STATISTICAL COMMISSION

SUB-COMMISSION ON STATISTICAL SAMPLING

Fourth session

## EDUCATION AND TRAINING IN SAMPLING

Report on the comments received at the Secretariat on
the syllabuses proposed for a course in sampling.

As requested by the Sub-Commission on Statistical Sampling, the syllabuses proposed by two of the members of the Sub-Commission at its third session in September 1949 (see Appendix A), were circulated to the national statistical offices and various statistical societies of all the countries.

Replies were received from twenty countries, six of which had no particular comments to offer. Detailed comments were received from the following fourteen countries: Brazil, Canada, Czechoslovakia, Denmark, Egypt, Finland, India, Ireland, Norway, Puerto Rico, Southern Rhodesia, United Kingdom, United States of America, and Venezuela. The comments are given below arranged according to countries:

### 1. BRAZIL

Comments from two of the members of the Brazilian Statistical Society are given below. The members are: O. Porto Carreiro and Octavio A. L. Martins.

1. You will understand my hesitation in offering any criticism of the work of such authorities. Fortunately, the points of disagreement, which I must in all frankness mention, are few. They are as follows:

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

Part 1 (Prerequisites in Statistical Methods).

I should prefer to arrange this part (Part 1, Section 1) in the following order:

Elements of probability theory; definitions; total probability and compound probability; binomial and multinomial probability; Poisson distribution; continuous probability; random variable; distribution function; mathematical expectation;

/Tchebycheff's
E/CN.3/Sub.1/21
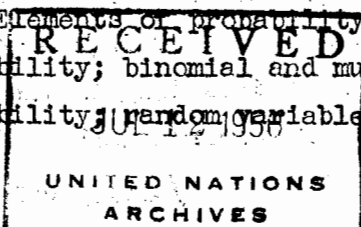
Tchebycheff's inequality; law of large numbers; Bernouilli's theorem; characteristic function; correspondence between sets of characteristic functions and sets of distribution functions (Laplace-Levy and Levy-Cramer theorems, without proof); central limit theorems; normal distribution.

Justification. These alterations would:

(a) avoid the presentation of the law of large numbers initially as a particular case, with full proof (Bernouilli's theorem) and later in a more general but still not entirely general way due to the fact that continuous probability is not considered first.

In the new order, Bernouilli's theorem is presented as an application of Tchebycheff's inequality;

(b) avoid presenting the concept of mathematical expectation as limited to discrete probabilities, as a result of its being taken before the study of continuous probabilities;

(c) avoid the discussion of normal distribution before the exposition of continuous probability;

(d) introduce the concept of characteristic function which I regard as indispensable in an Advanced Course;

(e) indicate the correspondence between sets of characteristic functions and sets of distribution functions (or cumulative distribution) which seems more comprehensable than central limit tending theorems;

(f) base the concept of normal distribution on a central limit tending theorem, thus demonstrating its importance.

In Part 1, Section 6 I think that t, F and z distributions should be taken after the $x^2$ distribution.

I have no other criticisms to make.

2. Both courses have mainly in view demographic and economic applications, fields outside of my specialization. I could (or I should) give this as a reason for abstaining from stating my opinion. However, even if their authors were not scientists of international reputation, only a cursory reading of the syllabi is enough to show that they have been based on a thorough and up-to-date knowledge of statistics theory and on an extensive acquaintance of practical field problems. Besides, the subjects are logically and gradually presented for teaching purposes. The only suggestion I would like to make is the advantage of including in the project similar courses having in mind the special needs of research workers in

/biological,

biological, psychological and sociological fields.

I do not wish, however, to make this brief statement without some comments on the desirability and possibility of the teaching of these courses in Brazil.

The contacts I have had with the teaching of statistics in Rio led me to believe that most of our courses could be classified in two mutually exclusive categories. In one class there are courses that barely go beyond the four arithmetical operations as a mathematical prerequisite; they emphasize unimportant theoretical and practical points and overlook all developments in the last thirty years of statistical teaching. The courses in the other class are at the opposite pole: they assume a thorough knowledge of mathematics based on set point theory far beyond the real needs of the bulk of prospective statisticians; they emphasize mathematical abstractions and overlook their practical bases and applications. This gives a kind of peculiar U-shaped distribution of statistical courses according to level. The middle-ground courses that one should expect to be the most frequent ones are almost absent. The statistical papers published here and the scant information I have on the teaching of statistics outside Rio de Janeiro lead me to believe that the same conditions prevail throughout Brazil.

