Quantitative Data Collection

As previously mentioned, you can use qualitative or quantitative methods to conduct data collection depending on your information needs and context. In this article, we will look at how the quantitative methods are structured, while in the next articles we will see how qualitative methods and sampling methods work.

The questionnaire

The main tool used to collect quantitative data is the questionnaire.

A questionnaire is a structured set of questions designed to elicit specific information from respondents. They are considered high-quality when they are able to get exactly the information you need. If you are going to design a questionnaire, there are some guidelines you will need to follow: first, limit the number of questionnaires developed. This is because given the financial and time resources when you begin to design your data collection you should look for ways to collect the information through as few tools as possible.

For example, if you are designing a questionnaire, try to identify opportunities to collect data on multiple PMP indicators using a single questionnaire. This is also in respect of the time the collection participants are giving you. Also, it's important that you use a consistent implementation: make the same questions to each respondent, in the same format and order. This helps ensure that the answers are clear, valid, and reliable. To do this, you should also use closed-ended questions, as they are a type of question that, by providing a predefined list of answer choices, generates answers that are easy to code and analyze. You will then need to include "Skip logic" in the design of your questionnaire, to allow respondents to skip a question based on their response to a previous question and include "I don't know". Also plan to use an exhaustive list of responses, designing the questions to include all appropriate responses. Since in many cases it is not possible to include every possible answer, plan to include an "other" category, with a space for the respondent to fill in.

Finally, a fundamental choice related to the design of a questionnaire is the communication medium that will be used to present the questionnaire to respondents and record responses. Traditionally, questionnaires have been paper-based, but today they are often administered digitally. When choosing the medium, carefully consider your target population and local logistics (e.g., if you decide to administer it digitally, make sure respondents have digital tools or an internet connection).

Type of questionnaire

- The facilitated questionnaire, which is administered by an enumerator and uses closed-ended questions answered by the respondent. However, a facilitator is needed as a support and resource for respondents. This typology has the advantage that respondents do not have to be literate and that there is a high level of cooperation with a low level of refusals. On the other hand, however, it has the disadvantage that it is time-consuming, expensive, and requires space and privacy for interviews. Also, it requires facilitator training and implies some travel costs.

- The self-administered questionnaire, which is completed by the respondent without the support of an enumerator, has the advantage of being easily deployed and accessible to a large

population across a wide geographic area. On the other hand, it has disadvantages because it requires respondents to be literate, as well as the fact that data entry can be cumbersome if responses are not collected through digital devices. In addition, in the absence of a facilitator to motivate participants, there is a possibility that response rates will be low. Using this type of questionnaire requires a good organization and budget for the distribution and collection of the questionnaires.

Let's now look at how qualitative methods are structured, to better understand the importance of knowing how to choose between them based on your information needs.

Data Collection Methods: Qualitative Data Collection Tools

When we talk about qualitative data collection, we are referring to the fact that interviewers use tools that allow them to explore responses in-depth and understand "hidden" processes that could not emerge through closed questions.

What tools can you use

You can use two tools to conduct qualitative data collection:

- Semi-structured Interview, that is, a guided discussion between an interviewer and an individual respondent. This tool is designed to bring out the depth and rich context of the interviewee's perspectives, opinions, and ideas and allow the interviewer to understand them.
- The second tool you can use to conduct qualitative data collection is the Focus Group Discussion, which is a guided discussion among respondents in a group, rather than between an interviewer and an individual respondent. This mode provides an opportunity for a small group of participants to interact with each other while comparing notes. A focus group discussion requires an experienced facilitator and annotator. Typically, a meeting includes 8 to 12 participants, but recruiting the right participants is critical. After narrowing down the topics and questions, you will have a better understanding of who should participate in the discussion. You will need to choose participants who can speak firsthand about perspectives, or who are stakeholders for a group.

Guidelines for Design

To conduct qualitative data collection you can follow guidelines that will direct your work. First of all, plan the questions: for solid qualitative data collection, it is critical to carefully plan the questions that will frame the conversation, so prepare them in advance, writing them down carefully. Then, you must use open-ended questions, which are questions that allow respondents to give a free response in their own words.

There are two types of open-ended questions: the Content-mapping questions, which are referred to as opening questions and are intended to initiate exploration of a topic by raising a question. To encourage rich discussion, facilitators often follow content mapping questions with content analysis questions; and the Content-mining questions, also known as probing questions, are follow-up questions that require more detail or explanation of an answer to a content-mapping question. These types of questions are completely unstructured. In this case, facilitators will need to have the ability and flexibility to adapt the flow of the conversation and ask the right content mining questions. Content mining questions allow the facilitator to explore a topic more deeply and investigate unanticipated topics.

