampling (a method used in the USSR): This is known as typical proportional sampling with mechanical selection of sampling units within typical groups.[6/] This method combines the elements of typical and mechanical selection. Before the selection, the population is divided into typical groups; within each group the units of the population are arranged in a definite order according to one of their most important characteristics, and the sampling units are selected mechanically so as to eliminate the possibility of human preferences to affect the selection.

6/ Typical groups in this sense refer to branches of industry and wage levels and are, therefore, domains of study or strata.

VII. STATISTICAL ANALYSIS

1. Estimate: This term is used technically to denote any quantity, calculated from the observed values, which is designed to give a value representing some numerical property of the population. Thus, in an unrestricted random sample the mean of the observed values provides an estimate of the mean of the whole population; the total of the sample values divided by the proportion of units sampled gives an estimate of the population total. Such population values are sometimes referred to as parameters of the population, but this term is better reserved for the quantities specifying distributions, e.g., the normal distribution, occuring in statistical theory.

   The appropriate method of estimation is in part determined by the method of sampling adopted; in part by the amount of supplementary information which is available and can be profitably utilized to improve the accuracy of the estimates; and sometimes also by the fact that more accurate estimates for comparative purposes are provided by biased estimates (which are usually also more easily calculable) for in this case the biases of the estimates to be compared are often of similar magnitude.

   For a full description of the various methods of estimation, and a discussion of when they are appropriate, reference must be made to the recognized text-books on sampling theory and methods.

2. Supplementary information: In many cases supplementary information, available from a complete census or other source, can be used to increase the accuracy of the estimates, often substantially. In the ratio method of estimation, for example, the sample is used to estimate a ratio which is then multiplied by a known total. The estimation of total production of an agricultural crop from yield per acre is an example of this. The direct estimates from the total yields of the sampled fields (i.e., the mean of all these totals multiplied by the total number of fields in the population) would be considerably less accurate.

3. Stratification after selection provides another example of the use of supplementary information. In some cases the number of sampling units in each stratum is known from census data, but stratification is not possible because the sampling units cannot be grouped by strata before selection. If a random sample is taken the numbers of units in the sample falling in the different strata will not be in the same proportions as in the population. This can be allowed for by differential weighting. This procedure will give nearly the same accuracy as stratification before selection with uniform sampling fraction.

4. Bias in observation: A tendency to errors in the same direction is termed a bias, or sometimes a systematic error. Specifically, the bias is the amount by which the mean value will be in error when a large number of observations are taken. In sampling work analogous biases due to non-response can arise when non-respondents as a class differ in some way from respondents. Similar biases due to omissions arise from incompleteness of the frame.

5. Bias of an estimate: An estimate is unbiased if the mean of its values for all possible samples selected from the population according to the specified sampling scheme is equal to the value for the population of the quantity estimated. Apart from bias due to bias in observation, etc. absence of bias in a mean or total is in general obtained by weighting the values of the variate obtained for each sampling unit inversely as the probability of selection of that unit. Thus, in an unrestricted random sample, or a stratified random sample with uniform sampling fraction, all sampling units are given equal weight. In a stratified sample with variable sampling fraction weights inversely proportional to the sampling fractions are assigned to the different strata. In sampling with probability proportional to size each unit is weighted inversely to its size. A further example is provided by the sampling of fields for yield of a crop, the area of the crop in a district being known. If the fields are selected with equal probability a virtually unbiased estimate of the total production of a district is obtained by giving equal weight to the total production of each field, which is equivalent to weighting the yield per acre (or hectare) of each field by the crop area of that field and multiplying this weighted mean by the total crop area: if the fields have been selected by point sampling the unweighted mean of the yields per acre is taken. (This estimate is in fact a ratio estimate, the required ratio being provided by the estimate of the total production of all sampled fields divided by the total area of all sampled fields: in point sampling the probability of selection is proportional to the size of the field).

   Biased estimates may, in certain circumstances, be considerably more accurate than unbiased estimates. In sampling for yield of a crop, for example, the main component of variation is from field to field and is often practically independent of size of field. While, therefore, weighted means of yields per acre may be required to provide unbiased estimates of total production comparisons of yields per acre between small geographic areas may be better made by unweighted means, as the gain in accuracy may outweigh the biases introduced, particularly as these biases are likely to be similar for different areas.