It is thus highly desirable that courses like those planned would be taught here. However, I should like to point out some of the difficulties of the problem. For one thing, I don't know of (English) textbooks covering adequately the subject matter, specially the practical applications that are scattered in scientific journals not easily available here. On the other hand, it would not be easy to find teachers here that could teach those courses with the desirable competency. I even do not believe that - outside English speaking and Scandinavian countries - there are very many statisticians able to teach adequately all the subject matter of the courses. If I am not mistaken in these rather bold assumptions, there are great obstacles to be removed before the efficient teaching of these courses in the countries that need them most takes place.

I do not know the kind of co-operation that is planned by the United Nations Organization and the countries for which the courses are intended. If the project is to be developed on a world-wide scale, I would venture to suggest what seems to me the easiest way to reach the intended purpose: the training (in the United States or in England) of a group of prospective teachers, selected (on a merit basis) from the countries included in the project, and the gathering of teaching material (comprising perhaps the writing of one or two textbooks) and its

/translation

translation into foreign languages. It seems to me that this first step (or something like it) is almost a prerequisite for the efficient accomplishment of the project on a world-wide basis.

## 2. CANADA

Comments by the Dominion Statistician, Dominion Bureau of Statistics:

The assistance which such outlines can give to teachers in the field of statistical training and education makes them a useful contribution.

As between the two courses outlined, that by Professor Mahalanobis deals more heavily in mathematics; Dr. Yates assumes either that the student has already covered the mathematical portions or does not require them. I wonder whether such matters as the distribution of the coefficient of correlation are not so difficult as to eliminate many otherwise capable samplers from the course.

The sections on design of sample surveys are in both outlines a satisfactory follow-up of the work of the Sampling Sub-Commission on the standardization of terminology. While it is not to be hoped that text books in statistics will abandon the old terms in any short period, a good syllabus widely distributed will hasten the change. An example of terminological progress is the elimination of the "universe"; "frame" is less pretentious and envisions more specifically the concrete material which is available for drawing the sample.

Dr. Yates specifies the various subject matters in which sampling might be used and suggests covering for each of these the practical aspects given in the preceding section of his outline. This seems to me useful since a discussion of sampling in the abstract is likely to be misleading at certain points and at best will fail to hold the interest of most students.

The Sampling Sub-Commission is to be congratulated on this work. While both of the syllabuses you are circulating are good, we favour Yates' briefer statement; its strictly practical emphasis omitting all but absolutely essential mathematical detail seems very sound.

## 3. CZECHOSLOVAKIA

Comments by F. Egermayer, Institute of Mathematical and Applied Statistics, Prague; J. Janko, Institute of Actuary and Mathematical Statistics, Prague; V. Vlach, head of the Study Department of the State Statistical Office in Prague:

1. "I welcome the efforts of the UN Sub-Commission on Statistical Sampling to broaden scientific sampling methods in statistical surveys. Of the recommendations I consider the syllabus of P. C. Mahalanobis more satisfactory, especially in Parts 2 and 3. The syllabus of Mr. Yates is too general and would

/need, in my

need, in my opinion, more detailed elaboration".

2.    "Both syllabuses contain all the essentials that should be included in a programme of training statisticians who desire to qualify as professional sampling statisticians. Considering that the course is designed to meet the requirements of advanced students who have already had basic training in statistical methods I prefer the syllabus of Yates. His presentation is formulated rather concisely so that it could be expanded as it is in the Mahalanobis syllabus. For example, Part 1 in the Yates syllabus (theoretical background) contains what constitutes in Mahalanobis syllabus Part 1, Items 5, 6, 7, 8. Also a number of other items of the Yates syllabus could be elaborated more in detail as in Mahalanobis syllabus.

3.    "Both syllabuses of training in sampling statistical methods will offer explanations to all the important problems which must be encountered by the statistician who desires to prepare, by himself or in co-operation with others, a sampling survey. The Mahalanobis syllabus is rather extensive since it includes even the basic course on mathematical statistics (law of large numbers, study of frequency data, measures of location, etc.). Yates evidently assumes that the students have acquainted themselves with these basic principles before entering the course of sampling methods. Otherwise both syllabuses are very similar, the programme Yates being formulated more concisely".

### 4.    DENMARK

#### Comments from the Department of Statistics:

The rapid development of the general theory of statistics during the last three decades and the corresponding development of the theory of sampling has created a need for courses in statistics both within and outside the universities. Many statisticians who took their university degrees ten years or more ago feel uncertain of the new theories - even if they in many cases are not essentially different from the older ones but only are expressed in a new terminology. Extraordinary measures must therefore be taken to fill out this gap between theory today and the theory taught not many years ago. In this respect the two syllabuses on sampling are very valuable and useful, as they give a detailed plan for such a course.

