OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

DISCLAIMER

FAIR USE POLICY

CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

D03533

Distr.
LIMITED

ID/WG.106/11
3 September 1971

ENGLISH
Original: FRENCH

United Nations Industrial Development Organization

Training Workshop in Methods
of Industrial Surveys

Dakar, Senegal, 13 - 24 September 1971

METHODS OF DATA COLLECTION 1/

by

Mr. Georges Godin
Director
National Statistics and Research Institute (INSEE)
Paris, France

1/ The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the secretariat of UNIDO.

id.72-500

Contents

When it has been established what the survey is to cover (activities and units) and what kind of information is to be obtained and collected, a decision must be made on the actual methods to be used in the survey: whether it is to be by correspondence or by interviewer, whether by exhaustive or by sampling methods.

I. Questionnaires and interviews

Two basic methods are used in industrial surveys: questionnaires sent out and returned by post, and questioning by visiting interviewers.

1.1 Advantages of the postal survey

It is the cheapest method.

The possibility of using automatic processes to print addresses, fill envelopes and check replies cuts down the time needed for collection to time taken to fill in the questionnaires.

1.2 Limitations of the postal survey

It presupposes the existence of a reliable and rapid postal service.

It presupposes the existence of a register of the addresses of enterprises and businesses.

The statistical and accounting methods of enterprises have to be harmonised, and the form and content of questionnaires have to be brought in line with these methods; otherwise there is no guarantee that the results will be collected in time or will be of the desired quality.

A system has to be brought into operation for keeping a central check on time-schedules and the quality of replies, involving postal and telephone communications, possibly also visits by interviewers.

1.3 Advantages of the survey by interviewers

It is the only option for a country where the postal service is slow and unreliable or where a register of addresses is not available.

The interviewers can help enterprises to prepare their replies, and can even take over responsibility for going into the firm's records in order to prepare the reply in cases where enterprises have incompatible statistical and accounting systems. This applies to both large and small enterprises.

Qualitative data can be collected on technical, economical or social aspects of growth or obstacles to growth, and through the personal contact which is established the interviewers acquire a knowledge of the industry outside the scope of formal questioning, which is needed when the results are processed.

The interviewers are able to encourage the introduction of standard accounting and statistical systems.

1.4 Drawbacks of the surveys by interviewers

An industrial survey based on the collection of information by an interviewer costs twice as much as a survey based on a questionnaire.

It can take two or three times as long.

1.5 Which method should be chosen?

(a) The postal survey should be given preference in all cases where the necessary conditions are met. This is the normal procedure in countries which have fully developed statistical systems and where the number of enterprises to be questioned precludes the use of interviewers as standard practice.

In particular, surveys made more than once a year (e.g. quarterly) and designed to provide material for short-term appraisals must be carried out by post because of the time factor.

(b) Surveys by interviewer should be chosen if this is the only way of assuring success. If this is the case the survey should be entrusted to highly qualified interviewers. The best use is made of their skill and time if they concentrate their observation on a small number of enterprises in one sector for the necessary length of time rather than spreading their observation over too many enterprises for a limited period. In other words, the choice of a survey by interviewers means giving priority to the quality and quantity of data rather than the cost.

Industrial surveys which have been made up to now in developing countries have been carried out entirely by interviewers.

1.6 Some variants of the two methods are mentioned here for the record;

(a) Interviewers visit enterprises, explain the aims of the survey, and give instructions for filling in questionnaires. Later, at an agreed date, the interviewers return to the enterprises to collect the completed questionnaires, which they look at on the spot so as to make sure that the questions have all been answered properly. At the same time they discuss with the head of the enterprise technical, economic or social aspects which are not quantifiable.

This variant of the survey by interviewers means an increase in the cost of the survey as against a yield that is at the most equal to that of posting as interviewer in the enterprise.

(b) Representatives of enterprises being surveyed are invited to go to a local statistical office where, after explaining the nature of their enterprise, they receive the appropriate questionnaire; they take this away and when they have completed it return it to the local office.

It is necessary to assemble the heads of enterprises in this way so as to prepare enterprises for the survey, and identify the enterprises to be surveyed if there is no other way of drawing up a register. But, at least for the first industrial survey, the drawbacks of postal return and advantages of direct enquiry by interviewers should be borne in mind.

