# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. Agenda</td>
<td>3</td>
</tr>
<tr>
<td>II. Standard Terminology</td>
<td>4</td>
</tr>
<tr>
<td>III. Application of Sampling Methods for the Measurement of National</td>
<td>5</td>
</tr>
<tr>
<td>Income and the Construction of Social Accounts</td>
<td></td>
</tr>
<tr>
<td>IV. Methods of Crop Estimating and Forecasting</td>
<td>10</td>
</tr>
<tr>
<td>V. Education and Training in Sampling</td>
<td>17</td>
</tr>
<tr>
<td>VI. Sampling Methods in Relation to Population Censuses</td>
<td>18</td>
</tr>
<tr>
<td>VII. Sampling Methods in Tabulation</td>
<td>23</td>
</tr>
<tr>
<td>VIII. Sampling in Manpower Statistics</td>
<td>28</td>
</tr>
<tr>
<td>IX. Sample Surveys of Current Interest</td>
<td>30</td>
</tr>
<tr>
<td>X. Methods to Determine Volume of Postal Traffic</td>
<td>31</td>
</tr>
<tr>
<td>XI. Other Subjects Discussed</td>
<td>33</td>
</tr>
<tr>
<td>A. Comments on Quota Sampling</td>
<td>33</td>
</tr>
<tr>
<td>B. Future Work Programme</td>
<td>35</td>
</tr>
</tbody>
</table>
INTRODUCTION

1. The Sub-Commission on Statistical Sampling held its third session in Geneva from 12 September to 23 September 1949. The Sub-Commission re-elected Mr. P. C. Mahalanobis as Chairman.

2. The following members were present:
   
   Mr. P. C. Mahalanobis
   Mr. G. Darmois
   Mr. W. E. Deming
   Mr. F. Yates

   Consultant: Mr. R. A. Fisher

   The representatives for the specialized agencies were:

   International Labour Organisation: Mr. R. M. Woodbury
   Food and Agriculture Organization: Mr. C. Tauber
   World Health Organization: Mr. J. Chombart de Lauwere
   International Refugee Organization: Miss H. Jeter
   Universal Postal Union: Mr. F. Radice
   Mr. F. Deprez

   Secretariat: Mr. W. J. Bruce (Secretary to the Sub-Commission)
   Mr. W. R. Leonard (representing the Assistant Secretary-General in charge of Economic Affairs)

3. The third session of the Sub-Commission was attended by the following statisticians who contributed substantially to the discussions:

   Mr. P. A. Thionet, National Institute of Statistics, France;
   Mr. A. B. Reisz, Government Statistician, Gold Coast;
   Mr. B. Barberi and Mr. F. Brambilla of the Central Institute of Statistics, Italy;
   Mr. J. R. H. Shaul of the Central African Statistical Office, Southern Rhodesia;
   Mr. Serifik Billur, Central Statistical Office, Turkey;
   Mr. B. Debevec, Mr. M. Macura, Mr. I. Bcimovic and Mr. J. Korencic of the Central Bureau of Statistics, Yugoslavia;
   Mr. M. H. Guenouille, University of Aberdeen;
   Mr. J. R. N. Stone, Mr. J. E. G. Utting, and Mr. J. Durbin of Department of Applied Economics, University of Cambridge;
   Mr. S. N. Roy, University of Calcutta;
   Mr. A. Linder, University of Geneva;
CHAPTER I

AGENDA

4. The Agenda for the third session of the Sub-Commission was as follows:
   I. Adoption of Agenda.
   II. Election of officers.
   III. Use of sampling in preparing national income estimates.
   IV. Development of standard terminology in sampling.
   V. Report on sampling methods in crop estimating and forecasting.
   VI. Aspects of quota and probability samples.
   VII. Education and training in sampling methods; preparation of syllabus.
   VIII. Bibliography in the field of statistical sampling.
   IX. Sampling methods in conjunction with complete enumerations.
   X. Sampling methods in tabulation.
   XI. Sampling in manpower statistics.
   XII. Sampling surveys of current interest, second report.
   XIII. Other business.
CHAPTER II

STANDARD TERMINOLOGY

5. The Sub-Commission considered comments (E/CN.3/Sub.1/W.6) which had been received from interested statisticians in reference to Recommendations Concerning the Preparation of Reports of Sampling Surveys, a memorandum prepared by the Sub-Commission at its second session and circulated by the Statistical Office to persons responsible for the conduct of sample surveys. (The Preparation of Sampling Survey Reports, Statistical Papers Series C. No. 1). The Sub-Commission was pleased to note that its recommendations on the preparation of reports and on terminology had been used in connexion with sample surveys carried out in France and Southern Rhodesia, and in at least one new text-book. On the basis of the written comments received, and the experience gained by applying the recommendations to specific surveys, the Sub-Commission re-examined the recommendations and recorded a number of changes intended to amplify and clarify some of the sections. In view of the nature of the proposed changes, it was not considered necessary to re-issue the recommendations, or to amend them by a supplementary statement, at this time. It was felt that the amended language could be conveniently introduced in a revised edition of the recommendations at a suitable time, taking account of the present stocks. The changes agreed to are given in Appendix B of this report.
CHAPTER III

APPLICATION OF SAMPLING METHODS FOR THE MEASUREMENT OF NATIONAL INCOME AND THE CONSTRUCTION OF SOCIAL ACCOUNTS

6. At its second session the Sub-Commission gave consideration to the preparation of estimates of national income by three alternative methods and made a number of suggestions for the application of sampling methods to the development and improvement of such statistics (E/CN.3/52, Chapter VI).

7. At its third session the Sub-Commission took note of the development of the technique of social accounting\(^1\) as a means of presenting the chief components of national income and related totals in a self-consistent framework, and gave special attention to the more general problem of the place of sampling in the estimation of the items in the social accounts.

8. The Sub-Commission had before it a report on the detailed definitions and concepts for the calculation of agricultural income, a note on plans in one country to conduct a sample survey of income payments and profits in manufacturing (E/CN.3/Sub.1/16), and a paper by Messrs. J. R. N. Stone, J. E. C. Utting and J. Durbin of the Department of Applied Economics, University of Cambridge, discussing proposals to apply sampling methods to the collection of data for the construction of social accounts.\(^2\)

9. The preparation of national income statistics, and the social accounting technique, as it has developed out of national income analysis require a very large amount of economic statistics. The accounts summarize all transactions of the economy so as to present in terms of an integrated system of concepts, those economic variables which are considered essential

---

\(^1\) A social accounting system consists of a set of balancing accounting statements for each of the sectors into which it is found desirable to divide the transacting entities of an economy (for instance, business enterprises, family households, government and the "rest of the world"). Transactions between the accounts are grouped into a not too large number of aggregates each representing a significant economic variable. For example, wages and salaries will appear on the payment side of the operating account of all enterprises, and similarly they appear on the receipts side of the household account. Cf. Measurement of National Income and the Construction of Social Accounts, Studies and Reports on Statistical Methods, No. 7, League of Nations. (Published by United Nations, 1947).

to an understanding of the operation and the structure of the economy. The
completion of an accounting statement of this type involves two steps:
first, the elaboration of necessary concepts and systems of accounts by
economists and, second, the obtaining of statistics required to make
estimates which will fulfill the requirements of the definitions.
10. Bearing in mind that the accounts summarize all economic transactions
including those which may not enter the price system - the task of the
statistician is a difficult one, the more so in the case of countries which
have not as yet developed the elements of a comprehensive statistical system.
Moreover, it is frequently true that the contents of many existing
statistical series do not conform to the definitions established by national
income economists. Official statistics are either collected for specific
administrative purposes, or are by-products of governmental processes, or
they are of a more general type - such as those resulting from censuses;
in these cases the statistics are not likely to correspond precisely with
national income categories. Thus even where statistical information
of the usual kind is available it is frequently impossible to specify the
margin of error of the estimates based on it.
11. From the statistical standpoint, therefore, the calculation of
national income requires an extensive body of statistics covering all
economic activities, and adjusted to be consistent with the analytical
categories. Income estimates may also require, and this is perhaps
especially true as regards social accounts, additional series obtained by
direct collection from the accounting entities concerned to provide data
not normally supplied even by highly developed systems of official statistics.
12. Statistical material for national income studies, as for all other
purposes, can obviously be obtained either by complete enumeration or by the
application of sampling methods. The Sub-Commission has previously
considered questions of sampling in relation to several broad subjects:
censuses of agriculture, censuses of population, and family budget studies
(E/CN.3/37 and E/CN.3/52). It is not necessary, therefore, to repeat here
the conditions under which statistical sampling can be successfully employed
in such fields. It should be noted, however, that if family budget enquiries
are to be most useful for estimating consumers' expenditures and savings for
the purpose of national income estimates and social accounts, it will be
necessary to design the sample so that it represents all groups of the
population.
13. To be of value for policy-making purposes, social accounting information
is required at frequent intervals, possibly annually, and it is important
censuses of industrial activity, upon which some of the estimates are based, are taken only at biennial or quinquennial intervals. Under these circumstances it would be possible by means of sampling enquiries to obtain the major items required for the intermediate years. The samples should be kept as small as possible, taking into account the accuracy required.