The two syllabuses cover approximately the same field, namely the statistical methods which today may be called standard methods from the point of view of theoretical statistics. Consequently, essential differences in opinion with regard

to the proposed collection of methods cannot be expected to occur but the important problem of how to teach these methods may perhaps give rise to divergent opinions.

The courses for statisticians who have finished their university studies and are engaged in practical work must necessarily be somewhat different from the regular university courses. Much emphasis must be laid in the principles of the sampling plans and these must be illustrated by carefully chosen examples of both good and bad plans. In many cases proofs can be omitted if sufficiently good examples are used to illustrate the theorems.

With regard to the regular university courses, however, our opinion is that there must be a reasonable balance between theory and practice and that this object cannot be accomplished by teaching theory without proofs. One of the main reasons for the present unsatisfactory positions of applied statistics (with regard to the use of sampling methods) is the low theoretical level of the statistics courses.

In the syllabus proposed by Professor Mahalanobis it is indicated where theorems are given with and without proofs. In most cases we agree upon this plan but in one important case, namely for the theory of estimation (Part 1, Section 5), we think that proofs are necessary for an advanced course. Also for the proper understanding of the analysis of variance and the design of experiments (Part 1, Sections 7 and 8) proofs are indispensable since these topics are of fundamental importance for the construction of sampling plans. Only if the statistician is familiar with the proofs he will be able to construct sampling plans, which fit his problems and prove that these plans are correct and efficient.

## 5. EGYPT

Comments by the Director General, Statistical Department:

These syllabuses are very useful and I have no comments. I incline to advise for training the syllabus prepared by P. C. Mahalanobis as this is more detailed.

## 6. FINLAND

Comments by the Director-in-Chief, Central Statistical Office:

I consider the matter in question very important, and I also think that the realization of the schemes concerning the university courses and professional training would be very desirable in the form presented in the syllabuses enclosed in your letter. In my opinion there is no greater material differences between the two alternative programmes in question. In Finland the tendency has already been

/towards

towards developing the university instruction at the highest level in the direction presented in the syllabuses, and likewise in the practical research work more and more attention has been paid to the use of adequate sampling methods. However, for the present there are so few specialists in this more and more extensive and manysided developing area that the realization of the scheme is possible only gradually. To the lack of students in this field is also due the fact that there are no signs of any specialization in professional sampling statisticians. But this is a phenomenon which may be perceptible also in other smaller countries.

## 7. INDIA

I. Comments from the Office of the Economic Adviser to the Government of India:

The two syllabuses appear to cover almost identical ground although the details and to a certain extent the emphasis vary in the different sections. Professor Mahalanobis has gone into greater details which should be very useful so far as the operational aspects of the surveys are concerned but at certain places he has expressed the syllabus in terms which may not be quite understandable to everyone concerned, in the absence of a standard terminology on the subject at present. The terms like "internal errors" and "external errors" in the description of interpenetrating network of samples may be omitted at this stage. No doubt references will be made to particular types of surveys but special emphasis should be given in practical work on some of the important surveys conducted by eminent statisticians in different parts of the world and for this purpose the classification adopted by Dr. Yates section (6) of his note will be particularly welcome.

II. Comments jointly from four members of the Calcutta Statistical Association are given below. The members are: Birendranath Ghosh, H. K. Nandi, Purnendu Kumar Bose, Manindra Nath Ghosh:

For convenience of reference we designate the syllabus entitled "Syllabus for an Advanced (Professional) Course in Statistical Sampling" by P. C. Mahalanobis as "Syllabus M" while that entitled "Syllabus for professional statisticians who intend to become experts in sampling census and survey methods" by F. Yates will be referred to as "Syllabus Y".

General Comments. From the title of the two syllabuses it is not quite clear whether the scope of both the syllabuses is meant to be the same. Specially the general term "Statistical Sampling" in Syllabus M comprehends more than what has been included under it and may be a little confusing. So it is considered desirable to clearly delimit the scope of the syllabus in the title as in Syllabus Y.

/Briefly

Briefly stated, the problem of framing a syllabus may be approached from two angles: (i) it may include a detailed and exhaustive listing of the various topics to be covered by the course of study; and (ii) it may indicate only the outlines (without much details) of the course. The first type of syllabus may be recommended as a typical one which, however, cannot be made imperative, because a uniform standard in all its details cannot be enforced irrespective of the considerations like facilities for training, immediate needs of a particular country, possible different approaches to a problem by different people etc. The second type of syllabus is to be preferred on the ground that it allows for natural variation in details and emphasis while indicating the necessary framework of training and hence it can be made imperative.

