



Co-funded by the
Erasmus+ Programme
of the European Union



LAB-MOVIE

Labour Market Observatory in Vietnam universities

WP 2

Transfer of knowledge and methodology to analyse the labour market

Outcome 2.1

Development of educational materials

THE STATISTICAL SURVEY

***KA2 - COOPERATION FOR INNOVATION AND THE EXCHANGE OF GOOD PRACTICES
PROJECT NUMBER: 609653-EPP-1-2019-1-IT-EPPKA2-CBHE-JP Lab Movie***

“The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.”

THE STATISTICAL SURVEY

The topics that will be examined:

1. Definition.
2. Steps for the implementation of a statistical survey.
3. The design of a statistical survey:
 - 3.1. Definition of the aims of the research.
 - 3.2. Definition of the target population, the variables of interest and the information already known about it.
 - 3.3. Analysis of available financial, human, technical and material resources.
 - 3.4. Definition of timing and accuracy.
 - 3.5. Search or construction of the list.
 - 3.6. Definition of the type of survey.
 - 3.7. Definition of the survey tools.
 - 3.8. Definition of the survey techniques.

1. Definition

A **statistical survey** is a collection of data carried out following a complex production process that aims to produce statistical information, that is the production of quantitative or qualitative summary descriptions of a collective phenomenon.

The object of observation of each individual phenomenon that constitutes the collective phenomenon is called **statistical unit**.

The collective phenomenon is studied through the observation or measurement of one or more characteristics of the statistical units.

Each characteristic is called a **statistical attribute or variable** and the domain in which the attribute varies is called **value**.

The set of statistical units constitutes the **population**.

2. Steps for the implementation of a statistical survey

1. Design.
2. Survey piloting.
3. Survey carry out (data collection, encoding and recording on computer support).
4. Data processing (revision and processing of collected data).
5. Presentation and dissemination of results (drafting of a final report and its dissemination).

3. The design of a statistical survey

1. Definition of the aims of the research.
2. Definition of the target population, the variables of interest and the information already known about it.

3. Analysis of available financial, human, technical and material resources.
4. Definition of timing and accuracy.
5. Finding or construction of the list.
6. Definition of the type of survey.
7. Definition of the survey tools.
8. Definition of survey techniques.
9. Definition of the sampling design. *(It will be developed separately)*
10. Definition of the questionnaire and its implementation on decided support. *(It will be developed separately)*

NOTE: The definition of the type of survey (including any sampling design and units selection) and the definition of the survey techniques constitute the **survey design**.

3.1. Definition of the aims of the research

First of all you have to:

- Precisely identify the information to be searched for, distinguishing exactly what is of interest from what is not, bearing in mind that the wider the range of topics covered is, the greater the complexity to be tackled on a conceptual and operational level will be.
- Exactly circumscribe the territory and the target period of the investigation.

It will be necessary, therefore, to decide the aims of the research:

- **Descriptive/exploratory research** (the researcher's interest is limited to the description of the phenomenon without formulating and verifying hypotheses on what has been observed).
- **Explanatory/analytical research** (the researcher's interest is not limited to the description of the phenomenon, but searches for possible relations between variables).

In addition, it will be necessary to decide which type of study is the most suitable for the purpose of the research:

- **Transversal study** (they are studies that are based on a single measurement over time, the aim is to detect the characteristics possessed by the target population in a single instant of time or referred to a period - they are studies that photograph a population in a given instant or in a given period)
- **Repeated study** (these are studies, generally transversal, which are repeated periodically over time).
- **Longitudinal study** (these are studies that follow a population over time in order to study the changes that have occurred over time).

3.2. Definition of the target population, the variables of interest and the information already known about it

3.2.1. The target population

The **target (or reference) population** is the set of statistical units to which the survey results will refer. It is, therefore, the population with reference to which we are interested in obtaining information.

The **statistical population** is the population with reference to which we effectively carry out the survey, since we have a list of the units that make it up. It is, therefore, the set of units to be subject to analysis.

Target population and statistical population should coincide, but in reality this is not always possible. The statistical population should be as close as possible to the target population.

The population should be defined:

- In content.
- In time.
- In space.

Eligibility conditions (the characteristics that determine the inclusion, or exclusion, of statistical units in the population) must be specified.

3.2.2. Variables of interest

The **variables** are characteristics, usually elementary, referring to the statistical units, therefore, characteristics of the statistical population.

The **values (set of responses)** are, instead, the possible outcomes of a measurement, i.e. the ways in which a variable can vary.