Trivial biases in estimation may be ignored. The ratio method of estimation, for example, is usually slightly biased, but in most practical cases this bias is of no consequence.

6. Actual errors: An estimate will usually differ from the quantity estimated. This difference, the actual error of the estimate, will generally be unknown.

7. Random sampling errors: If the observations (and the calculations) are without error, the actual errors in the estimates will be wholly due to the fact that a sample only of the population was observed. Apart from bias introduced in the estimation process, with a random sampling procedure positive and negative deviations in the individual observations will tend to balance one another. In a sample with an adequate number of randomly selected sampling units (or groups of sampling units) reliable estimates of the expected magnitude of the resultant errors in the estimates, which may be termed random sampling errors, may be made. The methods of error estimation appropriate to different types of sampling and methods of estimation of the population values are described in statistical text-books.

If the sampling is systematic, fully valid estimates of error cannot be made, though in certain cases, e.g., in a systematic sample from an alphabetically arranged list, the sample may be judged sufficiently nearly equivalent to some form of restricted or unrestricted random sample. Some forms of random sampling, e.g., stratified sampling with one unit from each stratum, cannot provide fully valid estimates of error, though in many cases reasonably satisfactory approximate estimates may be made.

8. Standard error: Estimates of variability and sampling error of an estimate are best expressed in terms of standard error. In survey work it is convenient to use "standard error" to refer to the error of an estimate and "standard deviation" to refer to the variability of a single observation (see item 10 below). The square of the standard error is termed the sampling variance of the estimate. In many cases, e.g., in an unrestricted random sample of units of approximately equal size, the actual errors of estimates of the population values in a series of samples will conform fairly closely to what is known as the normal distribution, and if the standard error is known the probability of an actual error of any magnitude can be determined by reference to a table of the normal distribution. Thus the actual error, positive or negative, will be less than the standard error in 68 per cent of all samples and less than twice the standard error in 95 per cent of all samples.

9. Limits of sampling error, margin of uncertainty, confidence limits: Instead of standard errors, limits of error or confidence limits at some assigned level of probability may be given. Thus limits of twice the standard error correspond to a probability level of 5 per cent (2 1/2 per cent in each direction). If such limits are used the level of probability and the factor by which the standard errors have been multiplied must be clearly stated, but the direct use of standard errors is usually preferable.

The contrasts between the estimates provided by the sub-samples of an interpenetrating network provide estimates of sampling errors which include any contributions due to differential bias between observers, computers, etc. If limits based on such sampling errors are calculated in the manner of confidence limits, then these limits specify the margin of uncertainty.

The symbol $\pm$ is used both in relation to standard errors and limits of error (and, in the past, for probable error); the actual use should be clearly indicated.

10. Standard deviation of a single observation: The standard deviation of a single observation is a measure of the variability of the sampled material when the sampling unit is the unit of observation. The square of the standard deviation is called the variance of a single observation. The relevant standard deviations for the estimation of errors of estimates of the population values depend on the type of sampling, e.g., in a stratified sample the standard deviations within strata (which may differ from stratum to stratum) are required. Information on the various standard deviations is also required for studies of efficiency and for planning future surveys, and they can therefore be usefully reported.

Clear distinction must be made between the standard deviation of a single observation and the standard error of an estimate, e.g., a mean.

The standard deviation of a single observation, expressed as a percentage of the mean of the observations, is sometimes termed the coefficient of variation (C.V.). This term is sometimes used to denote also the standard error of an estimate expressed as a percentage of the estimate. The phrases "percentage standard deviation of a single observation" and "percentage standard error of the estimate" are more explicit.

11. Non-random errors: Clear distinction must be made between components of error which are included in the estimates of random sampling error deducible from the survey results, and those which are not so included. Errors which are common to all investigators, and indeed any constant component of error (or bias) in the recorded information, (or in a stratified sample components of error which are constant for a particular stratum) or bias introduced by the estimation process, and also errors arising from incompleteness or inaccuracy of the frame, or lack of information for some of the units in the sample, will not be included in the estimates of random sampling error.

Not only are non-random errors (biases) non-estimable, except by special investigations designed for the purpose, but they will not for the most part be reduced by increasing the size of the sample, whereas random sampling errors can be reduced by any desired amount by sufficiently increasing the sample size, and can often also be reduced by improving the sampling design.