Comments on Syllabus M. This syllabus attempts to be a detailed one, but unfortunately it appears to be neither exhaustive nor a systematic presentation of the broad outlines. As for example, there is no mention of line sampling or area sampling, though the topic "Surveys for purposes of mapping" in 2 (11) obviously refers to such types of sampling as also the topic of space-field under 2 (1). Again the discussion of the method of selecting sample-unites (2(2)) and the problem of maximizing accuracy or minimizing cost (2(4)) cannot be undertaken without reference to sampling bias, sampling error etc. which appear later under 2(5). The section 2(7) and 2(12) have been considered together. The emphasis on interpenetrating samples (2(8)) seems to be unwarranted. So this syllabus cannot serve the purpose of either of the two types of syllabuses (i) and (ii) discussed in the preceding paragraph.

Comments on Syllabus Y. This syllabus is of the second type discussed above and viewed in this light it is considered to be adequate except for the following minor alterations or additions. If a detailed syllabus of the first type is to be recommended, each item in this syllabus has to be further elaborated.

Under "Theoretical background" we suggest the inclusion of an additional topic on preliminary notions of sampling, the renaming of the topics (a) and (b) under one single head and inclusion of analysis of co-variance in (c). Accordingly the first section will read:

1.  Theoretical background

    (a) Preliminary notions of sampling; subjective and objective sampling; random sampling numbers; choice of sampling units;

    (b) Theory of statistical inference in general; theory of estimation;

/(c) Analysis of

(c)  Analysis of variance and co-variance;

(d)  Principles of the design of experiments.

Include "double sampling" explicitly in 2(c); "maximisation of accuracy" in
3(c); and "optimum structure of sampling unit" and "space-variation in statistical
fields" in 4(f).  Inclusion of "interpenetrating samples" as a distinct type of
sample seems to be unnecessary for reasons stated in connexion of Syllabus M.
The item 4(e) may be changed to "Sampling with varying probabilities (e.g.,
probability proportional to size of unit), from the population and from within
strata".

Under "Practical aspects" there should be included a discussion of the
different types of errors likely to be encountered in survey work.  So the
existing 5(f) may be renamed as "Different types of errors in survey; control of
accuracy of field-work and computation".

Under "Types of surveys" an additional item should be incorporated, say, as
6(f) to be called "Multi-purpose surveys; master sample".  Here it is considered
difficult for the same individual to be conversant with all the types of surveys.
So it should be specified that only some of the types need be covered.

## 8.  IRELAND

Comments by the Director, Central Statistical Office:

My principal difficulty in replying is not in regard to the portions of the
two proposed courses which deal explicitly with sampling methods.  These in content
are practically the same.  They differ mainly in arrangement and emphasis and to
my mind either would form an excellent syllabus for the purpose of expounding
sampling technique.  I find myself in some difficulty, however, in envisaging the
type of person for whom the courses are designed.

The covering letter speaks of "university courses and professional training
at the highest level", while the material in the major portion of both syllabuses
could readily be followed with profit, as Professor Mahalanobis says, by "persons
with considerable practical experience of sampling surveys".  He also says that the
course is designed to qualify persons as 'professional sampling statisticians".
This is what I feel to be an inherent weakness of the approach.  Considered as
part of an integrated scheme for providing statistical training, either of the
suggested courses would be first rate but I fail to appreciate their value in
isolation, as training people ad hoc to become expert in the technique of a single
branch of the science.  To be a "professional sampling statistician" seems to be

/akin to being

akin to being a "death rate actuary". That is something that a well-organized profession such as the actuaries would not tolerate as a qualification and it seems to me to be a retrograde step to endeavour to qualify statisticians professionally in such a limited field. Either a person should be a statistician or not. He may be a specialist but he should, like a medical practitioner, first have obtained his general training, both theoretical and practical. How somebody can claim to be a statistician if he is not a "sampling statistician" is more than I can see.

I believe that Part 1 of Professor Mahalanobis's course - Prerequisites in Statistical Methods - fails inasmuch as it is too wide for someone who is purely an applied statistician and, with its emphasis on the presentation of results without proofs, is too empirical as a theoretical ground work. One point about it that strikes me in particular is that it has omitted any reference to Tests of Normality.