The **classification of variables** is the set of responses that a variable has (at the level of maximum detail).

The definition of variables should proceed through a progressive identification and refinement of the phenomenon of interest in its components, until the salient aspects are identified.

NOTE: The variables and their classifications will be studied in depth apart when we'll talk about data processing.

3.2.3. Recognizement of already known information on the population

Acquire existing knowledge on the subject to better define the objectives of the survey and develop the most appropriate methodology.

The information channels from which information can be drawn can be divided into five sources:

1. research experience gained both nationally and internationally,
2. bibliographical documentation,
3. statistical documentation (official sources),
4. surveys of users and/or professionals in the sector (interviews with opinion leaders),
5. previous similar surveys.

3.3. Analysis of available financial, human, technical and material resources

At this step a recognition must be carried out to understand the availability of:

- financial resources (amount of funds available to carry out the survey),
- human resources (researchers, computer scientists, specialised technicians, interviewers, telephone operators, etc.),
- technical and material resources (software, computers, printers, fax machines, telephones, internet connection, etc.).

NOTE: These are three aspects that heavily influence subsequent choices.

3.4. Definition of timing and accuracy

As regards the **timing** of the survey, the required timeliness in obtaining the results must be taken into account.

Timeliness can be induced both by the urgency of the information and the speed with which the observed phenomenon evolves over time.

Time and cost, in addition to influencing each other, strongly influence the quality of the information produced.

The **accuracy** of a survey is determined by non-sampling error.

The **non-sampling error** (accuracy) depends on several factors that are defined during the design phase. In particular, it depends on the quality of the list, the instruments and detection techniques that will be adopted, but it also depends on the data processing techniques used and the sample size.

In fact, the non-sampling error is directly proportional to the sample size and is positively influenced by the complexity of the survey.

Accurate surveys require the adoption of tools and techniques that can be expensive both in terms of time and the financial, human, technical and material resources required.

NOTE: The non-sampling error is determined by:

- Coverage error (depends on the quality of the list).
- Total non-response error (depends on whether statistical units participate in the survey).
- Partial non-answer error (depends on whether all questions are answered or not).
- Measurement error (depends on the tool used to detect/measure information/variables).
- Processing error (depends on data processing and processing techniques).

3.5. Search or construction of the list

At this step it is necessary to find or build the list of all the statistical units that make up the statistical population.

The list can be purchased (e.g.: lists of companies can be purchased through the chamber of commerce) or created by consulting specific portals (e.g.: in the case of companies, by consulting the portals of trade associations and specific portals that collect lists of companies).

The list should contain the identification information of the individual units and possible contact channels.

NOTE: Building the list can be a long and laborious job.

3.6. Definition of the type of survey

At this step, it is necessary to define which type of survey is best suited to collect information and to produce the relevant statistics.

The first step is to decide whether to proceed with a:

- **Direct survey** (information is collected by directly contacting the individual units making up the population).
- **Indirect survey** (information is collected through administrative data or data from official sources).

3.6.1. Direct survey

Advantages:

- Match between the objectives of the survey and the information collected.
- Possibility to contact the respondent.
- Low non-response rate (persuasion to cooperate).
- Possibility to provide explanations.

Disadvantages:

- Higher survey costs.
- Reduced timeliness.
- Greater statistical error/disturbance.
- Higher detection errors.
- Need for a widespread organization in the territory.
- Greater risk of conditioning.

3.6.2. Indirect survey

Advantages:

- Reduce the cost of detection.
- Reduces the statistical burden.

Disadvantages:

- The population to which the information refers may not coincide with that of administrative data collection (coverage error).
- The concepts, definitions and classifications adopted in administrative data collection may not coincide with those of statistical research.
- The quality with which information is collected in administrative data may not be sufficient for the objectives of the research.
- The availability of data in time for statistical information needs is not always guaranteed.

If you have chosen direct investigation, you will have to decide whether to proceed with one:

- **Census** (survey conducted on all the population units).
- **Sample survey** (survey conducted among a subset of population units, selected to be representative of the whole population).

3.6.3. Census

Advantages:

- Provides a real measure of the population (not affected by sample error).
- The information collected can be used as a basis for further studies.
- It provides the basis for selecting samples.
- The information can be linked to data from other sources (in particular administrative sources).
- High detail of analysis.

Disadvantages:

- Complicated from an organisational point of view.
- The release of information takes a very long time.
- High costs both in terms of resources.
- Considerable non-sampling errors.