Sampling Methods

A sample is a subset of the population or community you choose to study that will give you knowledge of the population or community as a whole, and you will need to use it to collect data from respondents. It is important to select a sampling method that will allow you to collect data from respondents who are in the best position to meet your information needs. Sampling a population is necessary because, since you have time and cost constraints, it would be impossible to collect data from all members of the population.

The goal of sampling is to identify a sample group that will give you high-quality data. There are primarily two sampling methods: random sampling and purposive sampling.

Random Sampling

Random sampling can be used when you plan to use quantitative methods and analysis, as respondents are selected at random from a list of the entire population of interest. Using this approach increases the likelihood that your sample will resemble a subset of the larger population or the entire population. In fact, when the sample is large enough, the results are more likely to represent the larger population. When we talk about a population we are referring to a set of similar people, objects, or events that are of interest to some question or experiment.

Four steps to identify a Random Sample

To construct random sampling, you must clearly identify the sampling unit. The unit is the individual person, category of persons, or object from which the measurement is taken (e.g., children under age 5, adolescents, women, men, and families). Next, you must choose a method to calculate your random sample: it can be a simple random sampling, in which every unit of the population has an equal chance of being selected; systematic sampling, which involves listing and numbering all potential subjects and then selecting one; or you can choose cluster sampling. In this case, the population is divided into naturally occurring clusters, such as geographic areas, schools, or workplaces.

All clusters are listed and a cluster sample is selected at random. If your data analysis plan includes subgroup disaggregation, your sampling method should specifically include those subgroups (strata) of the population. This is referred to as stratified sampling, which is a strategy that allows you to divide a population into separate strata to allow for the analysis of each subgroup. After choosing the calculation method, you will need to determine the sample size. This is important because the larger the sample size, the more likely the sample will accurately represent the population. Using the margin of error and the confidence level you can also determine how well a sample represents the population.

These two factors express the relationship between the result you would get if you could census the total population and the result you get using a population sample: the confidence level expresses how likely it is that the result found using a sample of the population will match the result that would be obtained by surveying the total population (the true result). The margin of error, on the other hand, expresses the maximum expected difference between the real result for the total population and the result found using a sample.

Lastly, you will need to select your sampling units, in which you must include men, women,

families, individuals, children, and adolescents. At this juncture, you will have a sampling frame, which is a specific list of units that you can use to generate your sample (e.g. census lists, a list of employed teachers, a list of project participants, etc.).

Purposive Sampling

This is a method that selects the sample group deliberately, rather than randomly. The purposive sample is chosen to reflect important characteristics of the groups within the sampled population and in accordance with the objective of the study. This type of sampling is primarily used to collect qualitative data in order to gain a deep understanding of a particular group's experience or perspective in a specific context. Because this sampling is not random, the data collected cannot be generalized to the entire population.

How can you build it

To begin building a Purposive Sampling, start by clearly identifying the population and sample structure you want to create (same step as a random sample). Then you'll need to establish the inclusion and exclusion criteria for the sampling, which are especially important when selecting purposive samples because they specifically define why you're choosing these particular people to participate in your data collection.

Next, you will need to select the sampling method you intend to use to identify your targeted sample: you can use Best and Worst Case Sampling, comparing communities or individuals that are considered the best and worst cases based on certain characteristics; or Typical Case Sampling, providing insight into the overall scenario by choosing those communities or individuals that are considered average.

Alternatively, you can choose a Critical Case Sampling, gathering information from the community or individuals that are significant for understanding a particular context or situation; or a Quota Sampling, in which you can attempt to select participants so that certain characteristics of interest are represented in the sample in proportions consistent with estimates of their proportions in the population. Finally, you may also choose Snowball (or Chain) Sampling, in which you collect information from participants in stages, starting with respondents known to the evaluators or partners and then asking these respondents for advice on who else to talk to. The advantage of this method is that it allows you to identify previously unknown sources of information.

The last step is to determine your sample size, which is calculated differently than random sampling. This is because, often, qualitative data is used to triangulate quantitative or other qualitative data. That is, they are used to validate other data through cross-validation using more than two sources. So enough focus groups or interviews need to be conducted to reinforce and confirm the patterns that are emerging.