## 9. NORWAY

### Comments from the Central Bureau of Statistics:

The Bureau does not formally provide training facilities, i.e., no regular courses in statistics are conducted. At present there exists, however, a certain type of positions in the Bureau that partly aim at giving possibilities for training in the practical field of statistics for students. These positions would give very little opportunity for attaining practice in the application of sampling methods, as very little work is done yet in this field in our country, and the work that is actually performed is still in the experimental stage.

As to the requirements of adequate sampling courses as outlined separately by Professor Mahalanobis and Dr. Yates, the training might formally be divided into the following four main types:

(a) General background in statistical methods,

(b) Specialized studies in the theory of sample surveys,

(c) Practical training covering problems generally encountered in any survey, whether on a sampling basis or an attempted complete count,

(d) Practical training dealing with problems particular to sample surveys.

Training in the two theoretical parts would in our country primarily rest with the universities.

As to the training on the practical level, we feel that at least in the case of our own country, adequate training facilities cannot be provided except in close

/co-operation with

co-operation with statistical agencies working in that field, without being unduly expensive. Some of the training could be given without connexion with a sample survey. For the more special problems encountered in sample surveys the training would to a large extent depend on the level of applications in this field in the particular countries.

We fully realize the importance of this matter raised by the Sub-Commission on Statistical Sampling and are looking forward to further development of their work with the greatest interest.

In our own case it would also be of great interest if information could be supplied, not only on the requirements on scope and detail of what would be considered adequate courses in sampling by the Sub-Commission, but also on the practical experiences with training of this kind in different countries at present. We would be particularly interested in the part played in training by statistical agencies, and also in details on the degree and form of co-operation in this field between universities and the agencies.

## 10. PUERTO RICO

### Comments by the Chief, Division of Statistics:

The syllabuses submitted by Messrs. Mahalanobis and Yates appear to me good enough for obtaining top level sampling experts.

In my opinion, however, there is also great need for a course of study for the more elementary students in sampling. This is particularly important in Latin America, not so in England or the United States. Only recently with the books of Yates and Deming have didactical material appropriate for wide distribution been prepared. The coming book of Hurwitz and Hansen will contribute greatly to relieve the lack of complete and organized reading material on this field. But for Latin America these books in English have rather limited use. So I think that not only should attention be given to an elementary course in sampling but also to the translation into Spanish of the recent works on the subject.

## 11. SOUTHERN RHODESIA

### Comments by the Director of Census and Statistics, Central African Statistical Office:

The syllabus prepared by Dr. Yates largely follows the order, structure and contents of his text book. This is an excellent practical foundation for the design and organization of sampling inquiries. I feel that for university

training on a high professional level insufficient importance is given to the mathematical and philosophical foundations of sampling technique. There is grave danger lest sampling should be regarded as a simple application of certain standardized rules and principles and lest the mathematical implications and limitations of each method be disregarded. I have frequently encountered the use of these methods and light-hearted references to the term sampling in cases where the methods used violated every sound theoretical principle and with the result that sampling as a useful statistical tool has been stigmatized in the minds of the public.

Examination of the syllabus prepared by Professor Mahalanobis shows that greater weight is given to the theoretical foundation and that the possession of a certain theoretical foundation on which to build training in sampling methods is emphasized. I would welcome Dr. Yates' syllabus if the same emphasis on theory was also stressed and see no reason why the best in both courses should not be abstracted by prefacing Dr. Yates' syllabus with a syllabus for a theoretical background very similar to that laid down by Professor Mahalanobis. Failing this I would prefer the syllabus drawn by Professor Mahalanobis.

## 12. UNITED KINGDOM

The comments from three members of the Royal Statistical Society are given below. The members are: J. Wishart, Faculty of Mathematics, Statistical Laboratory, Cambridge; R.G.D. Allen and M.G. Kendall, London School of Economics, London:

(1) We here note with great interest this development and have no criticisms to offer of the syllabuses by Professor Mahalanobis and Dr. Yates, which are very much alike. I hope the United Nations will be successful in getting something of this sort organized.

(2) We think that the syllabus suggested by Professor Mahalanobis and Dr. Yates is a good one and covers the ground very well. Presumably in the final version Dr. Yates' syllabus will be worked into the main syllabus suggested by Professor Mahalanobis.

There were two points on which we thought that perhaps a little more emphasis might have been laid, although we would not press them unduly.

(a) It is a point for consideration whether in the Mahalanobis syllabus something should be said about time-series other than graduation because of the importance of series of this character in nearly all economic work.

/(b) Dr. Yates

(b) Dr. Yates mentioned systematic selection, and here again perhaps a little more emphasis might be placed on the subject in view of the extent of its practical application.

