OCCASION

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

CONTACT

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

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

D03168

United Nations Industrial Development Organization

Training Workshop in Methods of Industrial Surveys

Dakar, Senegal, 13-24 September 1971

THE CONSIDERATION OF CONSTRAINTS OF COST, TIME LIMITS

AND THE RELIABILITY OF DATA IN ORGANIZATION AND

METHODS WITH REGARD TO INDUSTRIAL SURVEYS [1/]

by

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

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

id.71-496

CONTENTS

To meet the needs of programming and industrial development control, and also for the sake of its current industrial policy, every State sets up an official system of industrial statistics. This entails official organization of the various agents (government services and business concerns) responsible for putting into operation methods for the efficient production and utilization of a complete and coherent body of basic industrial statistics. Industrial statistics are called "basic" if they are essential for carrying out retrospective or prospective analyses and syntheses undertaken by the public authorities at intervals of more than a year, a year, or less than a year. The body of basic statistics must be complete (quantitative aspect) and coherent (qualitative aspect).

To achieve efficient organization and methods of production of the body of basic industrial statistics it is necessary to take into account the cost of the data, delays in their production, and their reliability. The field of responsibility of the director of an industrial survey programme lies in decisions made in the light of these three constraints: cost, delays and reliability.

I. Cost, in relation to organization and methods

Expenditure on the collection of exact figures is the main component in the cost of an industrial survey:

1.1 The effect on organization

As regards organization, it will therefore be incumbent on an Institute of Statistics:

(a) To provide the legal basis for industrial surveys in implementation of a law relating to statistics in general and dealing with compulsion, co-ordination and secrecy in statistical matters.

(b) To set up a Co-ordination Committee for industrial surveys, composed more particularly of representatives of the government and of business enterprises. The functions of this Committee will be:

- To examine survey programmes;
- To advise on appropriate methods of data collection;
- To give the necessary support regarding the information aspects of the surveys, the aims of the surveys, the legal provisions, and the way the surveys are carried out.

(c) To organize co-ordination on the basis of the following rules:

- To make complementary use of statistical inquiries and mandatory replies to official questionnaires;

- To decentralize data collection so as to facilitate relations between enterprises and the inquiry service, making sure that, as regards statistics, an enterprise is in touch with only one inquiry service;

- To take advantage of the economies of scale arising from concentration of the means of digitizing, punching, and processing.

(d) To incorporate checks that might lead to corrective action in enterprises, at the stage of data collection and not at the stage of processing. To that end, it is preferable that enterprises should do their own checking, after being fully informed of the checking plan to which their declarations are to be subjected.

1.2 The effect on methods

(a) An inquiry should not be carried out by an investigator unless the conditions for carrying out a proper postal inquiry are not met. This will usually be the case in countries where a system of statistics is in process of being set up It should be noted that if the statistical and accounting practices of enterprises are not harmonized or compatible with the inquiry, data collection by correspondence can still be possible, provided that information on the questionnaires is communicated to enterprises at the beginning of the business year so that they can bring their own data recording systems into line with the inquiry.

(b) The form and substance of the questionnaire should be designed in such a way that enterprises can consider it seriously and that it can serve as a single basic document for processing before punching. And, of course, ask only for data whose synthesis is programmed either in statistical tables or as parameters of a growth model, etc.

(o) The flow of information should be channelled throughout the government service.

(d) The method of simple stratified sampling should be used (detailed investigation of the strata of large undertakings, strata of small enterprises investigated by sampling) in the case of inquiry by correspondence. For an inquiry by investigator, the sample of small industries should be composed by two-stage area sampling.

(e) Expenditure should be planned. This requires viewing the operations as a whole, beginning with the "output" (the tables to be produced) and then going on to examine the "input" (the contents of questionnaires) that is necessary and adequate, and then adjusting the processes required to achieve the output by processing the input. This is tantamount to preparing production on the industrial model (design, research department, tests, putting into production). One of the advantages of this preliminary analysis is that it makes it possible to estimate operation times and thus to prepare and plan expenditure estimates with a good probability of not deviating more than 10 - 20 per cent from actual costs.

II. The time factor, in relation to organization and methods

Statistics are more or less useful according to the relative recency of the period they concern. This usefulness rapidly diminishes with time, to the point that the results of an inquiry available only a few years after it has been carried out are almost useless. Excessive delays, sometimes of as much as three or four years, required to correct the consequences of bad organization of processing, are more to be feared than slowness in data collection, which it is always possible to take energetic steps to rectify. Costs and delays are not independent of one another: though gaining time may entail additional expense, it can likewise be productive of economy; on the other hand, loss of time is always accompanied by a disproportionate loss of money. This shows the importance that the senior staff of statistical institutes should attach to considering the constraint of the time factor with regard to organization and methods.

2.1 The effect on organization

(a) A team of high-level staff, economists, statisticians, business management engineers, information specialists should be set up who for at least a year will be designing and studying the industrial survey in its multiple aspects, in each of its phases, down to the smallest details, examining foreign experience and never losing touch with reality. The tests carried out at the end of this year of study will mark the beginning of the preparation of the survey proper, which should not, as has often happened, appear to be a series of improvisations.