12. Accuracy, precision: The word "accuracy" is sometimes used to include both sampling errors and non-sampling errors and "precision" to indicate sampling errors alone. However, this use has not been standardized and the sense in which the words are used should therefore be made clear.

13. Permissible error: The error which, when planning a survey, it is considered should not be exceeded if the purposes of the survey are to be fulfilled. It is important, when a permissible error is given by those commissioning the survey, to ascertain whether standard error, or limits of error at some assigned probability level, are meant.

14. Fractile graphical method of assessment of error: A simple graphical representation can supply a visual (or geometric) means of assessing and controlling errors and also a measure of the margin of error (or of uncertainty) especially in the case of an interpenetrating network of independent sub-samples. The observed sampling units in each sub-sample are ranked in ascending order of any suitable variate (or even in order of time of observation) and the whole sub-sample is divided into a suitable number of fractile groups of equal size (that is, consisting of observed sampling units of equal number or of equal probability weight). For each sub-sample the average value, or median, or some other estimate of any variate for each fractile group is plotted on paper and the successive plotted points are connected by straight lines; these polygonal lines are called the fractile graphs of the sub-samples.

A fractile graph for the whole sample can be drawn in the same way, retaining the same number of fractile groups, as in the case of the sub-samples. The area between the fractile graphs for sub-samples provides a visual and geometric estimate of the error associated with the fractile graph of the whole sample. This method can be used not only to assess the significance of observed deviation from a statistical hypothesis relating to the results of the survey for any particular part or the whole range of observed data, but also for comparisons between results based on two or more surveys carried out in different regions or at different periods of time in the same region using the fractile graphical error for each survey. This method is being extensively used in India.7/

7/ Details are given by P. C. Mahalanobis in "A Method of Fractile Graphical Analysis." Econometrica, Vol. 28, No. 2 (1960), pp. 325-367; also published in Sankhyā, Series A. Vol. 23, Part 1 (February, 1961), pp. 41-64.

WHERE TO BUY UNITED NATIONS PUBLICATIONS
AND THE PUBLICATIONS OF THE INTERNATIONAL COURT OF JUSTICE

AFRICA

CAMEROON:
LIBRAIRIE DU PEUPLE AFRICAIN
La Gérante, B. P. 1197, Yaoundé.
DIFFUSION INTERNATIONALE CAMEROUNAISE
DU LIVRE ET DE LA PRESSE, Sangmelima.

CONGO (Léopoldville):
INSTITUT POLITIQUE CONGOLAIS
B. P. 2307, Léopoldville.

ETHIOPIA: INTERNATIONAL PRESS AGENCY
P. O. Box 120, Addis Ababa.

GHANA: UNIVERSITY BOOKSHOP
University College of Ghana, Legon, Accra.

KENYA: THE E.S.A. BOOKSHOP, Box 30167, Nairobi.

LIBYA: SUOKI EL JERBI (BOOKSELLERS)
P. O. Box 78, Istiklal Street, Benghazi.

MOROCCO: AUX BELLES IMAGES
281 Avenue Mohammed V, Rabat.

NIGERIA: UNIVERSITY BOOKSHOP (NIGERIA) LTD.
University College, Ibadan.

NORTHERN RHODESIA:
J. BELDING, P. O. Box 750, Mufulira.

NYASALAND: BOOKERS (NYASALAND) LTD.
Lontyre House, P. O. Box 34, Blantyre.

SOUTH AFRICA:
VAN SCHAIK'S BOOKSTORE (PTY) LTD.
Church Street, Box 724, Pretoria.
TECHNICAL BOOKS (PTY.) LTD., Faraday House
P. O. Box 2866, 40 St. George's Street, Cape Town.

SOUTHERN RHODESIA:
THE BOOK CENTRE, First Street, Salisbury.

TANGANYIKA: DAR ES SALAAM BOOKSHOP
P. O. Box 9030, Dar es Salaam.

UGANDA: UGANDA BOOKSHOP, P. O. Box 145, Kampala.

UNITED ARAB REPUBLIC:
LIBRAIRIE "LA RENAISSANCE D'ÉGYPTE"
9 Sh. Adly Pasha, Cairo.
AL NAHDA EL ARABIA BOOKSHOP
32 Abd-el-Khalek Sarwart St., Cairo.