14. In addition, where some of the items are of too complex a nature to be collected with ease in a complete census, the use of sampling, or in the case of a sample census, multiphase and multi-stage sampling, will enable more attention to be paid to the eliciting of this special information from smaller numbers of units in the sample.

15. It is obvious in general that the application of sampling methods is particularly suited to situations where only a broad total for an item (rather than a detailed sub-division) or an overall adjustment factor is sought. For example, total figures of the volume of individual savings or miscellaneous operating costs of business enterprises are types of statistics essential for income computations which may be obtained by sampling. It would not normally be necessary for any purposes to attempt to obtain such figures from all individuals or business enterprises.

16. Sampling methods may be particularly suitable in less developed countries to obtain information on the production and consumption of population groups living under conditions of a subsistence economy. Without entering into a consideration of the conceptual problems involved if the data are to be used for the purpose of measuring national income, the Sub-Commission wishes to draw attention to the application of sampling methods in this field.

17. Sampling methods may likewise be useful to obtain data for handicrafts and small scale industries usually omitted from or not adequately covered by censuses of production.

18. In countries with large numbers of local governmental units for which statistics are not readily available, sampling methods may be an appropriate procedure for obtaining information on their economic transactions.

19. A subject related to national income research and the construction of social accounting systems is the study of the distribution of income by size of individual incomes or of family incomes. Although fully aware of the conceptual and methodological difficulties involved, the Sub-Commission wishes to draw attention to the use that has been made of sample surveys to collect data on income distributions by size.
20. A special subject in this field is the estimation of the distribution of incomes below tax exemption limits. In those countries where national income estimates are derived also on the basis of taxation statistics, sampling methods have sometimes been used to estimate total income below the tax exemption limit.

21. In connexion with the estimation of agricultural income, the Sub-Commission noted that methods of measurement necessarily differ from country to country. As indicated in the report (E/CN.3/Sub.1/16) in countries with highly developed agricultural statistics, data regularly obtained, for example, in connexion with government regulations concerning agricultural output and prices, may have to be supplemented by information on various other items in order to evaluate agricultural income. Such additional information might be obtainable through sample surveys.

22. In less developed countries agricultural statistics are usually available for a small number of the main crops only. Because of high costs of surveys to obtain complete crop data as well as the limited resources in available personnel with sufficient statistical training, it is not practicable to carry out complete statistical enquiries. Under such circumstances sampling procedures may be the appropriate device for obtaining data such as the following:

   (a) the yield per unit of area of various crops, or for other units if appropriate; and data needed to make crop forecasts;
   
   (b) the number of livestock of various categories;
   
   (c) the average output of milk per cow, average output of eggs per hen, etc;
   
   (d) the average net weight of livestock delivered for slaughtering and quantities of livestock by-products;
   
   (e) the in-put of fertilizers per unit of area, use of fodder, seeds, water for irrigation purposes etc;
   
   (f) agricultural labour force and its utilization;
   
   (g) data on living conditions of the rural population.

23. The Sub-Commission also noted the work of the Statistical Office in preparing an operating manual and instructions for the calculation of national income, including the construction of social accounts. The manual will include intermediate and final table forms covering each step of the process from the basic series to the final estimate. In the opinion of the Sub-Commission it should be possible, on the basis of the manual, for different countries to compare their existing series with those necessary for national income estimates and related aggregates and on making this
recommends that the Statistical Office, in circulating the manual, draw attention to the usefulness of sampling methods in calculating the supporting details of national income estimates.

24. The experience of a larger number of countries in the application of statistical sampling to problems of income estimates could usefully be accumulated; the Sub-Commission therefore requests that the use of sampling procedures particularly designed for the purpose of national income estimates be reported to the Statistical Office from time to time so that the experience gained can be communicated to other countries faced with similar problems. Reports of surveys and experimental work in this field could be drawn especially to the attention of the Sub-Commission in subsequent documents concerning sample surveys of current interest.
CHAPTER IV

METHODS OF CROP ESTIMATING AND FORECASTING

25. The Sub-Commission had before them three papers submitted by the Food and Agriculture Organization, covering crop estimation and forecasting methods in the United States, India and Japan, and a further paper prepared by the FAO on the estimation and forecasting of rice production in South East Asia (E/CON.3/Sub.1/14; 17; 17/Add.1). These papers give a conspectus of the methods followed in countries which differ very markedly in their degree of development and in the methods of crop estimation and forecasting which are in use. While not wishing to comment on the papers in detail the Sub-Commission feels that it would be of value to make some general observations on the subject, especially in view of the approaching World Census of Agriculture around 1950.

Definitions

26. The estimation of total production of a crop may be conveniently divided into the estimation of the area planted or sown (or harvested) and the estimation of the yield per unit area.

27. The term forecast should be confined to estimates of the yield per unit area made before the crop has reached maturity, or in the case of area, to estimates made before the crop has been sown or planted. If first estimates are issued at harvest and are later revised these may be termed current and revised estimates respectively, the final revision being termed the final estimate.

Estimates of area

28. Estimates of area may be made on the basis of complete enumeration or by sampling. Complete enumeration will in general involve collection of information from each farmer individually. In the less developed areas such information will have to be collected by local officials, covering small localities such as villages. The recruitment and training of such officials, and their subsequent supervision, presents serious difficulties. Frequently, also, there are strong inducements for not making accurate returns; for example, such returns may form the basis of local taxes, or involve the contribution of a part of the crop to the central government.

29. The Sub-Commission are strongly of the opinion that the difficulties of obtaining accurate returns on crop areas by local officials are very commonly seriously under-estimated. Two examples showing uncritical acceptance of such returns may be cited. Thus in document E/CON.3/Sub.1/17, Appendix B, it is stated that "As all the fields have been surveyed and
mapped the total acreage under each of the crops for a given region is accurately known." Although the report makes passing comments on possible defects, it is clear that the statistics collected by this method have been regarded as acceptable without further checks. Similarly in document E/CN.3/Sub.1/14 there are numerous statements to the effect that returns on acreage are to be regarded as completely reliable, e.g. "The method used for collecting statistics of area .... seems to be entirely adequate and as reliable as any method could be".

30. A number of objective sample checks by independent authorities have recently been made on area statistics collected by local officials or by returns made directly by farmers. Thus the survey of jute in Bengal in 1944-45 showed an under-estimate of area by complete enumeration of the order of 20 per cent, and a survey of Japan in 1948 (E/CN.3/Sub.1/17, Appendix C) showed under-estimation of area ranging from 9 per cent for paddy rice to 26 per cent for sweet potatoes.

31. The Sub-Commission therefore wishes to state as its opinion that acreage returns obtained either directly from farmers or collected by local officials or investigators appointed for the purpose should not be accepted without some system of independent checks operated by a separate agency such as the Central Statistical Office of the country concerned. The frequency and intensity of the checks must depend on local circumstances but the checks to be satisfactory must be based on some form of sampling procedure. Sampling is also of use in determining the percentage of the area sown that fails before harvest, owing to drought and other causes.

32. Whether complete enumeration controlled by sampling, or sampling without any attempt at complete enumeration, is the best method of obtaining area estimates, depends on whether detailed area figures for individual farmers are required for administrative purposes. If they are not, there appears to be little argument for instituting any form of complete enumeration. The Sub-Commission is of the opinion that the merits of complete enumeration are greatly exaggerated in many countries. Thus in E/CN.3/Sub.1/17, Appendix B, the use of random sample surveys is dismissed without discussion. Instead the setting up of an extensive agency for the collection of area statistics by complete enumeration is recommended. It is further stated that the construction of cadastral maps was a necessary preliminary where they do not exist and which must be undertaken notwithstanding the cost, time, and other difficulties. While the Sub-Commission in no way questions the value of accurate cadastral or topographic maps for many purposes it is of the opinion that other methods are to be preferred.
cannot be obtained if they are not available, as completely unjustified. 33. The Sub-Commission does not wish to discuss the various methods of sampling that can be used for the purpose of obtaining area estimates. These have been fully discussed elsewhere. It would, however, emphasize that the methods which it is best to employ will depend very much on local conditions and that in any case pilot surveys should first be undertaken to test out whatever methods are proposed. 34. It should also be emphasized that a permanent organization collecting annual information on all the more important crops is likely to be much more effective than an ad hoc organization built up to deal with a particular crop in a single year.