## 13. UNITED STATES

I. <u>Comments by United States representative on the Statistical Commission:</u>

No attempt is made here to appraise the alternative presentations in the papers prepared by Messrs. Mahalanobis and Yates. There appears to be substantial agreement between the two as to the principal topics to be included in a syllabus. The differences in the two presentations lie mainly in the degree of detail given concerning the topics included and the sequence in which topics are considered. Either could serve, with adaptations to particular circumstances, as a guide to the minimum content of a course in sampling.

The question might well be raised whether a syllabus of any kind would fill a real need in the present stage of development of statistical education. It seems doubtful whether, in the United States at least, the absence in university curricula of adequate provision for "systematic instruction in the theory of sampling or for the application of sampling methods to practical problems" can be attributed in any degree to a need for guidance on the content of university courses and professional training.

A syllabus, in fact, would be of little use without a text book or text books covering the content envisaged by the syllabus, and without qualified teaching personnel to give the course. Anyone equipped to give instruction in the field, and having suitable text books or similar materials available, should be able to design his own syllabus in terms of the text to be used and the requirements and abilities of the students.

There appears to be a real need, on the other hand, for text books suitable for use in courses on the theory and applications of sampling. The Sub-Commission on Statistical Sampling has already taken some steps to stimulate work in this field, and a number of text books are now in preparation or nearing publication in the United States.

II. <u>Comments by members of American Statistical Association:</u>

The comments from eight of the members of the American Statistical Association are given below. The members are: William G. Cochrain, John Hopkins University, Baltimore; Jerome Cornfield, National Institute of Health, Bethesda;

/W. Edward Deming,

W. Edward Deming, U.S. Bureau of Budget, Washington D.C.; W. Duane Evans, U.S.
Bureau of Labour Statistics, Washington D.C.; Raymond J. Jessen, Iowa State College,
Ames; Leslie Kish, Michigan University, Ann Arbor; Frederick F. Stephan and
S.S. Wilks of Princeton University, Princeton.  It may be noted that although the
list of names has been arranged alphabetically, the comments quoted below have been
arranged in accordance with their dates of receipt.

(1)  To comment on the syllabuses prepared by Mahalanobis and by Yates is not an
easy task.  The following comments are forwarded in the hope that similar
appraisals from diverse members of the American Statistical Association can be
used jointly to improve on the excellent job already done.

    Unless I am ignorant of the intended meaning of the "distinction between
random-like and non-random (or patterned) space-fields" the topic belongs under
systematic sampling in 2.2.  The same thing is true of "procedures equivalent to
random selection".  In this connexion I should like to call attention to the
definition in section 3 of the UN recommendations which is amended now to read
"a process is properly described as random if to each unit has been initially
assigned a determinate probability of being selected".  Hence as amended the
qualification of "independent" in describing the probability of selection of units
does not appear.  Therefore systematic selection is also a random selection as
now defined.

    In 2.1 (of Mahalanobis' syllabus) I should like to see introduced the topic
of the intra-class correlation as a measure of the effects of clustering.  Section
2.7 should start out with a discussion of the total error and of its components.
These might be put into the three broad classes of sampling error, errors in
response and errors of non-response.  Section 2.8 might be devoted to the
measurement and control of errors in response.  Another section might be inserted
here devoted to the problems of measuring the non-response error and procedures
for dealing with it.

    Between sections 3.1 and 3.2 I should like to see a section devoted to:
Preparation of sampling materials, maps, listings, master samples; preparation of
sampling instructions; checking and controlling the correspondence of actual
sampling procedure to the requirements of the sample design.

    Between sections 3.5 and 3.6 insert a section concerned with:  Descriptions
of methods and of sources of errors in the report; preparation of generalized

tables of errors.

In section 4b (of Yates' syllabus) include the topic of optimum allocation and its relation to cost factors.

Between 4e and 4f include a section devoted to: Area sampling; the correspondence of elements of the desired population to the population of units of area segments.

I hope that the co-operation of many interested samplers will be useful to the United Nations Sub-Commission in helping them to increase the area and heighten the excellence of their contribution to sampling methodology.

(2) With regard to these or any other syllabus, I am always reminded of a remark from Professor Hotelling. A young woman who had been selected to teach a course in sampling in one of our universities wrote to him to ask what should she teach. His advice, in his kind way, was to "teach what you know". Syllabus or no, people can only teach what they know. A syllabus cannot make a course, but it may help to direct the studies of a teacher. A course, if it is good, is not good because of a syllabus, or just because the teaching is good, but because the teacher has something to teach.