ASIA

BURMA: CURATOR, GOVT. BOOK DEPOT, Rangoon.

CAMBODIA: ENTREPRISE KHMÈRE DE LIBRAIRIE
Imprimerie & Papeterie Sarl, Phnom-Penh.

CEYLON: LAKE HOUSE BOOKSHOP
Assoc. Newspapers of Ceylon, P. O. Box 244, Colombo.

CHINA:
THE WORLD BOOK COMPANY, LTD.
99 Chung King Road, 1st Section, Taipeh, Taiwan.
THE COMMERCIAL PRESS, LTD.
211 Honan Road, Shanghai.

HONG KONG: THE SWINDON BOOK COMPANY
25 Nathan Road, Kowloon.

INDIA:
ORIENT LONGMANS
Calcutta, Bombay, Madras, New Delhi, Hyderabad.
OXFORD BOOK & STATIONERY COMPANY
New Delhi and Calcutta.

INDONESIA:
PEMBANGUNAN, LTD., Gunung Sahari 84, Djakarta.

JAPAN: MARUZEN COMPANY, LTD.
6 Tori-Nichome, Nihonbashi, Tokyo.

KOREA, REPUBLIC OF:
EUL-YOO PUBLISHING CO., LTD., 5, 2-KA, Chongno, Seoul.

PAKISTAN:
THE PAKISTAN CO-OPERATIVE BOOK SOCIETY
Dacca, East Pakistan.
PUBLISHERS UNITED, LTD., Lahore.
THOMAS & THOMAS, Karachi.

PHILIPPINES:
PHILIPPINE EDUCATION COMPANY, INC.
1104 Castillejos, P. O. Box 620, Quiapo, Manila.
POPULAR BOOKSTORE, 1573 Doroteo Jose, Manila.

SINGAPORE:
THE CITY BOOK STORE, LTD., Collyer Quay.

THAILAND:
PRAMUAN MIT, LTD.
55 Chakrawat Road, Wat Tuk, Bangkok.
NIBONDH & CO., LTD.
New Road, Sikak Phya Sri, Bangkok.
SUKSAPAN PANIT
Mansion 9, Rajadamnern Avenue, Bangkok.

VIET-NAM, REPUBLIC OF:
LIBRAIRIE-PAPETERIE XUAN THU
185, rue Tu-do, B. P. 283, Saigon.

EUROPE

AUSTRIA:
GEROLD & COMPANY, Graben 31, Wien, I.
GEORG FROMME & CO., Spengergasse 39, Wien, V.

BELGIUM:
AGENCE ET MESSAGERIES DE LA PRESSE, S. A.
14-22, rue du Persil, Bruxelles.

BULGARIA:
RAZNOÏZNOS, 1, Tzar Assen, Sofia.

CYPRUS: PAN PUBLISHING HOUSE
10 Alexander the Great Street, Strovolos.

CZECHOSLOVAKIA:
ARTIA LTD., 30 ve Smečkách, Praha, 2.

DENMARK: EJNAR MUNKSGAARD, LTD.
Nørregade 6, København, K.

FINLAND: AKATEEMINEN KIRJAKAUPPA
2 Keskuskatu, Helsinki.

FRANCE: ÉDITIONS A. PEDONE
13, rue Soufflot, Paris (V^e).

GERMANY, FEDERAL REPUBLIC OF:
R. EISENSCHMIDT
Schwanthaler Str. 59, Frankfurt/Main.
ELWERT UND MEURER
Hauptstrasse 101, Berlin-Schöneberg.
ALEXANDER HORN, Spiegelgasse 9, Wiesbaden.
W. E. SAARBACH, Gertrudenstrasse 30, Köln (1).

GREECE: KAUFFMANN BOOKSHOP
28 Stadion Street, Athens.

HUNGARY: KULTURA, P. O. Box 149, Budapest 62.

ICELAND: BÓKAVERZLUN SIGFÚSAR
EYMUNDSSONAR H. F., Austurstraeti 18, Reykjavik.

IRELAND: STATIONERY OFFICE, Dublin.

ITALY: LIBRERIA COMMISSIONARIA SANSONI
Via Gino Capponi 26, Firenze,
and Via Paolo Mercuri 19/B, Roma.
AGENZIA E.I.O.U., Via Meravigli 16, Milano.