Yield per unit area

35. The estimation of the average yield per unit area, and hence of the total production, is a much more difficult problem than the estimation of the area. Current estimates can either be made by means of reports (by crop reporters, village headmen etc.) of what in the opinion of the reporter is the average yield of their district, or they may be based on sample harvestings (usually of small areas). The method of crop reporting must always be subjective, and there is no way of assessing its reliability other than by the use of crop-cutting or (in instances in which the whole of the crop is marketed) by reference to market returns. 36. Even in fully developed countries with long established systems of crop reporting doubts are frequently cast on the accuracy of the estimates provided, particularly in times of stress when accurate estimates are of greatest importance. In 1948 a check on the yields of rye instituted in Western Germany by sample harvesting (with proper allowance for loss of harvesting, variation in moisture content, and loss in threshing) showed discrepancies ranging from 8 per cent to 20 per cent underestimation in the various Länder. In less developed countries, and particularly if taxation depends on the estimated yields, reporters estimates are likely to be even less reliable. 37. The Sub-Commission therefore suggests that in under-developed areas generally, and in other areas where there is reason to doubt the existing crop estimates, the proper course is to set up a rigorous sampling procedure for estimating the yields by harvesting sample areas immediately prior to harvest. In this case also the exact methods will depend on the existing conditions. The Sub-Commission does not wish to discuss the various alternative methods which are available, but emphasizes in this case.
surveys for development and testing of methods.

38. It also emphasizes that an organization for the estimation of crop yields can only function properly if it operates on a permanent basis so that field workers can be trained and experience accumulated.

39. If a system of yield estimation by crop reporters is in existence this should be continued unchanged until results from sample harvestings have been accumulated over a number of years, so that an adequate test of the accuracy of the crop reporting can be made. To be of value the comparison should cover both good and bad seasons as the errors in crop-reporters’ estimates frequently vary considerably with the level of yield of the crop.

40. The Sub-Commission emphasizes that the total number of fields from which sample harvestings require to be taken to determine the average yield of a crop with reasonable accuracy is not large. Except in the case of mixed crops the variation per field is not likely to be greater than about 50 per cent (standard deviation), in which case a random sample of 400 fields will give an estimate of average yield with a sampling standard error of 2 1/2 per cent (limits of error approximately plus and minus 5 per cent). The use of a two stage sampling process for the selection of fields may increase the number of fields required to a certain extent. If separate estimates are required for different districts, it may also be necessary to increase the total number of fields in order to obtain the required accuracy in the district estimates.

41. In estimating total production from sample harvestings corrections will be required in the gross yield per unit area in order to allow for such factors as losses in full scale harvesting and threshing, changes in moisture content, discrepancies due to the inclusion of uncropped areas (edges of fields, etc.) in the area statistics etc. These can be quite large. Thus in sample harvestings of potatoes carried out in a number of counties of England and Wales in 1948 (approximately 1500 fields) the following results were obtained:1

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross yield of samples</td>
<td>10.8</td>
</tr>
<tr>
<td>Corrections for potatoes left in the ground, uncropped areas, and change of riddle size</td>
<td>1.5</td>
</tr>
<tr>
<td>Net yield from samples</td>
<td></td>
</tr>
<tr>
<td>Comparable estimate of Ministry of Agriculture</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td>8.1</td>
</tr>
</tbody>
</table>
Special investigation may be required to determine the relevant corrections. If commercial yields are subsequently available, as for example figures on marketing for a sub-sample (not necessarily random) of the sampled fields these may also be compared with the corresponding sample yields, and give a valuable check on the corrections. In the above example of potato yields the average difference between the gross sample yields and commercial yields was 1.6 tons per acre, agreeing closely with the total correction of 1.5 tons per acre obtained by investigating the loss factors.

42. Apart from these corrections the sample harvestings will give an objective and unbiased estimate of production if, but only if, scrupulous attention is paid to the sampling procedure. In particular:

(a) the fields must be selected at random (subject to restrictions such as stratification);

(b) the areas for sample harvesting must be located by some procedure which is equivalent to random location, and which is uninfluenced by the appearance etc. of the crop in the locality;

(c) the areas must be accurately demarcated;

(d) the time of cutting must be sufficiently near to the time of harvest for the amount of additional growth to be negligible.

In general, although successful results can be obtained by trained investigators with very small harvested areas, with untrained investigators, it is best to use areas of such a size that the danger of bias arising under headings (b) and (c) is thereby reduced. In any case objective tests should be made from time to time against full-scale harvestings.

43. Sample harvestings can be used in conjunction with eye estimation, a sub-sample of the fields of which eye estimates are made being subjected to sample harvesting. This is a two-phase sampling scheme and will serve to calibrate the eye estimates so that the effects of bias in these estimates can be removed. This procedure, however, requires to be tested out on an adequate scale before its practical value can be assessed.

**Crop forecasting**

44. Crop forecasting has hitherto been almost exclusively based on reports of crop reporters etc. of their opinions of the prospects of the crop. Such forecasts are necessarily subjective, though not for that reason valueless. On the other hand, it is usually difficult to assess their accuracy owing to the absence of objective estimates of yield. Frequently forecasts and estimates are made by the same reporters, and factors which influence them in arriving at their forecasts are likely to influence them
45. Recently meteorological data have been used in the United States in conjunction with crop reports to improve the forecasts. In special cases also, e.g. in areas in which the wheat yield is mainly dependent on the rainfall of the previous autumn, the winter forecasts based on meteorological data alone have proved of value.

46. It is to be expected that in many cases better forecasts could be made if the condition of the crop at the time of the harvest was determined by physical measurements on suitably chosen samples which could be used either alone or in conjunction with meteorological data. Little work, however, has been done in this field. The measurements which are most suitable will vary from crop to crop and possibly among different districts for the same crop. They can only be determined by carrying out trial schemes of measurement over a number of years, and investigating the relation with final yield. The evaluation of these relations and of the relations with meteorological data requires somewhat complicated statistical analyses, and the advice of a statistician experienced in this type of work should therefore be sought before this kind of investigation is undertaken.

47. Since objective estimates of the yields will be required for any plan of this kind, the institution of a scheme of sample harvesting provides a good opportunity for setting up a scheme of crop measurements with a view to seeing whether a satisfactory method of forecasting can be devised. In the initial stages it is frequently better to carry out the observations on a special sample of fields, rather than on the more scattered sample which will be suitable for crop estimation proper. More intensive investigations can also be carried out on experimental plots, on which different varieties are grown, with different sowing dates, cultural treatments etc. Experimental plots, however, are not likely to provide all the information required for the evolution of a suitable forecasting scheme.

48. It must not be assumed that it will in all cases be possible to evolve a suitable method of forecasting even if extensive work on the above lines is undertaken. The yield of a crop is influenced to a greater or less extent by meteorological and other events up to and sometimes after harvest, and this may introduce too great a degree of uncertainty into yields predicted one or more months before harvest, to make the predictions of any value.
Recommendation about future reports

49. The Sub-Commission wishes to thank the Food and Agriculture Organization for preparing the three papers under review (E/CN.3/Sub.1/14, E/CN.3/Sub.1/17 and 17/Add.1). The Sub-Commission would appreciate receiving similar reports in future. It desires that in preparing such reports special attention be directed to the actual level of accuracy of figures relating to crop areas obtained by complete enumeration or equivalent procedures. It is not sufficient to give purely subjective opinions. It has been recommended in this report that independent checks should always be made to assess the accuracy of the reported figures. The Sub-Commission requests that mention should be made of the results of such objective checks.
CHAPTER V

EDUCATION AND TRAINING IN SAMPLING

50. The Sub-Commission further considered certain aspects of education in statistical sampling methods, having noted that university curricula generally did not provide adequately for systematic instruction in the theory of sampling or for the application of sampling methods to practical problems. It feels that while the primary responsibility for introducing and conducting courses along these lines must rest with universities, institutes and other national and private organizations, some encouragement could be given by international organizations. In particular, in the field of sampling, the Sub-Commission feels that attention would be drawn both to the need for sampling courses and to standards for them if it prepared syllabuses setting forth the minimum content of university courses and professional training at the higher level.