(3) You ask for comments but really I have very little to say. I think that Mahalanobis' outline is quite ambitious and I don't know very many people who could qualify on all points touched upon. Certainly my own interests and feeling of competence are limited to a very small portion of the area that he covers. Yates', on the other hand, is a little closer to the area that I feel more familiar with, and in our own teaching we follow a programme probably much closer to the line laid down by Yates. In fact during the past quarter we used Yates' new book as an alternative text from our notes. My conclusion so far is that Yates' book is probably, for our purposes, more suitable as a reference book than as a text, but this is quite a tentative conclusion.

(4) As you know, such an outline gives only hints as to the content which may be behind it. Certainly enough topics are mentioned to provide pegs for most important notions in sampling theory. Nevertheless, three possible omissions occur to me.

(a) Surely the United Nations is interested primarily in social and economic surveys. In this field particularly, universes are finite and in many cases quite limited in size. It is therefore especially important that all sampling

/problems be

problems be considered within the context of finite populations. There is no indication in the syllabuses that this is the case, and judging from other information it is doubtful whether Yates would have it in mind.

(b) It has been my experience that many students are troubled by the problem of "estimating" the parameters which are required before a specific sample design can be selected. For example, in a stratified design, how to estimate the within and between stratum variances on which size of sample and allocations are to be based? This topic might be given explicit consideration.

(c) Much attention is given to the alternative forms of design, but little to the many types of parameters which may be of interest. So far as the outlines indicate, one might never have a problem which did not simply require estimating a mean value. There should be some consideration of the special problems involved in estimating ratios, percentiles, regressions, distributions, and perhaps other non-moment parameters.

Finally, I feel called on again to protest the perversion of the word "census". If this word has a meaning, it surely refers to a complete count, and so in many ways is an antonym for "sample". I think Yates debases our terminology by referring to "sampling census and survey methods".

I hope these comments will be of some use to you.

(5) (a) The level of prior training assumed by Mahalanobis seems to me to be excessive. In general only the first element in Part I is necessary and can be incorporated in the sampling course itself. The others do not seem to me to be essential to even a fairly thorough theoretical grasp of the subject.

(b) I agree thoroughly with the preceding comments - (4)a - on the finite nature of the population. It seems to me this must be central to the theoretical formulation.

(c) If one attempts to cover in detail the theory behind sampling practice, it soon becomes apparent that a full year is required to develop the following topics:

    (i)   Probability, expectation, some simple distributions.

    (ii)  Simple random sampling for means, differences, ratios, regression coefficients, etc.- determination of sample size.

   (iii)  Stratification and allocations among strata.

    (iv)  Sampling units that include more than one unit of analysis, sub-sampling.

                                            /(v)  Sampling units

(v)  Sampling units of varying sizes - probability proportionate
to size.

(vi)  Special topics - such as systematic sampling and double
sampling.

From this point of view both syllabi appear too long.  The practical aspects in
both syllabi might very well be omitted.  This would cover Parts 3 and 4 of
Mahalanobis' and Parts 5 - 8 of Yates'.  Practice is, of course, at least as
important as theory.  I would therefore, favour a separate course for practice,
which could very well proceed the course in theory and for which the only
prerequisite would be some common sense and possibly a course in statistics.

(d)  I should like to object to the terminology proposed by the United Nation;
Sub-Commission on Sampling and used in both syllabi.  It has both the virtues and
defects of all logical languages, the decisive defect to my mind being that it is
almost wholly unrelated to the terminology actually used by current practitioners.
(6)  Both the syllabuses are good and will be useful to any university teacher
who is planning a course on sample survey methods.  If they are to be regarded as
giving "minimum content" they are a little unrealistic, in that I do not know of
any place at present where such comprehensive training is available.  For this
reason I slightly prefer Yates' outline to that of Mahalanobis, because the
prerequisite training given by the latter seems to contain quite a lot of material
that is not very relevant in the training of a sampling statistician.

Two topics occur to me as deserving specific mention, though both are, I
believe, covered by implication.  In the prerequisite training I would suggest an
item "Sampling with and without replacement from finite populations".  In the
design of sample surveys I would suggest specific mention of "The problem of
non-response".

If a suggestion may be transmitted to the Sub-Commission on Statistical
Sampling, it would be very helpful to have a syllabus for a course that is to be
given not to embryo professional sampling experts, but to people who are in other
fields where sampling is one of the main methods for getting information, and who
wish to know what are the main principles by which they can be guided.  Since
sampling is to a considerable extent a matter of common sense, I believe that such
a course can be given successfully.

