How to prepare a questionnaire or a form

Ulisse Di Corpo

Data is the row material of statistical analyses, but not all data is suitable for statistical analyses; only data which is numerical and gathered in a systematic way can be used:

- *Coded data*: only quantitative information and information which can be transformed into numbers can be treated statistically. All other information is of little use.
- *Systematic gathering of data*: statistical analyses require data which has been gathered systematically. When preparing a form or a questionnaire it is therefore necessary to check if data can be obtained systematically.

Coded data.

Statistical databases use extensively coded information, information which can be translated into a number. The more we want to use statistics the more we need coded data. For example:

Commercial database	Statistical database

Name:
 Address:
 Town:
 	
 Marital Status: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
 	
 Instruction: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _
	Name:
	 Address:
	 Town:
	 uses the official list of towns
	 Marital status:
	  1. *un-married*
	  2. *married*
	  3. *separated*
	  4. *widow*
	 Instruction:
	  1. *Illiterate*
	  2. *Elementary diploma*
	  3. *High school diploma*
	  4. *University degree*

In all the fields, except for *Name* and *Address,* a code is associated to the information; for example code 2 is associated to married, code 4 to university degree. Each field refers to a list of labels (check list). When paper archives are used the differenced is secondary, but when archives are electronic advantages are immediate:

1. *Dimensions*: the physical dimension of the archive is reduced; while a number (from 0 to 255) requires one byte of memory, a text requires as many bytes as the number of characters which have been reserved, for example MARR (Married) requires 4 bytes.
2. *Errors reduction*: the risk of committing errors is reduced, the program accepts only labels which are in the list, it is not possible to enter a different label or a code outside the range of codes listed;
3. *It is easier to use*: data-entry is easier and faster. Labels can be completed by the program, and codes can be memorized and entered only with the pressure of one key. Searching data becomes easy, thanks to the lists of labels which remember which information exists in the archive;
4. *Easier analyses*: when a group of records is selected, frequencies are automatically produced on all coded fields. For example if we select only the married people with a university degree only one search is necessary, and all frequency distributions are produced automatically.
5. *Territorial analyses*: it is possible to produce territorial distributions thanks to the numerical code which is associated to each town.
6. *All information available*: coded information permits easy access to all the information of the archive and instant distributions on all coded fields. When information is not coded search is complex, for example MARR (for married) might be found as MAR or MAR.; counting is difficult and time consuming.
7. *Scientific research*: beside producing frequency distributions, statistical archives allow to study relations between variables using cross tables, correlations, and factor analysis.

How to choose items

There is an interplay between theoretical hypothesis and data gathering. No one would ever gather information out of thin air without making any hypothesis. There is always previous inductive, even if unconscious, reasoning which generates the hypotheses which are used to select the variables which need to be observed. The progress of research is based on a **Inductive → Hypothesis → Deduction** spiral and the quality of the information is strictly linked to this. If little attention is given to this process the quality of the data and the relevance of the results will inevitably be low.

Preparing a questionnaire or a form is one of the most delicate moments of the entire work. If insignificant information is asked and important information is omitted, results will be irrelevant and of little use. Before starting preparing a questionnaire or a form it is important know that:

1. It is necessary to distinguish among key variables, explicative and structure variables:
   1. *key variables* are those which describe the topic under investigation, for example if the study is relative to cancer, key variables will be relative to cancer;
   2. *explicative variables* are all those variables which might be correlated (linked) to the key variables, for example in the case of cancer it could be the environment, stress, food, and so on;
   3. *structure variables* are age, sex, education, profession; variables which are

      usually used to describe the sample of the study.
2. In order to avoid choosing explicative variables which might be irrelevant it is important to be helped by experts who have a good knowledge of the subject.
3. It is important to remind that a study becomes useful when it compares different hypothesis and theories. The aim of scientific research is that of testing different hypotheses, comparing them, and finding which and when it works. Scientific research is a process of continue change and evolution during which hypotheses are revisited, changed and often abandoned.

The process of formulating a form can be divided in several steps:

1. first, start declaring which is the aim of the study (key variables);
2. second, continue listing all those variables (explicative variables) which might be correlated to the key variables. It is very important to keep track of the hypothesis which support them, in this way it will be easy to interpret the results, otherwise it will be difficult to say which hypothesis have been proved and fall in errors as, for example, paying attention to secondary information and producing interpretations which are totally irrelevant and of little scientific value;
3. finally prepare the form (questionnaire, observation grid, …) and test it in order to assess if it works or if it can be improved and optimized. It is necessary to continue testing the form until it reaches a high levels of quality and efficiency;
4. when the form concerns information flowing in a organization, it becomes necessary to involve, in the designing process, all the people which are involved in the organization.