51. Some members of the Sub-Commission therefore prepared syllabuses considered applicable to different circumstances (see Appendix A). The Sub-Commission requested the Statistical Office to circulate the syllabuses to appropriate organizations and individuals for comment and to prepare a report for its consideration at the next session.

52. It also draws attention to the need to bring up-to-date periodically the bibliography in sampling prepared by the Food and Agriculture Organization early in 1947. An important purpose would be served by periodic revision particularly if it could be widely circulated to statisticians in all parts of the world. The Sub-Commission therefore requests the Statistical Office to work out arrangements for a revised bibliography in co-operation with other organizations and individuals interested in the field of sampling.

53. The Sub-Commission also draws attention to the usefulness of a standardized and adequate classification of statistical subjects. Since this proposal is considerably broader than the field of statistical sampling, the Sub-Commission requests that this matter be taken up by the Statistical Commission at its next session.
CHAPTER VI

SAMPLING METHODS IN RELATION TO POPULATION CENSUSES

54. In the report of its first session (E/CN.3/37), the Sub-Commission indicated the different ways in which the sampling methods can be usefully employed in substituting for or in conjunction with complete enumerations. These ways include:

(i) a sample survey may be carried out by the same agency as an integral part of a complete census to obtain supplementary data,

(ii) a sample survey may be carried out independently at or about the same time as a complete census to assess the margin of error, comparative speed, cost and convenience or organization,

(iii) a sample survey may be carried out as a substitute for a complete census where funds, physical facilities and suitable personnel are inadequate, and

(iv) sample surveys may be carried out in inter-censal years for the study of demographic changes.

In addition, sampling methods may be used in tabulation to speed up the results and to maintain control over the accuracy of the statistical analysis. This has been discussed in Chapter VII.

55. The report of the first session noted that all these recommendations were subject to the important provision that, "a sample survey should be carried out only under the technical guidance of professional statisticians not only with adequate knowledge of sampling theory but also with actual experience in sampling practice, and with the help of a properly trained field and computing staff."

56. More than 30 countries are planning to take population censuses (complete enumerations) around 1950. Since the usefulness of sampling methods in connection with complete censuses has been discussed widely in the last few years, it is likely that a number of countries will carry out some type of sample survey. It is possible that some of these countries will have relatively little experience in census taking or in sampling, and therefore it may be useful to review some of the considerations which should be kept in mind in organizing sample surveys in such countries.
Sample surveys as an integral part of a complete enumeration

57. The number of subjects which governments will include in a census questionnaire will vary according to their needs for statistics. Generally, however, countries which have no elaborate census organization and a somewhat limited technical staff will find it convenient to collect information on a relatively small number of specially important topics, such as total population, age, sex, marital status, and some economic characteristics, such as (a) total economically active and inactive population, (b) occupation, industry and industrial status, (c) population dependent on various types of economic activities and (d) additional details regarding the agricultural population. These items were considered particularly important by the Statistical Commission and the Population Commission of the United Nations. In addition to these, a number of items of considerable demographic interest were included in the list recommended by the Population Commission such as, place of birth, citizenship, mother tongue, educational characteristics, and fertility data.

58. In deciding what questions to include in the questionnaire, a government may distinguish between those questions which are required for certain legal or administrative purposes and those which are not required in full detail. The required questions may be asked of every inhabitant and the supplementary ones may be the subject of a sampling enquiry. Sample surveys can conveniently be used in cases where estimates for fairly large regions or for the whole country are usually adequate. Fertility statistics are an example.

59. The proper organization of census operations requires much preparatory work. This includes the determination of territorial regions within a country and then the demarcation of smaller civil divisions. The divisions may then need to be divided into further sub-sections in one or a number of successive steps to define the operational areas of the field units. The areas in every stage of division should be clearly delimited and complete lists of such areas prepared.

60. The smallest area, for census purposes, will be a unit of area which will be suitable in size for an enumerator to cover within a specified period of time. This will be called here an "enumeration unit". The units will not be equal in size all over the country. In areas where the density of population is high, the units will be
61. In a sample survey which is an integral part of a census the process of complete enumeration creates the frame for the sample survey. One of the easiest ways of obtaining a sample is for the enumerators to secure additional items for every n-th household or every n-th individual starting with a number chosen at random. Under proper safeguards this method of selection of units is convenient in practice and some or all the information for the sample-units can be tabulated separately as described in Chapter VII.

62. In the above method the additional information will be collected by the same enumerators who are filling out the forms for complete enumeration. Sometimes it may be considered advisable that a different set of enumerators should obtain the additional information. In this way the additional information can be collected by more qualified enumerators. There are circumstances under which the enumeration-unit may be used as a sample-unit, to be sampled in one or more stages. A purposive selection of the sample-units (with a view to making the sample representative of the total population) will not be a correct method, as the results will usually be biased. In order to prevent bias the selection, of course, must not be left to the enumerator. A simple plan, again, will be to select every n-th enumeration-unit starting with a random number. Alternatively the enumeration-units can be stratified before selection by a random procedure.

63. The size of a sample suitable to obtain the results of a desired level of accuracy within required confidence limits can be determined, provided some previous information about the variance of the sample units in respect of the characteristics under study is available. When such information is not readily available, it is often worthwhile before undertaking a sample survey on a large scale to organize a pilot survey to obtain, among other particulars, an estimate of the standard error of the mean of one or more of the important characteristics under study. The necessary sample size can then be determined on the basis of (a) the estimated standard error and (b) the limits of accuracy required of the results.

64. It will usually be advisable to adopt a sample size which will be organizationally convenient and suited to available resources. From the point of view of minimizing additional work, a sample of 1 in 50, or 1 in 20 may be found quite convenient. If separate
ered important, then a higher fraction like 1 in 10 may be necessary. Also where there are special reasons for closer study of certain areas a high sampling fraction such as 1 in 10 may be used. A simple procedure is to adopt a uniform sampling fraction of, say 1 in 20 for all regions and all civil districts.

Independent sample survey concurrently with the complete enumeration

65. There are great advantages in organizing a sample survey at about the same time as a complete enumeration and in collecting the information by an entirely independent set of investigators. Such an independent sample survey may take two forms. The first may extend over the whole of the geographical area of a country covered by the complete enumeration. In this case the sample survey supplies independent results together with valid estimates of the margin of error with the help of which it is possible to assess the accuracy of the complete enumeration. If the complete enumeration is considered to be unreliable in any respect, it is also possible to obtain useful results in the form of aggregates from the data collected in the sample survey. In addition it is possible to use the independent sample survey to collect information for items not included in the census questionnaire.

66. The second form of the independent sample survey is to use it in certain selected regions but not over the whole country. In this case, the results of the sample survey can be used to assess the accuracy of the complete enumeration in such regions, and hence assess the general reliability of the complete enumeration. Such restricted sample surveys cannot, however, yield results for the whole country, and are not suitable for collecting supplementary data (except such information as may be of regional interest). Either form of the independent sample survey may be used with advantage to assess the cost, speed, adaptability and efficiency of organization (in addition to the accuracy) of the sampling methods in comparison with complete enumeration.

67. In organizing an independent sample survey it is possible to use the same frame as that of the complete enumeration. Although the same frame is used, the information will be collected by an entirely independent agency.

68. Another possibility is to use an entirely different frame from
of households. The size (i.e. the area) of such sample-units can naturally vary, and such variations may have to be taken into consideration both at the time of selection and at the time of statistical analysis. Another possibility particularly in rural areas is to locate sample-units of definite size and shape in a random manner on large-scale maps, and to enumerate all households or individuals falling within the demarcated area. An area sample-unit of either type will almost of necessity have to be prepared and selected in a central or branch office possessing facilities for such work.

69. Sample surveys may be, of course, carried out in the way indicated above, as a substitute for a complete census, to collect required demographic information. It is also possible to conduct such independent sample surveys in intercensal years in which case it is convenient to organize the work on a comparatively small scale and repeat it at frequent intervals preferably every year. The successive series of results obtained in this way can supply valuable and up-to-date information on changes in demographic conditions from year to year.
CHAPTER VII

SAMPLING METHODS AND TABULATION

Early Tabulations of Complete Returns Through Sampling

70. It is possible to speed up the tabulations that result from a complete census through the applications of comparatively simple sampling techniques. A complete national census is a huge work. Unless sampling methods are introduced, months and even several years may elapse beyond the date to which the census refers before some of the tabulations of the greatest economic and social interest are finished. This lag may be expected particularly in countries which are large and which do not take censuses regularly and frequently. Even in a small country which takes censuses of population or agriculture regularly, and which may have become a model of efficiency, there is often a finite need for early tabulations of some of the results. Moreover, it has often happened that funds appropriated for censuses were exhausted before the tabulations could be completed.