(b) All necessary measures should be taken to limit delays, for example:

- Provide rapid liaison arrangements and thus ensure a rapid reaction in face of unexpected contingencies, by decentralizing data collection and centralizing processing.

- Not to allow the manual entry of questionnaire data on intermediate forms in the course of processing, except in the case of rapid provisional processing.

(c) Before the questionnaire is drawn up, the requester should be asked to furnish information about the use he is going to make of the data. It is useless to pay special attention to the time factor and, in general, it is disadvantageous to carry out the survey, if the tables are to be left lying in cupboards because there are no competent persons and methods to analyse and make use of them.

2.2 The effect on methods

(a) The questionnaire should be suitable in relation to the possibilities for its completion, and therefore to the size of the enterprises concerned. The average duration of inquiry by enterprise, that is to say, the time the enterprise takes to reply, plus the time taken for data collection, plus the time taken for processing, is proportional to the following:

Number of questionnaire items
Size of the enterprise

It will be immediately inferred that, for small enterprises, it is necessary to plan a simplified questionnaire including the basic items of the general questionnaire which will permit standardized processing for the whole field.

The second consequence is that, in any case, the number of questions should be reduced to the sole data necessary and sufficient for securing a clearly defined "output".

(b) Tests should be carried out and their results analysed. The first aim of the tests will be to test methods of data collection, checking, and processing. Their effect will be, of course, to start the preparation stage of the survey proper, which could last six months in order to take account of all the lessons learned from the tests.

(c) A detailed plan should be established, and the progress of operations kept under review. Presentation, in the form of a Gantt progress chart, for example, of the operations to be carried out simultaneously or successively, reflects the results of the

integrated "design" phase. Keeping a check on progress includes the use of progress
indicators such as rates of response and representativeness of the data collected,
established on the basis of lists drawn up when the survey was launched, the percentage
of punch cards created, etc., and the application of a reminder strategy for data collec-
tion, including, if necessary, approaches by letter, by telephone, and by the investi-
gator, the serving of formal notices, and the filing of reports on those who have not
replied, which sets in motion the procedure of official penalties.

(d) Arrangements should be made for the production of essential provisional
results, by means of counting and summarizing done by hand during the collecting stage
(if the survey is carried out by investigators), and at the time when the questionnaires
are being prepared for processing, (if the inquiry is carried out by correspondence).
In the organizational and methodological conditions described above, it should be possible
to adhere to a calendar of the following type for the production of industrial survey
results.

Time pattern of surveys	Provisional data	Final results (J = reference period)
Every several years (i.e.: R.I)	middle of J + 1	middle of J + 2
Annual	middle of J + 1	end of J + 1
Monthly	middle of J + 1	end of J + 1

III. The reliability of data, in relation to organization and methods

The task of statisticians is to produce data that are sufficiently reliable, at a
minimum of cost, and within present time limits. This is particularly true of the
specialists responsible for industrial surveys. The degree of reliability of the results
of an industrial survey will depend on the relative importance of three types of errors:
random errors, systematic errors and errors in replies or processing.

3.1 The three types of errors

(a) Random errors arise in the main from employment of the sampling method.
An estimate arrived at by extending to small industries as a whole a measurement made on
a sample only is fraught with uncertainty: in that case one speaks of the variance of
an estimate. This variance can be calculated and depends in particular on the size and
method of composition of the sample.

(b) **Systematic errors** spring mainly:

- From faulty composition of the sample, due sometimes to an unskilful sampling plan but more usually to shortcomings in the basic index;

- From a considerable number of "non-replies" and the bias introduced by corrections;

- From the processing of simplified declarations supplied by small enterprises.

(c) **Errors in replies or processing**, for instance: total of purchases declared with taxes included, whereas these should be given without tax, entry of a production activity in the wrong heading of a pre-coded questionnaire, improper recording of a number with its decimals when these should have been left out, punching error through faulty centring, failure to remain within the field, etc. The manual and automatic checking positions are designed to detect these errors and lead to their correction.

3.2 Relative degrees of importance, and consequences

(a) Most importance should be attached to errors of the third type. Experience shows that these play the greatest part in determining the degree of confidence that can be placed in the results of an industrial survey. With manual control, the attention paid to different returns must vary in direct proportion to the size of the enterprise, not overlooking the fact, however, that the importance attaching to a return from a small enterprise will depend on the weight given to it in the sampling.

(b) Be careful in using the method of "proportional increase". If industrial surveys are to be reliable, there must be a directory of enterprises and concerns. If this is the case, systematic errors remaining after checking will be due to the method of processing the simplified returns drawn up by small enterprises. The simplified questionnaire in fact comprises only the main totals, whereas the constituent elements are obtained by means of standard questionnaires. Processing frequently consists of applying the rule of three to increase the results in the distribution obtained by processing according to the categories of the standard questionnaires, in order to make this distribution consistent with the grand total, taking into consideration the results from simplified questionnaire summaries. This method introduces an error if the proportions are strongly influenced by the size. When it is applied, this possible effect should therefore be taken into consideration.