A census should be carried out if the target population consists of a few units (e.g. survey of industrial districts in a certain region).

From a purely theoretical point of view, with a census it is possible to obtain perfect estimates of what is of interest. In practice, however, it is easy to see that this is in fact impossible. The problems are various, in particular due to the financial and time costs involved and related to the quality of the collected data (usually lower than that of data collected through a sample survey).

3.6.4. Sample survey

Advantages:

- Abundance and depth of content.
- Speed of execution and timeliness of results.
- More accurate detection (non-sampling error).
- Design flexibility.
- Lower cost.
- Fewer respondents involved.
- Necessary when the sample units are destined to be destroyed (e.g.: lamp life).
- Practicable in developing countries.

Disadvantages:

- Not suitable for detailed analysis, non significant estimates for smaller aggregates.
- Coverage is often incomplete.
- Often not considered valid for policy-maker decision-making purposes.
- Invasive for interviewers.
- Sample error (accuracy).

By reducing the burden of the survey in terms of time, human and financial resources, the sample survey allows greater attention to be paid to the quality of the data collected (accuracy).

On the other hand it is to consider the extension (inference) of the results obtained which are directly related to the sample, and "only" inductively (in the case of probabilistic samples) related to the population of interest.

3.7. Definition of the survey tools

At this step it is necessary to decide the tool to collect the information:

- Interview with privileged witnesses (experts of a certain subject).
- Questionnaire.

Interview with privileged witnesses

- The **privileged witnesses** are people who are recognised as having a particular knowledge and/or competence on the target subject and a particular ability to interpret the facts.
- They are interviewed using sophisticated techniques to obtain points of view that are analysed by an experienced researcher.
- The result is a qualitative, and sometimes even quantitative, representation of the phenomenon.

NOTE: this survey tool will be analyzed in depth separately.

3.8. Definition of the survey techniques

At this step (having decided to collect the information through a questionnaire) it is necessary to decide how the statistical units should be contacted and how the information should be collected. The choice of the survey technique is a fundamental aspect of the planning and therefore of the execution of the survey, being closely linked, for example, to the training of personnel to be used on the field, costs and expected time.

First of all it will be necessary to decide whether to proceed with a:

- Questionnaire filled out by the respondent.
- Questionnaire filled out by the interviewer.

3.8.1. The role of interviewers

- In the administration of a questionnaire, the interviewer conditions the outcome of the survey by interacting with the interviewee both with his physical presence (face to face) and, albeit to a lesser extent, with his way of conducting the interview (by telephone).
- Interviewers must be discreet but at the same time get reliably all the information they need.
- Interviewers should be able to give clarifications to respondents and to induce them to answer as correctly and truthfully as possible (not rushing the answer, etc.).
- Interviewers should be professionals, and they should be specially trained on the specific topics of the survey to which they have to take part.

3.8.2. Questionnaire filled out by the respondent

In the case of questionnaires filled out by the respondent the survey techniques are:

- Hand-delivered.
- Delivered by mail..
- E-mail/Web (CAWI) delivered.

The questionnaire is delivered by hand or sent by mail, fax or e-mail to the respondent who fills it out, independently, by answering the questions.

The completed questionnaire is hand-delivered or sent back by mail, fax or e-mail.

Sending, filling in and returning the questionnaire can be done with the help of a computer and internet (e-mail) in this case we are talking about CAWI (Computer Assisted Web Interviewing) survey.

Main advantages of hand-delivered, mail delivered, web delivered survey

- Low costs (no staff must be hired or trained to carry out interviews) especially via e-mail/web.
- A smaller organisation is required.
- The questionnaire tends to be long because it must contain instructions on how to fill it in.
- Low risk of conditioning, the absence of the interviewer eliminates possible distortions due to the interviewer (clothing, tone of voice, gender are all factors that can influence the interviewee).
- Suitable for asking questions on sensitive topics (due to the absence of an interviewer).
- The respondent is given the opportunity to consult any documents (e.g. organisation chart) calmly before answering.
- You can reach every place by mail at the same cost (you can also reach people without a telephone or internet and you do not have to pay for the travel of the interviewer).
- With CAWI, the data-entry is also performed at the same time as the data collection.