71. National or regional totals and averages of many characteristics are often of great economic and social utility, provided they are not long delayed. For example, in a census of agriculture, some of these characteristics might be the number of farmers aged 30-39, the number of workers who were hired for cash, and the total cash paid to them, the number of acres in certain major crops, the number of acres in these crops in the previous year, the number of sheep, cows, horses, etc., the yield of certain crops, the number of sheep or cows sold during the past year, the total population living on farms, the average population per household living on farms. It should be pointed out that for purposes of forecasting and framing economic and social policy, either by government or private enterprise, such tabulations may be obtained speedily and readily by proper procedures of sampling.

72. Sampling cannot be used for determining the total number of people having any particular characteristic for very small areas. Such information is more economically come by tabulation of the complete count. The number of people for each small area of the country is fortunately a comparatively simple tabulation which for speed may be carried out by hand before the schedules are sent into the field. This is the practice in many countries.

73. The frame and the domain of study will have been definitely settled by the complete census, and all that remains is to develop methods of sampling or arrive with sufficient common knowledge.
74. Such a sample is a sample of the complete census. The complete census will be the result of applying a series of operations such as the hiring, training and supervising of the enumerators, and the development of a particular questionnaire. The returns will contain biases of response and non-response. There will be incompleteness of coverage and a certain amount of duplicated coverage. There will be incomplete returns and other failures. The sample cannot overcome these difficulties. The sample is a sample of the returns as they are.

75. The sampling techniques to be applied will depend upon the nature of the tabulations that are to be derived. In this report it is not intended to develop the theory for the various techniques of sampling, but a few general remarks may be made.

76. The design of the sample and the inter- of the results obtained are problems in theory and practice which for several reasons may be comparatively simple: (a) advantage may be taken of the efficiency of small sample units, which may be the individual person, the household, the farm, etc., as the smallest units are actually the easiest to define in the returns; and (b) the sampling may in most cases be carried out at a single stage at little or no extra cost; (c) there is no field-work to be carried out; (d) geographic stratification by area is usually automatic, as the returns are filed in this manner anyhow. Further stratification, as by type and size of farm, may usually be accomplished readily.

77. It may be that a portion of the population covered will have been designated as a sample, for broadening the scope of the census for other purposes. This sample may be ideal for early tabulations. An example is furnished by the census of population in the United States for 1940. A 1:20 sample by individual people had been designated in the enumeration, to broaden the scope of the census. This sample served admirably for early returns on the employment characteristics and work-status of the adult population. The same sample also served for detailed cross-classification later on for the analysis of the family and household characteristics.

78. Competent statistical advice should be obtained and followed in the selection and tabulation of the sample of the returns, and in the interpretation of the results.

Reducing the Cost of Detailed Tabulations for Large Areas through the Use of Sampling

79. Sampling may be used not only for early tabulation, but to effect savings in the cost of detailed tabulations for large areas. The limiting factors are the amount of work to be done and the time in which it is to be done. The sample must cover the entire area or at least as much of it as is necessary for the purpose in hand. It may be possible to reduce the sample size by careful selection of the areas to be included, or by the use of a smaller sample in the earlier stages of the investigation.
tabulation is large, fairly detailed tabulations of numerous population and agricultural characteristics may be carried out by sampling the returns of a complete coverage, thus effecting considerable savings and permitting many more tables to be published than would otherwise be possible.

80. The detailed tabulations will ordinarily be made after the editing has been accomplished. In countries where mechanical tabulation is used, these sample tabulations may well be delayed until the cards are punched.

81. The sample that is to be used for the detailed tabulations may or may not be the same as the sample that was selected for the early returns. In any case, competent advice should be sought.

82. In the census of population in the United States for 1940, the sample already mentioned was used for obtaining the detailed cross-classifications of the characteristics of the population, and it was used also for tabulating the family and household characteristics. Because of the savings in funds and time thus accomplished, as well as because of the decreased load of tabulation, it was possible to publish many more volumes of tables than would otherwise have been possible.

**Speeding and Lowering the Costs of Processing the Returns of a Census Through the Introduction of Sampling Inspection**

83. The sheets of paper on which the data of a census are collected must go through many stages of processing before they may be summarized in the form of statistical tables. Thus, editing may be required in an attempt to correct errors, misunderstandings and occasional omissions of entries, on the part of the enumerators, and to bring the terminology into uniformity. In later stages, whether the final tabulations are to be carried out by machine or by hand, some of the information will usually require coding. Thus, wheat may be coded 0, oats 1, rice 3, etc. Finally, the data are summarized by hand or are punched into cards for sorting and tabulating by machine.

84. Definite rules and definitions must be written down for each stage in the processing of the data. There should be criteria by which any worker may judge whether any detailed part of his work meets the specifications. Two workers will occasionally show differences in editing or coding, which cannot be classified as errors but rather as difference in judgment. Nevertheless the procedure should be made as definite and as uniform as possible.

85. It is inevitable that some real errors should occur in the processing. Hence, in an effort to control these errors, it is customary to combine each stage of editing, coding, tallying, or punching with a review or inspection.
the number of people or the number of farms as recorded in the census should be made as exact as reasonably possible and in fact may be required by law so to be.

86. On the other hand, since the numerical material handled is itself in general subject to variation and errors of various kinds, absolute numerical accuracy need not necessarily be attained at the early stages of the calculations. Thus, the error of classifying a few acres of pastureland as crop land may not be serious, unless it is repeated persistently. For this reason it is sometimes practicable to impose sample checks on such operations as punching and coding in large-scale work. If such checks are relied on, however, it is essential that steps be taken to prevent gross errors of the type that will produce material errors in the results. Thus a single misclassified card in a count will usually produce an entirely trivial error in the results, but the mispunching of a number representing a quantitative character, e.g. 610 for 010, may produce a serious error in the resultant mean or total of the class in which the unit falls.

87. The amount and character of these errors can be determined by sampling. It is easy to overdo the amount of verification for characteristics for which exact conformity with the enumeration is meaningless and useless. The aim of verification should be to ensure that the data are satisfactory and to do so at minimum cost. Stated negatively, the aim of inspection should not be to correct small errors. The frequency of such errors from each worker should be maintained so low that correction is not necessary, in which case the inspection is entirely an operation of control, and not an operation of correction.

88. To this end, instead of reviewing every stroke of work that is done in the editing and punching, it has been found more practicable in some cases to introduce methods such as are used for the statistical control of quality, the aim of which is the maintenance of some desirable and economic level of performance. In the statistical control of the processing, the verification of the work that is carried out at any stage has two main functions:

(a) to determine the proportion of errors made by each worker during some prescribed period of time, such as two weeks;
(b) to provide methods for improving the work of each worker:
   (i) by bringing to their individual attention both the nature and the quantity of the errors that they are making,
   (ii) by supplying a basis for bonuses.

89. The savings in both time and money which may be accomplished by the use of these methods may assume large proportions.
90. Whatever system of checking is adopted, it is important to recognize that one of the main functions of a checking system should be to maintain a high standard of accuracy throughout the work. If this is not done, some material errors are almost certain to escape correction (or to be corrected wrongly). For this reason a minimum standard should be laid down for all workers and those not attaining this standard should be transferred to other work.
CHAPTER VIII

SAMPLING IN MANPOWER STATISTICS

91. The Sub-Commission discussed a note (E/CN.3/Sub.1/13) prepared by the International Labour Office requesting advice upon the application of sampling methods to obtain statistics on the extent and trends of unemployment and the extent and development of under-employment and job openings. The statistics would relate to economically under-developed countries desirous of developing their labour statistics but not now possessing extensive statistical resources. The statistics should be adequate for policy-making purposes. The Food and Agriculture Organization also manifested interest in this problem since almost invariably a high proportion of the population of an under-developed country is dependent upon agricultural occupations.

92. The question has two aspects. The first relates to the problem of formulating precise definitions in regard to unemployment, under-employment and job opportunities, the second to sampling and its applicability to the statistics sought. Definitions are necessary irrespective of whether the information is collected by complete enumeration or by sampling. It may also be mentioned here that existing conceptions dealing with the various aspects of labour statistics under discussion have generally tended to evolve in advanced industrial economies. Some of these advanced countries, such as Canada and the United States, are carrying out labour force surveys at quarterly or monthly intervals, but even in these countries there are certain aspects of the definitions of unemployment and under-employment which are subject to continuous research and investigation. In under-developed countries where the nature of the organization of the productive activities is different, the existing conceptions used in the industrialized countries may lose their sharpness and as a result become unsuitable. The formulation of precise definitions is, therefore, a very important problem.