LUXEMBOURG:
LIBRAIRIE J. TRAUSCHSCHUMMER
Place du Théâtre, Luxembourg.

NETHERLANDS: N. V. MARTINUS NIJHOFF
Lange Voorhout 9, 's-Gravenhage.

NORWAY: JOHAN GRUNDT TANUM
Karl Johansgate, 41, Oslo.

POLAND: PAN, Palac Kultury i Nauki, Warszawa.

PORTUGAL: LIVRARIA RODRIGUES Y CIA.
186 Rua Aurea, Lisboa.

ROMANIA: CARTIMEX, Str. Aristide Briand 14-18,
P. O. Box 134-135, Bucuresti.

SPAIN: AGUILAR S.A. DE EDICIONES
Juan Bravo 38, Madrid 6.
LIBRERIA BOSCH
Ronda Universidad 11, Barcelona.
LIBRERIA MUNDI-PRENSA, Castelló 37, Madrid.

SWEDEN:
C. E. FRITZE'S KUNGL. HOVBOKHANDEL A-B
Fredsgatan 2, Stockholm.

SWITZERLAND:
LIBRAIRIE PAYOT, S. A., Lausanne, Genève.
HANS RAUNHARDT, Kirchgasse 17, Zürich 1.

TURKEY: LIBRAIRIE HACHETTE
469 Istiklal Caddesi, Beyoglu, Istanbul.

UNION OF SOVIET SOCIALIST REPUBLICS:
MEZHDUNARODNAYA KNYIGA
Smolenskaya Ploshchad, Moskva.

UNITED KINGDOM:
H. M. STATIONERY OFFICE, P. O. Box 569, London, S.E. 1
(and HMSO branches in Belfast, Birmingham,
Bristol, Cardiff, Edinburgh, Manchester).

YUGOSLAVIA:
CANKARJEVA ZALOŽBA, Ljubljana, Slovenia.
DRŽAVNO PREDUZEĆE
Jugoslovenska Knjiga, Terazije 27/11, Beograd.
PROSVJETA, 5, Trg Bratstva i Jedinstva, Zagreb.
PROSVETA PUBLISHING HOUSE
Import-Export Division, P. O. Box 559,
Terazije 16/1, Beograd.

LATIN AMERICA

ARGENTINA: EDITORIAL SUDAMERICANA, S. A.
Alsina 500, Buenos Aires.

BOLIVIA: LIBRERIA SELECCIONES, Casilla 972, La Paz.
LOS AMIGOS DEL LIBRO
Calle Perú esq. España, Casilla 450, Cochabamba.

BRAZIL:
LIVRARIA AGIR
Rua Mexico 98-B, Caixa Postal 3291, Rio de Janeiro.
LIVRARIA FREITAS BASTOS, S. A.
Caixa Postal 899, Rio de Janeiro.
LIVRARIA KOSMOS EDITORA
Rua Rosario 135/137, Rio de Janeiro.

CHILE:
EDITORIAL DEL PACIFICO, Ahumada 57, Santiago.
LIBRERIA IVENS Casilla 206, Santiago.

COLOMBIA:
LIBRERIA AMERICA, Calle 51 Núm. 49-58, Medellin.
LIBRERIA BUCHHOLZ, Av. Jiménez de Quesada 8-40, Bogotá.

COSTA RICA: IMPRENTA Y LIBRERIA TREJOS
Apartado 1313, San José.

CUBA: CUBARTIMPEX, Apartado Postal 6540, La Habana.

DOMINICAN REPUBLIC: LIBRERIA DOMINICANA
Mercedes 49, Santo Domingo.

ECUADOR:
LIBRERIA CIENTIFICA, Casilla 362, Guayaquil.
LIBRERIA UNIVERSITARIA, Calle García Moreno 739, Quito.

EL SALVADOR:
LIBRERIA CULTURAL SALVADOREÑA
2a. Av. Sur, San Salvador.
MANUEL NAVAS Y CIA.
1a. Avenida Sur 37, San Salvador.

GUATEMALA:
LIBRERIA CERVANTES, 5a. Av. 9 39, Zona 1, Guatemala.
SOCIEDAD ECONOMICA-FINANCIERA
6a. Av. 14-33, Guatemala.

HAITI: LIBRAIRIE "À LA CARAVELLE", Port-au-Prince.