(c) The questionnaires are sent to the enterprises by post and the interviewers fetch the completed forms, for example, a fortnight later. This solution, when practicable, as it generally is for large enterprises, is comparable to a direct survey by interviewers. It has the possible advantage of enabling the enterprise to prepare the documentation in advance.

II. Surveys by sampling

2.1 General

It is possible to reduce the overall cost and the time spent on the survey collecting data on the basis of a carefully selected sample of enterprises, but sampling should not, on these grounds, be considered a universal solution. Samples are more difficult to carry out and check than exhaustive surveys. The difficulties of sampling increase in

proportion to the complexity of the sampling plan, so that if there is a choice between two plans, one of which is more efficient but also more complex than the other, the simpler one is to be preferred. A simple plan for collecting data from all establishments over a certain size but only from a sample of smaller firms is generally preferable, even if, from a theoretical point of view, this is not the optimal sample.

The sampling method also has the disadvantage that it reduces the amount of detailed information which can be used with full confidence. The problem can be seen by considering the probable number of units which would be represented under each column of the resulting statistical tables. It is known that in the first approximation the random error is inversely proportional to the square root of the number of units represented. If this number is small the results have no significance for the column in question. This is the case, for example, when the columns represent industries situated in small territorial sub-divisions.

Finally, the sampling method does not eliminate the usual problems of industrial surveys, but only reduces the number of items to be dealt with.

2.2 Probability sampling and partial surveys

Firstly a distinction must be made between probability sampling surveys and partial surveys. In the first type of survey, each enterprise in the population being considered has a known probability (not zero) of being selected. In a partial survey the population under consideration is only represented by enterprises satisfying certain conditions - generally, those over a certain size.

Furthermore, it should be remembered that a sample survey is not possible without a basis. This basis is the register of enterprises and businesses. If such a basis does not exist, it is better to set about establishing one and postpone the survey until later.

2.3 Stratified probability sampling

In theory, stratification produces groups which are homogeneous (i.e. with low internal variance) but which are as different as possible from each other (i.e. with high between-group variance). It is also known that, even if this aim has not been achieved, the accuracy of results is not diminished by using a stratified sample rather than a Bernoulli one. The usual practice in industrial surveys is to sub-divide the population into one or more groups according to enterprise size. For instance, the

sample may be made up of all enterprises employing more than 5 people and a tenth of the enterprises employing up to 5 people. The sampling ratio can also be varied according to the different activities, so as to take into account the varying concentration of different sectors, but it is preferable for processing purposes to adopt a single sampling ratio in each dimensional category. This method has the great advantage of simplicity. The selection of the samples can be done in a systematic manner by taking, for example, every tenth card in the register of enterprises.

Stratified samples are suitable when the survey is carried out by correspondence. In the case of a survey by interviewers, area or cluster sampling is nearly always more efficient, mainly because travel costs are reduced.