93. The two aspects are inter-related. The answers to questions of whether or not an individual or a household or a group of households should be the sample-unit of enquiry and how the survey should be designed depends on the specific nature of the particulars to be collected. On the other hand, the proper planning of surveys from the sampling point of view will encourage clear formulation of definitions, which are applicable to the conditions under investigation. Definitions of employment and
organized on a small scale and at the same time can be designed to bring out the significant points of difference between different methods of approach, questionnaires of various types may be tried out to test the validity and usefulness of the results.

94. So far the household has been the sample unit in all countries where labour force surveys have been carried out on a sample basis. Such a sample unit will be suitable for under-developed countries also. Under certain circumstances, depending upon the nature of the economic organization, an entire village may be selected as the sample unit for employment studies. In the initial stage of the investigation it will perhaps be helpful to obtain a complete economic pattern for the village as a whole. Although the efficiency of sampling for the provision of overall employment statistics is likely to be reduced, information will be obtained on the village economy as a whole. It is possible that all households or, alternatively, a random sample of households within each village will be investigated. In pilot investigations a few villages from a number of contrasted regions such as agricultural and industrial areas in various stages of development may be taken up for intensive study.

95. In general, information is likely to be required on the following points relating to individuals:

(i) the number and types of occupations (including unpaid family labour) pursued during a given period of time,

(ii) the relative importance of these occupations in terms of time and earnings,

(iii) the places of work,

(iv) industrial status (employee, employer, self-employed, etc.)

96. It may be mentioned finally that certain concepts may be found to be ascertainable while some others may be statistically unascertainable. As a first step, properly designed pilot surveys must be undertaken to clarify the various problems. It may be in certain cases that these pilot surveys will show that full-scale surveys are impracticable.

97. It should be emphasized that because of the varying degrees of economic development the concepts in regard to employment, unemployment, and under-employment will be found to be very different in various countries. The statistics obtained from all the countries will not therefore be comparable and many precautions are necessary when figures on one country are studied in comparison with another.
CHAPTER IX

SAMPLE SURVEYS OF CURRENT INTEREST

98. As a part of its continuing responsibility to "examine the use which is being made of statistical sampling in different countries and in different fields of subject matter", the Sub-Commission examined the second report of the Statistical Office on Sample Surveys of Current Interest (E/CN.3/Sub.1/15). An extensive report on the Problem of Sampling African Populations was presented by the Director of Census and Statistics, Central African Statistical Office, Southern Rhodesia (E/CN.3/Sub.1/W.10). In addition the Sub-Commission heard an oral report on the recent experience of the Institut National de la Statistique et des Etudes Economiques of France in planning and conducting national sample surveys (E/CN.3/Sub.1/SR.32). Members of the Sub-Commission and observers supplemented the documentary information with comments on other important sample surveys known to them.

99. The Sub-Commission wishes to express its appreciation to governments and research agencies for their response to its request for reports on sample surveys in order that such information may be made available to all statisticians interested in sampling. From the reports available it is apparent that increasing use is being made of sampling methods throughout the world and that sampling techniques are being applied in a great many subject fields. Sampling techniques which are both sounder and more efficient are rapidly coming into use.

100. The Sub-Commission requests that the Secretary-General continue to make these reports on sample surveys of current interest available to interested statisticians and statistical organizations. The Sub-Commission noted that the matter of selection of the sampling unit is not explained in many of the reports. It is desirous that this should be done. The Sub-Commission also requests that the attention of those who are preparing reports on current sample surveys be drawn to the recommendations of the Sub-Commission concerning such reports, especially regarding the method of selection of the sample units.
CHAPTER X

METHODS TO DETERMINE VOLUME OF POSTAL TRAFFIC

101. The Universal Postal Union requested the Sub-Commission on Statistical Sampling to express preliminary views on the methods to be applied in the collection of statistics on postal traffic (E/CN.3/Sub.1/4, 9). The postal statistics compiled and published by the Universal Postal Union are collected and supplied by the postal administrations of the Member States. Under the Postal Convention of 1947, individual Member States have responsibility for the way in which their statistics are collected; therefore no uniform standards are applied.

102. The Sub-Commission agreed that sampling methods can be expected to produce great improvements in the reliability of postal statistics. Sampling methods have been tried out in India in 1946 with promising results and are being used for the investigation of special problems in the United States. It is, of course, only possible to make general recommendations because detailed comments and suggestions of possible sampling plans can be made only by a statistician who is close to the work in one or another of the Member States, and thoroughly familiar with it and is in a position to undertake pilot investigations.

103. The Sub-Commission agreed further that:

(a) A sampling plan for the counting and weighing of pieces of mail should be operated at suitable, specified, and frequent intervals throughout the year. Thus in the pilot study in India every 13th day was used for the counting and weighing of some particular class of mail in certain selected post offices. The present custom of selecting a particular period regarded as "representative" is hazardous; such practice is subject to unknown and unknowable biases.

(b) The offices at which the weighing and counting are to be carried out should be selected according to a proper sample design. In particular, counting and weighing would be required only in a small fraction of small offices, in some larger fraction of the medium-sized offices, and possibly in all of the largest offices.

(c) The selection of the particular mail bags which are to be investigated should be specified in the sampling plan.

(d) The procedure of computation for forming the final estimates and for computing the errors would also be included as part of the sampling plan.
(e) Such procedures would:

(i) prevent biases due to human preference and judgment in the selection of the sample;

(ii) produce improved reliability and permit the magnitudes of the errors in the estimates to be calculated and published for the benefit of users of the data;

(iii) reduce the burden of collecting the postal data and improve the quality of the work. This result can be expected because the work will be spread out over the year, instead of being concentrated in a short period;

(iv) speed up the results; for example, when proper methods are developed, the statistics on counting and weighing for any year or other period might well be ready within a few weeks following the close of the period.

(f) The compilation of detailed information about postal notes, money orders, registered letters, etc., may also be speeded up and spread uniformly over the year by the introduction of sampling techniques.

104. The adoption of sampling methods in the collection of postal statistics would permit the Postal Administrations to have fuller information on which to base their operations as well as to supply the Universal Postal Union with more comparable statistics for international comparisons. Wholesale and rapid changes cannot be expected and are not practicable. Except for some scattered experience, the practical aspects of sampling in this field remain to be developed. Development of new statistical methods should preferably start on a small scale in such countries as are able to obtain the services of qualified sampling statisticians. Successful experience would undoubtedly encourage others to modify their present methods.

105. If a rigorous sampling scheme is instituted for trial over a period, the existing scheme for the collection of data should be continued, so that comparisons can be made should any discrepancies be revealed in the existing statistics.

106. With regard to the current publications of postal statistics published by the Universal Postal Union, the Sub-Commission further agrees and wishes to point out that the publication of any statistical figure is incomplete unless it is accompanied by some sort of statement regarding the methods used and the possible errors and sources of error in the figure. The Universal Postal Union might therefore profitably consider a programme by which the Member States would be urged to evaluate the possible errors and sources of error in their figures. Such a programme would not only improve the usefulness of the current figures but would also pave the way for
CHAPTER XI
OTHER SUBJECTS DISCUSSED

Quota and Random Sampling

107. The Sub-Commission had before it a paper on quota and random sampling (E/CN.3/501.18) reporting the results of the Committee on Analysis of Pre-election Polls and Forecasts set up in the United States following the failure of these polls at the last Presidential election.

108. The Sub-Commission has already expressed its opinion on quota sampling in general terms in the Memorandum on the Preparation of Sampling Survey Procedures (Statistical Papers, Series C, No. 1) prepared at the second session. It was then stated

"...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.

"The reporter should describe the procedure used in selecting sample-units, and if it is not a random selection he should indicate the evidence on which he relies for adopting an alternative procedure. Purposive selection and quota sampling cannot be regarded as equivalent to random sampling."

109. The Sub-Commission does not wish to enter into a detailed discussion of the weaknesses of quota sampling, which are well described in the report of the Committee referred to above. It may however, be of value to indicate the main defects to which quota sampling is subject.