HONDURAS: LIBRERIA PANAMERICANA, Tegucigalpa.

MEXICO: EDITORIAL HERMES, S. A.
Ignacio Mariscal 41, México, D. F.

PANAMA: JOSE MENENDEZ
Agencia Internacional de Publicaciones,
Apartado 2052, Av. 8A Sur 21-58, Panamá.

PARAGUAY:
AGENCIA DE LIBRERIAS DE SALVADOR NIZZA
Calle Pte. Franco No. 39-43, Asunción.

PERU:
LIBRERIA INTERNACIONAL DEL PERU, S. A.
Casilla 1417, Lima.
LIBRERIA STUDIUM S. A.
Amargura 939, Apartado 2139, Lima.

URUGUAY: LIBRERIA RAFAEL BARRETT
Ramón Anador 4030, Montevideo.
REPRESENTACION DE EDITORIALES,
PROF. H. D'ELIA
Plaza Cagancha 1342, 1° piso, Montevideo.

VENEZUELA: LIBRERIA DEL ESTE
Av. Miranda, No. 52, Edf. Galipán, Caracas.

MIDDLE EAST

IRAN: MEHR AYIN BOOKSHOP
Abbas Abad Avenue, Isfahan.

IRAQ: MACKENZIE'S BOOKSHOP, Baghdad.

ISRAEL: BLUMSTEIN'S BOOKSTORES
35 Allenby Rd. and 48 Nachlat Benjamin St., Tel Aviv.

JORDAN: JOSEPH I. BAHOUS & CO.
Dar-ul-Kutub, Box 66, Amman.

LEBANON: KHAYAT'S COLLEGE BOOK COOPERATIVE
92-94, rue Bliss, Beirut.

NORTH AMERICA

CANADA:
THE QUEEN'S PRINTER/L'IMPRIMEUR DE LA REINE
Ottawa, Ontario.

UNITED STATES OF AMERICA:
SALES SECTION, UNITED NATIONS, New York.

Puerto Rico:
PAN AMERICAN BOOK CO., P. O. Box 3511, San Juan 17.
BOOKSTORE, UNIVERSITY OF PUERTO RICO, Rio Piedras.

OCEANIA

AUSTRALIA:
U. N. ASSOCIATION OF AUSTRALIA
McEwan House, 343 Little Collins St., Melbourne C.1, Vic.
WEA BOOKROOM, University, Adelaide, S.A.
UNIVERSITY BOOKSHOP, St. Lucia, Brisbane, Qld.
THE EDUCATIONAL AND TECHNICAL BOOK AGENCY
Parap Shopping Centre, Darwin, N.T.
COLLINS BOOK DEPOT PTY. LTD.
Monash University, Wellington Road, Clayton, Vic.
COLLINS BOOK DEPOT PTY. LTD.
363 Swanston Street, Melbourne, Vic.
THE UNIVERSITY BOOKSHOP, Nedlands, W.A.
UNIVERSITY BOOKROOM
University of Melbourne, Parkville N.2, Vic.
UNIVERSITY CO-OPERATIVE BOOKSHOP LIMITED
Manning Road, University of Sydney, N.S.W.

NEW ZEALAND:
GOVERNMENT PRINTING OFFICE
Private Bag, Wellington (and Government Bookshops in
Auckland, Christchurch and Dunedin).

WEST INDIES

BERMUDA: BERMUDA BOOK STORES
Reid and Burnaby Streets, Hamilton.

BRITISH GUIANA: BOOKERS STORES, LTD.
20-23 Church Street, Georgetown.

CURAÇAO, N.W.I.:
BOEKHANDEL SALAS, P. O. Box 44.

JAMAICA: SANGSTERS BOOK ROOM
91 Harbour Street, Kingston.

TRINIDAD AND TOBAGO:
CAMPBELL BOOKER LTD., Port of Spain.

[64E1]

United Nations publications may be purchased or ordered from booksellers throughout the world and paid for in local currency. For further information write to United Nations, Sales Section, New York, N. Y. 10017, or to United Nations, Sales Section, Palais des Nations, Geneva, Switzerland.

STATISTICAL OFFICE OF THE UNITED NATIONS

SELECTED LIST OF PUBLICATIONS JUNE 1964

STATISTICAL YEARBOOK 1963
(E/F) 714 pages. Paperbound $9.00; clothbound $11.50, Sales No. 64.XVII.1.