(7)  I have read over with great interest the two syllabi for advanced training
in statistical sampling.  They seem to me to be very good outlines and I agree

with the range of topics and emphasis on training in rigorous methods.

My general impression is that it will be difficult to cover all these topics thoroughly unless approximately half a year of full time work is devoted to them. For this reason, I would hesitate to suggest an additional topic but if there is adequate time I believe one should include a serious and systematic study of the problem of presenting the results of sample surveys to persons who are not trained in statistics and, in particular, explaining how such people should utilize estimates of sampling error.

(8) I have read the syllabuses with considerable interest. Since courses leading to professional training in sampling survey have not been very widely introduced into universities, it is my feeling that syllabuses of such courses should be rather detailed. Thus, I am inclined to believe the greater detailed plan prepared by Mahalanobis might be a bit more useful at this stage than that prepared by Yates.

I do not believe either syllabus gives enough consideration to non-sampling errors, although Mahalanobis does mention such errors in (7) and (8) of Part 2. I believe that the principal limitations of sample surveys at the present time lie in the reliability and validity of the information actually obtained from the respondents. There is too much of a tendency to assume that answers made to questions on questionnaires are correct or true or valid. In a course on sampling survey, it is desirable to deal at some length with this matter, discussing types of information which can and types which cannot be reliably elicited from respondents, and dealing with procedures for determining reliability and validity of responses.

In preparing a syllabus on sampling, I should think it would be valuable also to provide a carefully selected list of references on which the course could actually be conducted.

## 14. VENEZUELA

I. Extracts are given below from a letter from the Director General de Estadistica, Ministerio de Fomento, Venezuela: -

Re: Prof. Mahalanobis' syllabus: -

It is evident, as the author himself points out, that this syllabus can apply only in places where there are competent statisticians who have made advanced studies of the subject. This implies that superior and well organized courses in

statistics already exist, and this is not true yet, to say the least, in the majority of Latin American countries....

... When it comes to sampling, the importance of which is fundamentally based on the application of theory, the lack of facilities for practical experience is still made more apparent in the programmes of basic statistical courses intended to impart the complete knowledge of the technique....

... As a (transitory) solution, until the superior course in sampling suggested by Mahalanobis has been established successfully with newly acquired experience and consequently, competence, it would be advisable to suggest the preparation of a programme dealing with less theory but much more with the practical end of sampling. It is evident that this programme should include a certain minimum of theory which would have to be supplemented by intensive practical experience.

Re: Dr. Yates' syllabus: - This programme, which appears to be less advanced in theory than Mahalanobis' programme, gives a more important part to practical experience and contains only one chapter on theoretical background, covering the theory of estimation, of errors of estimates, tests of significance and analysis of variance...

... Granted that the implementation of this programme requires from the part of the students the same statistical knowledge that is necessary to study Yates' treatise mentioned above, it would be quite feasible to adopt it in most South American countries with possibly certain deletions or adaptations as necessary due to local lack of practical experience in conducting sampling surveys or called for to meet the special conditions of a given area.

II. Comments from Dr. Othmar Winkler, Faculty of Social and Economic Sciences, Caracas:

Both projects are to be considered a good basis for special training of professional statisticians.

Yates' programme which provides a condensed treatment of the theoretical and practical aspects of sampling appears to be very adaptable to complete programmes of statistical instruction which so far have not included that subject. It also makes it possible to consider the creation of a complementary university course of say, 2 hours a week, during the semester. The general composition of this programme recommends it for teaching statistics in any field of studies (economics, biology, technology, etc.).

/Mahalanobis' programme

Mahalanobis' programme deals also with sufficient length with the theoretical end, but it emphasizes the practical part of sampling, giving many useful details on preparation, collection and tabulation of the sample with the greatest control and scrutiny possible throughout the various phases of the process. For this reason, this programme seems to be excellent for training sampling experts in statistical offices and as a basis for practical scientific work carried on by a group (as a university seminar). As a means of perfecting statistical knowledge in other sciences such as economics, medicine, etc., this exacting programme seems less to be recommended. Which of those two programmes--both essentially pretty similar--should be recommended, will depend upon the type of instruction with which it is intended to be used.

## APPENDIX A.

## 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 (without proof).

(2) Least square methods; adjustment of observations; accuracy of 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 $\rho = 0$ and $\rho \neq 0$; tests of significance for regression coefficients (with proof); Fisher's z-transformation for r; tests for significance for multiple and partial correlations (without proof).

/(7) Analysis of

(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

(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 results and control over accuracy.

(5) Preparation of the report: presentation of results for general 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

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.)

### 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;

    (b) Methods of collecting the information;

    (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.

------