(c) <u>Random error depends on the sampling plan</u>. Simplicity has been preferred to a possible gain in variance that would not greatly influence the reliability of the results as a whole. Indeed, on account of the relatively insignificant weight (in production, for instance) of the small industries investigated by sampling, a considerable error in an estimate will probably have an only slight effect on the total, as the following table shows:

	Actual production	Estimated production	Percentage of error
Large enterprises	98	98	–
Small enterprises	2	3	+ 50
	100	101	+ 1

In this example, although the production of small enterprises has been overestimated by 50 per cent, the effect of this error on the total is only 1 per cent. That percentage is within the acceptable limits.
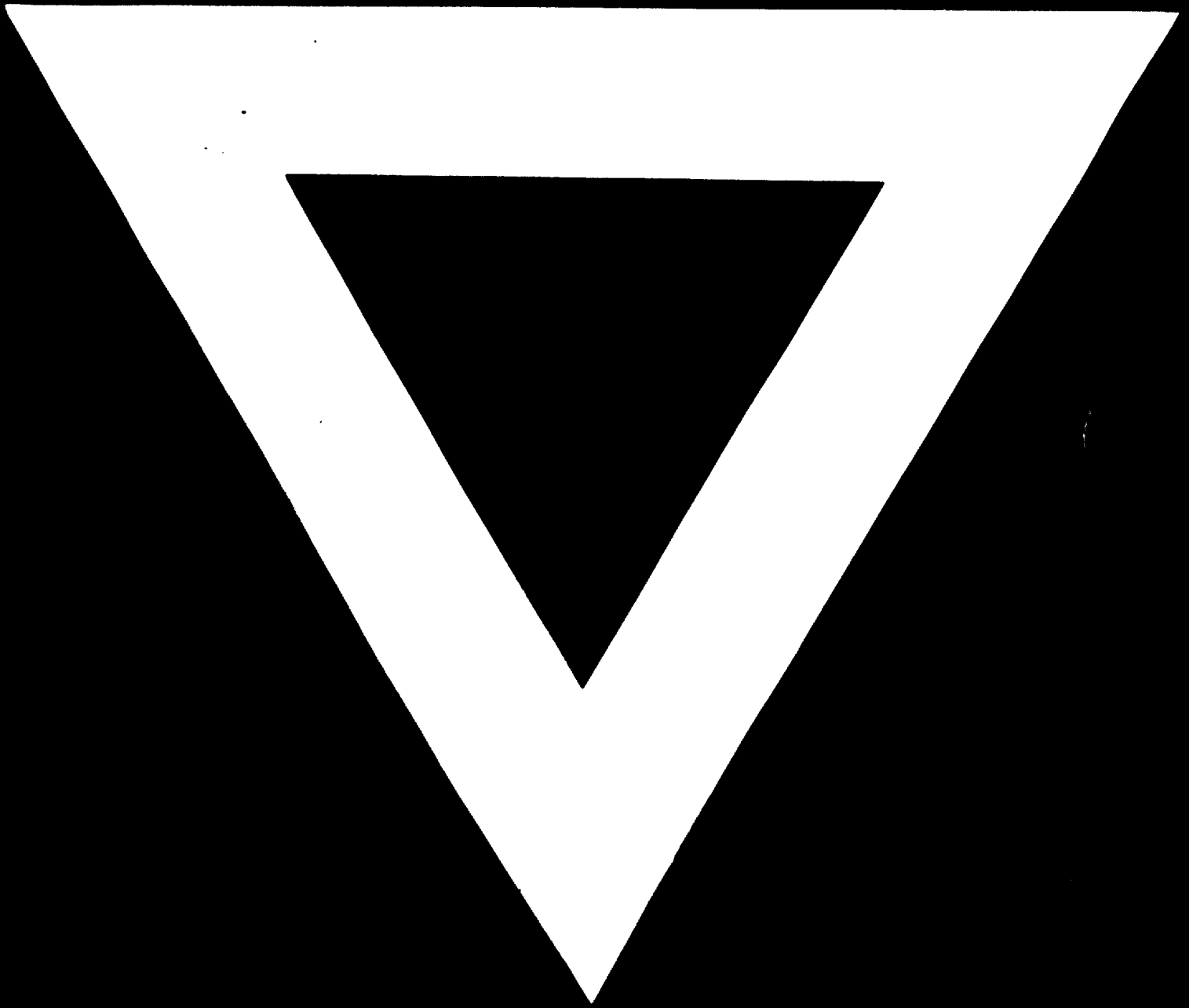
(d) In order to detect <u>errors in processing</u>, the tabulated data should be checked by subjecting them to the manual checking plan for individual data, and to comparison with outside data.

*

* *

Suitably chosen and strictly applied principles of organisation and methods for the preparation of industrial surveys, the collection, checking and processing of data guarantee reliability in the results of a survey. However, if, despite the precautions taken, the statistician cannot answer for the quality of an item of data, it is his duty not to disseminate it.

7.8.73