Main disadvantages of hand-delivered, mail delivered, web delivered survey

- Long and uncertain timings for answers' collection.
- Impossibility to certainly identify the respondent (the questionnaire could be filled in by another person or the respondent could get someone to suggest the answers).
- Strong self-selection of respondents (only particularly motivated people answer). The fact that not everyone answers brings the problem of understanding who answered and who did not and what are the (different) characteristics of the two groups. It is important to distinguish between non-responses and wrong addresses (unreachable person).
- Only a small percentage of people respond (usually around 20/30%). To overcome at least part of this problem, you can send one or two letters/e-mails or make one or more reminder calls, reiterating the importance and objectives of the research. At the same time you can send the questionnaire again.
- Non-verbal behaviour (attitudes) cannot be observed and therefore the respondent's level of interest cannot be detected.
- It is more difficult to help the respondents in understanding the questions and filling in the questionnaire (hence the importance of the questionnaire graphics).
- It is impossible to check that the order of the questions is followed.
- Many questions may remain unanswered.
- The date of completion is not certain.
- By e-mail/web it is not possible to reach those who do not have an e-mail (however, this information may be difficult to find in some cases, there is no equivalent of a telephone book). Those who do not have an e-mail or who are excluded from lists often have special characteristics.

3.8.3. Questionnaire filled out by interviewer

In the case of questionnaires filled out by the interviewer the survey techniques are:

- Face to face (CAPI).
- Telephone (CATI).

Face to Face

The interview is conducted in person by an interviewer who reads the questions and the answer options in the exact order and with the same language of the questionnaire.

The interview can take place with the help of a computer, in this case we speak of a CAPI (Computer Assisted Personal Interviewing) survey.

Main advantages of the face-to-face interview

- Greater opportunity to contact and convince the respondent to cooperate.
- It greatly facilitates the identification of the respondent.
- Increased interviewer-respondent interaction (it is possible to explain the meaning of the questions and the correct way to answer them, so more complex questions can be asked).
- Facilitates long interviews.
- Allows the use of graphic supports such as so-called "tags".
- If the interview is CAPI, the data-entry is also performed at the same time as the data collection.

Main disadvantages of the Face-to-face Interview

- Very expensive, particularly if the places to be reached are far away or the territorial boundaries of the research are wide.
- It is time consuming.
- It requires a detailed organization on the territory.
- Difficulty in controlling the work of the staff.
- There is a risk of influencing the answers. the way the interviewer conducts the interview is particularly important. A good interviewer must ask the questions exactly as they are formulated on the questionnaire and in the established order; he/she must not provide clarification on the meaning of the questions that may alter their meaning, he/she must not try to anticipate the answer or comment on it; he/she must not give the impression that it is an exam and that there are right or wrong answers.
- Possible difficulty in finding the respondent. In some cases, you may be forced to work in non-canonical hours (such as late in the evening or very early in the morning). In other cases, while finding the person, it may be difficult to be welcomed into the home (elderly people alone).

Telephone

The interview is conducted on the phone by an interviewer who reads the questions and answer options in the exact order and in the same language used in the questionnaire.

The interview can be carried out with the help of a computer, in this case we speak of a CATI (Computer Assisted Telephone Interviewing) survey.

Main advantages of the telephone interview

- Significantly less expensive than the face-to-face interview. Statistical units can also be spread over a large area. The lower costs make it possible to make several attempts to trace respondents. Lower training costs for interviewers.
- Very short data collection time. In addition, if the interview is CATI at the same time as the data collection, data-entry is also performed.
- No organisation on the territory is needed.
- Possibility to easily control the work of the interviewers.
- Facilitates contact with people who are not at home during canonical hours.
- Low risk of bias in the answers due to the absence of distortions due to the physical appearance, facial expressions and the interviewer's gestures.
- Greater opportunity to ask questions on sensitive topics.

Main disadvantages of the telephone interview

- Unable to reach those who are not in the phone book. Those who do not have a phone or who are excluded from the lists often have special characteristics. In addition, the spread of mobile telephony that does not have telephone books should not be underestimated. Random phone number generation techniques can be used to try to overcome such problems.
- The identification of the respondent is not certain.
- It is more difficult to stimulate the respondent's collaboration.
- More difficulty in explaining the meaning of the questions and how to answer them correctly.
- It is very difficult to conduct long interviews.
- The options of closed answers must not be too many, otherwise they can be forgotten.
- The development of telemarketing can lead to the belief that the aim of the interview is to sell some products or services.
- It does not allow the use of graphic media such as photos or tags.
- Non-verbal behaviour (attitudes) cannot be observed and therefore the respondent's level of interest cannot be detected.

3.8.4. Mixed techniques

They are used when only one survey technique does not perform well in all practical situations (e.g.: Telephone or web survey + face-to-face survey for those who do not have a telephone or internet).