110. Superficially a quota sample has the same structure as a stratified sample, the quota classes constituting the strata. Selection from within these strata, however, is not necessarily by any means random. Consequently in a quota sample the danger of bias is always present. The quota method must therefore be ruled out as a suitable method of investigation for precise enquiries in which unbiased results are required.
It is frequently claimed that a quota sample is in effect a stratified random sample, the quota classes constituting the strata. This is not the case, for two reasons. In the first place selection from within the quota classes is not by any means random. It is well known, and has been abundantly demonstrated, that haphazard or judgment selection does not give the equivalent of a random sample. In the second place the quota classes are very frequently ill-defined, or if properly defined, are such that the numbers of the population falling in these classes cannot be, or are not, accurately determined. The definitions, even if precise, are often such that the investigators cannot classify the numbers of the population accurately in the field. Thus social classes are vague and ill-defined; income classes can be precisely defined but the numbers in the various classes in the population are not usually known, nor can the investigators in practice determine at all accurately the incomes of the selected individuals.

In consequence of these two facts bias is always present in a quota sample, and the amount of this bias is often large. Nor does multiplicity of quota classes help matters. The quota method must therefore be ruled out as a suitable method of investigation for precise enquiries in which unbiased results are required.

The use of quota sampling in particular circumstances is often justified by comparison of the results with those obtained subsequently from other sources, e.g. pre-election polls have in the past been justified by the fact that they successfully predicted the actual polling. The Sub-Commission feels that it is not sufficiently recognized that if there is a change of conditions a quota sample which has previously adequately reproduced the characteristics of the population may cease to do so. Consequently the fact that a quota system has consistently given reliable results over a period of years is no guarantee that it will also do so in the future.

It is recognized that if used with skill the quota method may give indication of preferences and changes of opinion in simple enquiries where no great accuracy is required. No measure of the accuracy attained is, however, possible and the results of quota sampling must therefore be regarded as subjective and should never be relied on when objective and unbiased information is required.
115. The Sub-Commission considers that in addition to the items in which it has expressed a continuing interest such as crop reporting, national income and the reports of current sample surveys the members should explore the documentation available on operational research. The Sub-Commission will also consider the application of sampling to other fields of tabulation at its next session. During the first three sessions the Sub-Commission has dealt with a number of specific subjects and requests the Secretariat to bring new developments in these subject fields to its attention.
APPENDIX A

Syllabus for professional statisticians who intend to become experts in sampling, census and survey methods

by F. Yates

1. Theoretical background
   (a) Theory of estimation;
   (b) Errors of estimates and tests of significance;
   (c) Analysis of variance;
   (d) Principles of the design of experiments.

2. Principles of sampling
   (a) Selection of sample;
   (b) Biases in selection;
   (c) Estimation of values relevant to the population and parts thereof; alternative estimates; bias in estimation;
   (d) Use of supplementary information; ratio and regression;
   (e) Sampling errors of estimates and their estimation.

3. Principles of design
   (a) Determination of size of sample necessary to give results of the required accuracy;
   (b) Estimation of the relative efficiency of different types of sample;
   (c) Minimization of costs;
   (d) Minimization of costs plus losses due to errors.

4. Types of sample (the aspects outlined in Sections 2 and 3 should be covered)
   (a) Random sample;
   (b) Stratified sample with uniform and variable sampling fractions, and sample stratified after selection; multiple stratification;
   (c) Multistage samples;
   (d) Multiphase samples;
   (e) Sampling with probability proportional to size of unit, from the population and from within strata;
   (f) Point and line samples, and area sampling generally;
   (g) Systematic samples from lists and from areas;
   (h) Interpenetrating samples;
   (i) Sampling on successive occasions.

5. Practical aspects
   (a) Information to be collected: planning of observations and questionnaires;
(c) The frame: frames suitable for various types of survey;  
(d) Selection, training and supervision of field investigators;  
(e) Methods of abstracting and summarizing the data, including the  
    use of punched card machines and methods of coding;  
(f) Control of accuracy of field work and computations;  
(g) The use of sampling in statistical analyses;  
(h) The analysis of costs;  
(i) The preparation of reports.  

6. Types of survey (the aspects outlined in Section 5 should be covered)  
(a) Censuses of human populations; demographic problems;  
(b) Surveys of social and economic conditions of human populations;  
(c) Surveys of industry and economic institutions;  
(d) Market research and opinion surveys;  
(e) Agricultural censuses and surveys, including crop estimation  
    and forecasting.  

7. Pilot surveys  
The design and use of pilot surveys in the planning of large scale  
surveys, and in the evolution of field procedure and training of investigators.  

8. Critical statistical analysis of survey results  
More advanced methods of handling survey data. Comparison of survey  
results with information from other sources. Methods of estimating the  
effects of a factor, freed from the effects of extraneous factors. Limitation  
of this procedure and of conclusions that can be drawn from survey data.  

---  

Syllabus for an Advanced (Professional) Course in  
Statistical Sampling  
by P. C. Mahalanobis  

This course is designed to suit the requirements of advanced students who  
have already had basic training in statistical methods, and desire to qualify  
as professional sampling statisticians. The level and scope of such previous  
training have been indicated in Part 1 of the syllabus. The topics are, of  
course, neither mandatory nor exhaustive; and persons with considerable  
practical experience of sampling surveys should also be able to take the  
subsequent parts of the course with profit.  

Part 1 Prerequisites in Statistical Methods  
(1) Elements of probability theory; definitions; total probability and  
    compound probability; binomial and multinomial probability; Bernouilli's  
    Theorem; Poisson and normal distributions; mathematical expectations; law of  
    large numbers; continuous probability; central limit tending theorems
interpolated and extrapolated results.

(3) Graduation of frequency data by Pearsonian and other types of curves; graduation of time-series like data by polynomials etc.

(4) Study of frequency data: measures of location, dispersion, skewness, and kurtosis; binomial, Poisson, and normal distributions with important properties; bivariate and multivariate normal distributions; total, multiple and partial correlation and other measures of association and regression.

(5) Theory of estimation: basic concepts of consistency, efficiency, sufficiency, and maximum likelihood, important formulae and theorems (without proof); the theory of information (without proof); fiducial probability and confidence interval (without proof); basic concepts of tests of hypotheses, and important theorems of statistical inference (without proof).

(6) Tests of significance: mathematical derivation of the distribution of t, F, z, and \( X^2 \) (chi-square) with applications; distribution of \( r \) (coefficient of correlation) and its use for both \( z = 0 \) and \( z \neq 0 \); tests of significance for regression coefficients (with proof); Fisher's z-transformation for \( r \); tests of significance for multiple and partial correlations (without proof).

(7) Analysis of variance: basic concepts; orthogonality; important theorems.

(8) Design of experiments: basic concepts; randomized blocks; Latin squares; factorial experiments; confounding.

(9) Statistical quality control: basic concepts and procedures; control charts; elements of sequential analysis (without proof).

Part 2 The design of sample surveys

(1) Basic concepts: field of enquiry; "frame" for surveys; elementary units; sample-units of different types and varying sizes; sample as an aggregate of sample-units; domains of study. Distinction between random-like and non-random (or patterned) space-fields.

(2) Method of selecting sample-units: the principle of random selection; procedures equivalent to random selection; systematic (determinate-interval) selection; procedures not equivalent to random sampling such as quota or purposive selection.

(3) Various types of design: stratification, multistage; multiphase; correlated variate; combination of multistage and multiphase sampling.

(4) Cost (in terms of labour, money, and material resources) and accuracy (or margin of error) determined by the design of the survey. Relation between cost and accuracy. The general problem of design, namely, to secure either

(a) maximum accuracy at any desired level of cost, or

(b) minimum cost at any desired level of accuracy.
(5) Estimation problems for different types of design; sampling bias; sampling errors. Study of the variance function (including the case of varying size of sample-units).

(6) Components of cost: enumeration; journey; inspection; supervision; statistical analysis; overhead. Cost in terms of labour; different grades of labour. Cost in relation to material equipment and resources.

(7) Non-sampling errors: investigator bias; discrepancies arising from the form of ascertainment; physical fluctuations; mistakes at the stage of processing of the material.

(8) The use of inter-penetrating networks of samples for the control of non-sampling errors; analogy with design of experiments; residual errors (by the analysis of variance); internal errors; external errors.

(9) Pilot surveys: series of surveys on a gradually expanding scale for the improvement of the design and for training of staff and organization of the sampling organization; pilot studies in conjunction with standard surveys; use of model sampling experiments.

(10) Successive (repeated) surveys; different forms with duplicated sample-units over time; the use of control charts in connexion with successive surveys.

(11) Surveys for purposes of mapping; problems of stratification; utilization of information becoming available at each step; the design of experiment with a continuous variate.

(12) Accuracy of the results: comparison of internal and external errors; errors arising from adjustments; comparison with information from independent sources; deficiencies arising from non-response, lack of records, etc.