Objective and subjective information.

It is well known that when people answer using social masks. It is though important to note that the relational methodology used in Sintropia-DS (Syntropy 2005, 1: 117-113) works well even with data which has been distorted by masks. For example, unsatisfied people tend to present a positive image, not only to others but also to themselves.

Psychologist have tried to solve this problem by developing tests which are designed to get information in an indirect way. Secondary questions which are correlated to the trait which is being measured, and which the person does not distort with social masks, are used. In this way it is possible to obtain an indirect estimation of anxiety, depression and other information which otherwise would be masked.

It is interesting to note that even when data is distorted, by the effects of social masks, and cannot be used in a quantitative study (which uses average values or frequency distributions), it can be used successfully with correlation analyses. This happens because the mask is coherent: it is not used only on one variable, but on all correlated variables. For example, if people who say not to be depressed, instead are depressed, they will also say not to feel anxiety, when instead they feel it. The relation between depression and anxiety will shows up, even if the subjects are using masks.

In the case of a study with 200 students, the following results could be obtained:

	Depressed	Not depressed	Total
Anxiety	15	3	18
No anxiety	2	***180***	182
Total	17	183	200

Because of the social mask the answers are mainly grouped in the cell "no anxiety" and "no depression". Depression and anxiety come out as correlated (concomitance of the answer No). Using, on the same sample, psychological tests which provide objective measures of depression and anxiety the results change into:

	Depressed	Not depressed	Total
Anxiety	***158***	10	168
No anxiety	2	30	32
Total	160	40	200

the answers are mainly grouped in the cell "yes anxiety" and "yes depression". Because correlation are concomitances, depression and anxiety remain correlated (concomitances of the answer Yes).

This example shows that if a relation exists, this relation shows up even if the data has been masked. This happens because the mask is used in a coherent way on all the answers. This last consideration underlines that, when asking someone to answer a questionnaire or a form where there is the risk that the social mask might distort the answers, it is important not to change time or setting, so that the use of the mask remains coherent over all the answers which have been given.

Redundancy of information

Sintropia-DS allows to work with complex phenomena where the same information can have more than one meaning. For example it might be find difficult to decide if it is better to ask "*do you feel alone?*" or "*do you feel* lonely". It is always good habit to introduce more items (redundancy).

One item one information

It is important to remember that the aim of a scientific research is that of finding relations. Data analysis will tell which information are linked, correlated. For example it is wrong to ask: *Loneliness causes anxiety?* because in the item the relation between loneliness and anxiety is introduced. Data analysis will not be able to tell us if this relation really exists. In order to study the relation between loneliness and anxiety it is necessary to formulate two different items: *Do you feel lonely? Do you experience anxiety?* It is also important to ask information

in a clear and direct way. Asking information in a indirect way can lead to the total absence of information.

The following question is not correct:

Did the family receive State Aid?
1. Yes
2. No
3. It is a one parent family
4. It is a two parents family

The lists shows two items, one about State aid (Yes/No) the other one about Family type (one parent family, two parents family). The correct version would therefore be:

Did the family receive State Aid?
1. Yes
2. No

Family type:
1. One parent
2. Two parents

It is very important to distinguish between different variables (items): each variable = one data-entry field (or more).

Organization and information

Information is one of the basic ingredients of any organization. There is a moment when the information is collected and someone is responsible for the different kinds of information. When it is known what information is to be taken and who will be responsible for it, we are making clear which are the roles, and how different roles interplay. For example, which information is gathered by the doctors, which by the psychologists, which by the nurses, which by the secretaries and the administration office and to what extent this information may be shared. Introducing the scientific approach in an organization can sometimes be almost a game. Working on the forms, which will be used to collect information and the procedures used to gather this information, can be transformed into a game in which every participant expresses his different needs, ideas and theories. Therefore it is possible to clarify the way in which each person works and gets the information. In this way a more efficient method of work, which avoids repetitions, uses less resources, avoids frustration and is perceived as higher quality by the user is achieved.

This transparency of the information and of the roles requires that the people involved in producing the services share the aims of the organization. People who are working only to acquire personal power and influence, will perceive the work on the information as a threat to their power, and if the process of developing the forms is not adequately conducted, conflicts can generate, and the process can come to a halt. Better information means better organization, services, efficiency, effectiveness and use of resources. Power struggles always result in economical costs.