DEMOGRAPHIC YEARBOOK 1962
(E/F) 700 pages. Paperbound $8.00; clothbound $10.00, Sales No. 63.XIII.1.

YEARBOOK OF NATIONAL ACCOUNTS STATISTICS 1962
(E/F) 319 pages, $4.00, Sales No. 63.XVII.2.

YEARBOOK OF INTERNATIONAL TRADE STATISTICS 1961
(E) Sales No. 62.XVII.8, 710 pages, $8.00.

MONTHLY BULLETIN OF STATISTICS
(E/F) Annual Subscription: $10.00; $1.00 per copy. Subscribers receive the 1959 Supplement (E).

CURRENT ECONOMIC INDICATORS
(E/F) Annual Subscription: $4.00; $1.00 per copy, issued quarterly.

STATISTICAL PAPERS

Series A:
POPULATION AND VITAL STATISTICS REPORTS
(quarterly) (E) $0.30 per copy. Annual subscription $1.00

Series B:*
STATISTICAL NOTES (current events in international statistics) (periodically), (E)

Series C, No. 10:
SAMPLE SURVEYS OF CURRENT INTEREST, Ninth Report
(E, F, S) 33 pages, $0.50, Sales No. 62.XVII.3.

Series D:
COMMODITY TRADE STATISTICS
(E) Annual Subscription: $10.00.

Series J, No. 7:
WORLD ENERGY SUPPLIES, 1959-62
(E) 115 pages, $1.50, Sales No. 64.XVII.6.

Series K, No. 2:
COMPENDIUM OF SOCIAL STATISTICS
(E/F) 586 pages, $7.00, Sales No. 63.XVII.3.

Series M, No. 4, Rev. 1:
INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION OF ALL ECONOMIC ACTIVITIES
(E, F, S, R) 27 pages, $0.50, Sales No. 58.XVII.7.

Series M, No. 4, Rev. 1/Add. 1: (Indexed Ed.)
INDEXES TO THE INTERNATIONAL STANDARD INDUSTRIAL CLASSIFICATION OF ALL ECONOMIC ACTIVITIES
(E) 359 pages, $3.50, Sales No. 59.XVII.9.

Series M, No. 8:*
INTERNATIONAL STANDARD DEFINITIONS FOR TRANSPORT STATISTICS (1950) (E,F,S,)

Series M, No. 14/Add. 3:
RETAIL PRICE COMPARISONS FOR INTERNATIONAL SALARY DETERMINATION
(E, F, S) 51 pages, $0.75, Sales No. 62.XVII.4.

Series M, No. 17, Rev. 1:
INTERNATIONAL RECOMMENDATIONS IN BASIC INDUSTRIAL STATISTICS
(E, F, S) 57 pages, $0.75, Sales No. 60.XVII.8.

Series M, No. 17, Rev. 1/Add. 1:
INTERNATIONAL RECOMMENDATIONS ON THE 1963 WORLD PROGRAMME OF BASIC INDUSTRIAL STATISTICS
(E, F, S) 43 pages, $0.75, Sales No. 60.XVII.8, Add. 1.

Series M, No. 19:
PRINCIPLES FOR A VITAL STATISTICS SYSTEM
(E, F, S) 28 pages, $0.30, Sales No. 53.XVII.8.

Series M, No. 21:
WORLD WEIGHTS AND MEASURES
(E) (Preliminary issue), 225 pages, $2.00, Sales No. 55.XVII.2.

Series M, No. 22, Rev. 1:
DIRECTORY OF INTERNATIONAL STANDARDS FOR STATISTICS
(E, F, S, R) 39 pages, $0.75, Sales No. 60.XVII.5.

Series M, No. 26:
INTERNATIONAL RECOMMENDATIONS IN STATISTICS OF DISTRIBUTION
(E, F, S) 16 pages, $0.20, Sales No. 58.XVII.4.

Series M, No. 27:
PRINCIPLES AND RECOMMENDATIONS FOR NATIONAL POPULATION CENSUSES
(E, F, S) 29 pages, $0.30, Sales No. 58.XVII.5.

Series M, No. 28:
GENERAL PRINCIPLES FOR A HOUSING CENSUS
(E, F, S) 14 pages, $0.20, Sales No. 58.XVII.8.