2.4 Area or cluster probability sampling

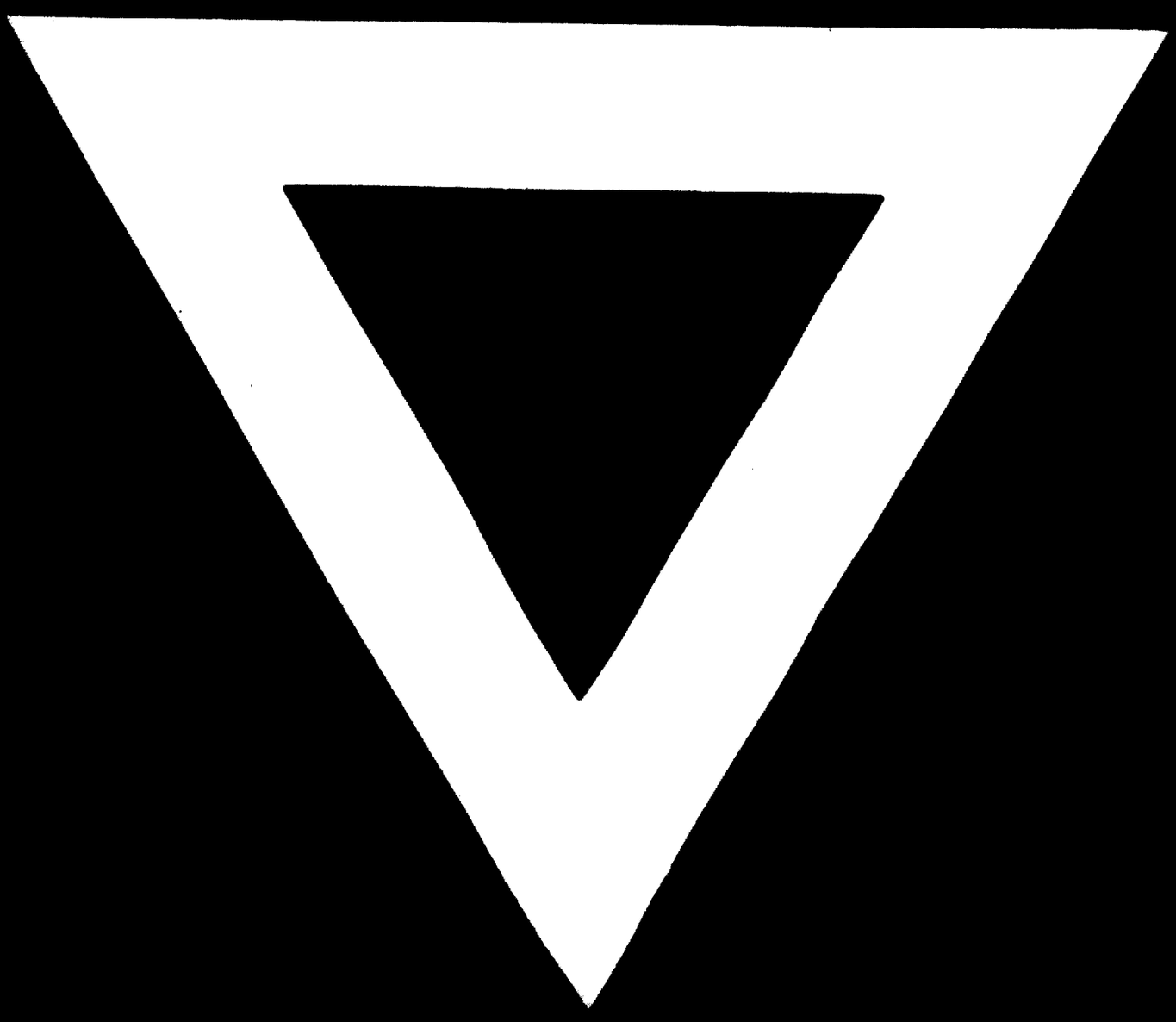
Samples of this type are suitable for surveys covering a large number of small enterprises which are questioned by interviewers. Area samples can produce totals which are significant for the territory as a whole, but not for sub-divisions of it, except perhaps for the districts included in the sample.

In the case of industrial surveys, stratified or area sampling could be used as an additional method. Thus, in the example given, the stratum of enterprises employing more than five people would be covered by an exhaustive survey and the method of area or cluster sampling would be used to make up the sample of enterprises employing up to five people. These "clusters" of enterprises correspond to territorial sub-divisions. The first thing to be done in preparing a simple cluster sampling plan is to draw up a list of districts classified according to their industrial importance. The large areas are all included in the survey, but for the small areas a systematic sample of, for example, one area in ten is made.

In the simplest form of such a plan, all the enterprises within each area are covered by the survey. It is also possible to take a second-stage sample, i.e. a sample within the cluster. The units of the second-stage sample are the enterprises; the total sampling ratio is then the product of the sampling ratio for the areas or districts and the sampling ratio for the enterprises within the areas. Thus one can, as in the previous example, take one district in ten ($t_1 = 1/10$) and all the enterprises within each such district ($t_2 = 1$). One can also take one district in five ($t_1 = 1/5$) and every other enterprise within these districts ($t_2 = 1/2$). The enterprise sampling ratio, which has to be taken into account in processing, is $t_1 \times t_2 = 1/10$.

2.5 Partial surveys

In this type of survey only enterprises over a certain size are questioned. The major draw-back of this method is that it is impossible to determine the standard error in the results. The inclusion in the over-all results of an unknown quantity of data relating to the small enterprises that are supposed to be disregarded is generally the result of making rough judgements "by eye" and these data are all the more unreliable if there is little in the way of outside information, all the more serious if they relate to sectors with low concentration.



4. 12. 73