Part 3 Organization of the survey and preparation of reports

(1) Preparation of the questionnaire, forms, and schedules for the collection of the primary material and input of labour; concepts and definitions; standards; arrangements of items; coding, etc.; preparation of instructions for different sectors of work.

(2) Organization of the field staff: general structure; methods of inspection and supervision; controls at the collection of the primary material; the training of the field staff.

(3) Organization of statistical compilation and analysis: arrangements for flow of data from the field; preliminary scrutiny; preparation of tabulation forms; organization of various stages of processing and tabulation including cost analysis; inspection and supervision. Integrated and multi-purpose surveys.

(4) Use of sampling methods in tabulation for purposes of speeding up of
purposes; technical reports; discussion of results.

(6) Preparation of plans and budgets for sample surveys.

Part 4. Practical Work

Each candidate should be given a great deal of practice in working out simple theoretical exercises and numerical examples. He should also have practice in the preparation of forms and questionnaires and trying these out in the field; and participate in field work as investigator and inspector or supervisor in pilot or large-scale surveys. He should learn how to prepare forms for tabulation and processing of the material; have some actual practice in the scrutiny of the primary material, the organization and supervision of computational work, and the preparation of reports and plans with budgets. If possible he should be given opportunities for handling small-scale projects, and participating in the sample surveys in progress.

(N.B. No mention has been made of particular subject fields as these would be determined by local circumstances. It is desirable that the candidate should have experience of practical work in more than one subject.)
APPENDIX B

Section 1. GENERAL DESCRIPTION OF THE SURVEY

1. The first paragraph has been replaced by the following:
   A general description of the survey should be given for those who are primarily interested in the results rather than in the technical statistical aspects of the sample design, execution, and analysis. These technical aspects should be described in fuller detail in separate sections or appendices; the ground to be covered has been indicated in the subsequent sections of this memorandum. The general description should include information on the following points:

   After point e. of Section 1, the points have been re-arranged and slightly amended as follows:

   f. Repetition: State whether the survey is an isolated one undertaken without intention of repetition, or is one of a series of similar surveys.

   g. Date and Duration: The date or period of time to which the data refer should be stated, and also the starting date and period taken for the field work.

   h. Accuracy: A general indication of the accuracy attained should be given.

   i. Cost: An indication should be given of the cost of the survey, under such headings as preliminary work, field investigations, analysis, etc.

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

   k. Responsibility: The name of the organization sponsoring the survey should be stated; also the name of the one responsible for conducting it.

   l. References: References should be given to any published reports or papers relating to the survey.

Section 2. DESIGN OF THE SURVEY

2. Points b. and c. have been altered as follows, and the two last paragraphs below marked (x) and (y) have been added to point c.

   b. Elementary Units: Units which are the smallest part of the material capable of possessing a particular characteristic. Elementary units may, therefore, be of different types (e.g., age of children is a characteristic of a family, age is a
c. Sample-units: The units which form the basis of the sampling process. The sample-units may be (i) the same as the elementary units or (ii) groups of such units capable of possessing all the characteristics required. A group may consist of a cluster of contiguous elementary units, or a number of units arranged in an assigned configuration. A systematic pattern of elementary units may, for instance, constitute a sample-unit.

It is conceptually convenient that the sample-units be so defined that the totality of sample-units contains every elementary unit once and once only.

(x) 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.

(Here follows the last paragraph of point c. in the original text)

(y) Sample-units may be of the same or differing size. They may contain the same or approximately the same number of elementary units or they may contain widely differing numbers. The estimation of the population values and of the sampling errors is simplest when the sample-units within a stratum are of approximately the same size, or contain the same number of elementary units.

3. Point h. has been altered to read as follows: h. Stratification: The totality of elementary units included in the frame may be divided into groups or "strata", each stratum being sampled separately so that a specified number of sample-units is obtained from each stratum. Such strata may be geographic sub-divisions, divisions depending on some quantitative or qualitative variate appertaining to the sample-units, etc.

4. Point j. has been altered to read as follows: j. Multiple Stratification: In certain cases the totality of elementary units may be divided simultaneously according to two or more classifications, each of which depends on one geographic, quantitative, or qualitative variate. Each cell determined by the two or more way classification itself potentially constitutes a stratum of the totality of elementary units. If each of the cells is sampled separately, as in an ordinary stratified sample, the term multiple stratification may be used without qualification. If the available information is not adequate for this to be done, so that the numbers of sample-units in the main strata only can be pre-determined,
5. Point k. has been altered as follows: In the second lines of the first and second paragraphs the words "sample units" have been replaced by the word "units".

6. After the last paragraph of point k, a new paragraph has been inserted as follows: 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 adjusted so as to give a uniform overall sampling fraction. If selection is with probability proportional to size at the first stage (see Section 3), the sampling fractions at the later stages may be similarly adjusted to give equal probability of selection to all elementary units.

7. After the first paragraph under point l, the following sentence has been added: It may be noted that in multi-phase sampling, the different phases of observation relate to sample-units of the same type, while in multi-stage sampling, the sample-units are of different types at different stages.

8. At the end of the second paragraph of the original version of point l, the following sentence has been added: This is sometimes called correlated variate sampling.

9. The following paragraphs will replace the last paragraph of point l:

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

A more complicated example is the following: 100 villages are selected at random out of 5,000 villages (first-stage selection); and the number of trading establishments is counted in each selected village (first-stage observation). A selection is then made of a fraction of (or perhaps all) villages which have one or more trading establishments (first-stage second-phase selection); and in each such village, a selection is made at random of 10 farms (second-stage selection based on first-stage observation). An enquiry is made about crops grown in the selected farms (second-stage observation). A selection is then made of a fraction of (or perhaps all) farms which grow cotton (second-stage second-phase selection); and an enquiry is made among these farms whether the cotton is sold through the
10. After the first paragraph in point m. the following paragraph has been added:

It is possible to keep some of the sample-units the same in two (or more) interpenetrating networks of samples so that information for such sample-units are 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 sample-units which are observed more than once are called duplicated (triplicated, quadruplicated, etc.) sample-units.

Section 3: METHOD OF SELECTING SAMPLE-UNITS

11. The section with two paragraphs added, has been revised to read as follows:

The reporter should describe the procedure used in selecting sample-units, and if it is not a random selection he should indicate the evidence on which he relies for adopting an alternative procedure. Purposive selection and quota sampling cannot be regarded as equivalent to random sampling.

A process is properly described as random if to each unit has been initially assigned a determinate probability of being selected. 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. 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.

Ordinarily all units within a given stratum are assigned an equal probability of selection, sampling being "without replacement", i.e. no unit is included more than once in a sample. In certain cases it may be advantageous, or convenient, to select the sample-units with probabilities proportional to some known quantitative characteristic of the units, such as size; 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.

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.
Section 4. PERSONNEL AND EQUIPMENT

12. The first paragraph has been altered to read as follows: 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 may be made of the equipment used.

13. The second paragraph has been included in Section 7 - ACCURACY OF THE SURVEY as point g.

14. A new Section 5 has been added with the heading 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.

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.

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. A numerical example is often of assistance in making the procedure clear.

It frequently happens that the 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 harvest. In such cases adjustments may have to be made, the amount of which is estimated by subsidiary observations, or otherwise. Account should be given of the nature of these adjustments and the ways in which they were derived.

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. Mention should be made of further tabulations which have been
analyses which failed to yield results of interest and which are therefore not considered to be worth reporting in detail.

The inclusion of ancillary 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 material).

15. The Original Section 5 COSTS has been numbered as Section 6 and called COSTING ANALYSIS, and, the original Section 6 ACCURACY OF THE SURVEY has been numbered as Section 7.

Section 7. ACCURACY OF THE SURVEY

16. Add to Section 7 (a) as the second paragraph: The term "sampling error" has been used to denote both sampling standard error and limits of sampling error at some assigned level of probability, e.g. twice the standard error corresponding approximately to the 1 in 20 level of probability. The sampling standard errors and the limits of sampling error should be clearly differentiated and the multiples used in the latter should be specified.

17. Point c has been revised to read as follows: Other Non-Sampling Errors: Errors which are common to all investigators, and indeed any constant component of error (or "bias") in the recorded information, will not be included in the estimates of the random sampling errors deducible from the survey results. Another source of error is that due to incorrect determinations of the adjustments (referred to in Section 5) arising from observation of quantities which do not correspond exactly to the quantities of which estimates are required.

The existence and possible effects of such 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 therefore be fully discussed. Any special checks instituted to control and determine the magnitude of these errors should be described, and the results reported.

18. The last paragraph of the original Section 4 PERSONNEL AND EQUIPMENT has been made point 8. of Section 7.