Series M, No. 29:
METHODS USED IN COMPILING THE UNITED NATIONS PRICE INDEXES FOR BASIC COMMODITIES IN INTERNATIONAL TRADE
(E, F, S) 66 pages, $0.60, Sales No. 58.XVII.9.

Series M, No. 30:
CUSTOMS AREAS OF THE WORLD
(E, F, S) 28 pages, $0.40, Sales No. 59.XVII.5.

Series M, No. 31:
STATISTICAL SERIES FOR THE USE OF LESS DEVELOPED COUNTRIES IN PROGRAMMES OF ECONOMIC AND SOCIAL DEVELOPMENT
(E F, S) 56 pages, $0.50, Sales No. 59.XVII.10.

Series M, No. 34:
STANDARD INTERNATIONAL TRADE CLASSIFICATION, REVISED
(E, F, S) 135 pages, $2.00, Sales No. 61.XVII.6.

Series M, No. 36, Rev. 1:
BIBLIOGRAPHY OF INDUSTRIAL AND DISTRIBUTIVE-TRADE STATISTICS
(E) 64 pages, $0.75, Sales No. 63.XVII.7.

Series M, No. 37:
STATISTICAL INDICATORS OF HOUSING CONDITIONS
(E, F, S) 21 pages, $0.35, Sales No. 62.XVII.7.

Series M, No. 38:
COMMODITY INDEXES FOR THE STANDARD INTERNATIONAL TRADE CLASSIFICATION, REVISED
(E) (2 volumes) Vol. 1, 463 pages, Sales No. 64.XVII.2, $5.00; Vol. II, 349 pages, Sales No. 64.XVII.3 $4.50.

Series P, No. 2:
GROWTH OF WORLD INDUSTRY, 1938-1961; NATIONAL TABLES
(E/F) 849 pages, $10.00, clothbound, Sales No. 63.XVII.5.

STUDIES IN METHODS

Series F, No. 1:
INDEX NUMBERS OF INDUSTRIAL PRODUCTION, 1950
(E, F, S) 60 pages, $0.25, Sales No. 50.XVII.4.

Series F, No. 2, Rev. 2:
A SYSTEM OF NATIONAL ACCOUNTS AND SUPPORTING TABLES
(E) 45 pages, $0.75, Sales No. 64.XVII.5.

Series F, No. 3:
CONCEPTS AND DEFINITIONS OF CAPITAL FORMATION
(E, F, S) 20 pages, $0.25, Sales No. 53.XVII.6.

Series F, No. 4:
INDUSTRIAL CENSUSES AND RELATED ENQUIRIES
(2 volumes). Vol. 1, (E, S) 384 pages; Vol. II, (E) 359 pages, $9.00 complete; single volume, $4.50. Sales No. 53.XVII.11.

Series F, No. 5, Rev. 1:
HANDBOOK OF POPULATION CENSUS METHODS
(3 volumes) (E, F, S) Vol. 1, 164 pages, $1.75; Vol. II, 79 pages, $0.80, Vol. III, 78 pages, $0.80, Sales No. 58.XVII.6.

Series F, No. 6:
HANDBOOK OF STATISTICAL ORGANIZATION
(E, S) 138 pages, $1.50, Sales No. 54.XVII.7.

Series F, No. 7:
HANDBOOK OF VITAL STATISTICS METHODS
(E, F, S) 258 pages, $3.00, Sales No. 55.XVII.1.

Series F, No. 8:
METHODS OF NATIONAL INCOME ESTIMATION
(E, F, S) 58 pages, $0.50, Sales No. 55.XVII.5.

Series F, No. 9:
A SHORT MANUAL ON SAMPLING
(E, F, S) Vol. I, 214 pages, $3.00, Sales No. 61.XVII.3.

*A limited number of these publications are available without charge from the Statistical Office, United Nations, New York, U.S.A.

All United Nations sales publications may be purchased in local currencies. Please see list of U.N. sales agents on inside back cover.

E=English; F=French; S=Spanish; R=Russian; E/F=bilingual text.

Litho in U.N.
08540—June 1964—2,825
ST/STAT/SER.C/1/Rev.2

Price: $U.S. 0.35
(or equivalent in other currencies)

United Nations publication
Sales No.: 64.XVII.7