The sample size is decided on:
- The budget constraints and resource limitations, which can affect your sample size decisions;
- And whether the data analysis plan in your PMP requires you to compare subgroups. The more subgroups you have the larger the size increases; and of. For example, you may have to limit the number of subgroups you compare if you lack the financial resources to hold more meetings.

How to do a Successful Data Collection

In this brief focus, we will look together at how you can start thinking about how to collect data, after designing your collection tool and sampling strategy.

To ensure the success of the data collection you will need to take some precautions: the first step is to translate your data collection tools if your project works in a region that uses multiple languages so that you are not influenced by a particular language; then train and update data collectors and test your tools. To do this you will need to write instructions to lead your collection tool and train the employees.

This training should include an explanation of the basic ethical principles of good data collection. Training those who will be collecting data serves two purposes: to build the skills of your data collectors and to ensure that your tool works as it should. Also, make sure the way you have written your questions is understandable to respondents. The third step to conducting successful data collection is to review and finalize your tools: after testing your tool, incorporate any revisions into your final document. Finally, plan your data collection activities. That is, make sure your plan includes enough time for each data collection event.

Remember that questionnaires should take no longer than 45 minutes, and interviews and focus group discussions can take up to 90 minutes. As mentioned earlier, once the interview is complete, plan to ask the participant if they have any questions. Choose a location for interviews and focus groups that provide privacy and an appropriate level of comfort.

A brief overview of Data Management

When we talk about data management, we are referring to the process of managing data through the stages of its life. By creating an efficient data management system, you will be able to effectively analyze, interpret, and use the data you collect. There are four main components to comprehensive data management:

Data entry
Data entry consists of identifying the type of database for data entry and then entering the data. Using a database effectively enhances your ability to access, manage, and share data; improves its security, and facilitates timely decision making.

Based on the types of data you need to manage (qualitative or quantitative), you will inform your information and communication technology (ICT) strategy for M.E.A.L.: in fact, there are special software to manage both qualitative and quantitative data and can vary depending on the size of the data collection.

There are 2 steps required to enter data: you will need to create a data entry protocol, as data entry errors can compromise your data, analysis, and M.E.A.L. results; then identify the requirements for who enters the data. In fact, when digital devices are used, data are often automatically uploaded to the M.E.A.L. project database. However, in some environments, it may be necessary to enter data by hand. For this, it is necessary to specify whether or not the person doing it has prior experience.

Cleaning Data

As the use of digital devices to collect data has increased, the risk of transcription errors has decreased, Nonetheless, your team should still invest in data cleansing to ensure it is accurate and error-free. When we talk about data cleansing, we are referring to identifying and removing errors and inconsistencies from a data set to improve its quality. To clean the data do quality checks by randomly selecting and comparing the raw data with the electronically entered data to check for data entry and coding errors.

You can also identify outliers, that is, check for unexpected entries in the data. The presence of an outlier could mean that the person entering the data doesn't understand the process and has made a coding error. Once you have identified the errors remove the duplicate entries. In addition, it is recommended to keep a log of errors to review with the data entry team any trends and patterns observed in the errors to improve future results.

Data storage and security

This component addresses the importance of ensuring that data is secure and protected from unauthorized modification, copying, tampering, unlawful destruction, accidental loss, improper disclosure, or unauthorized transfer. The European Union in 2018 enacted an EU General Data Protection Regulation (GDPR) to protect its citizens. It seems irrelevant if you're working outside the EU, but remember that the regulation affects any organization based in the EU. The main benefit of the regulation is that consent for data use has been strengthened, so you need a request for consent that is clearly stated using plain language that can be understood by everyone. Efforts to protect the individual's privacy and their data must be part of the initial design of any data management system.

Data retention and disposal

There are several reasons why an organization may retain data, including that of complying with government regulations, donor requirements, or organization policies. Regulations on data retention requirements vary by project, but they are important to be identified before M.E.A.L. data collection activities begin, and when data are no longer needed, all records must be disposed of or arranged so that it is impossible to identify data respondents. You may also choose to retain the data after the end of the project activities, but you will need to conduct a de-identification process to maintain the anonymity of the respondents. There are two techniques for doing this: anonymization, which involves stripping the data of any identifiable information; or pseudonymization, which involves replacing information fields with a code that protects the identity of the respondent.

If you are interested in knowing more about project writing and evaluation, and would like to have the assistance of professionals, you can email us at ssr@signis.net. At SIGNIS Services Rome we are experts in the sector and have been involved in project writing for the creation and development of communications projects all over the world for decades.

*This content was curated by Valeria Appolloni and inspired by the materials published on Kaya, published for non-commercial and educational use.