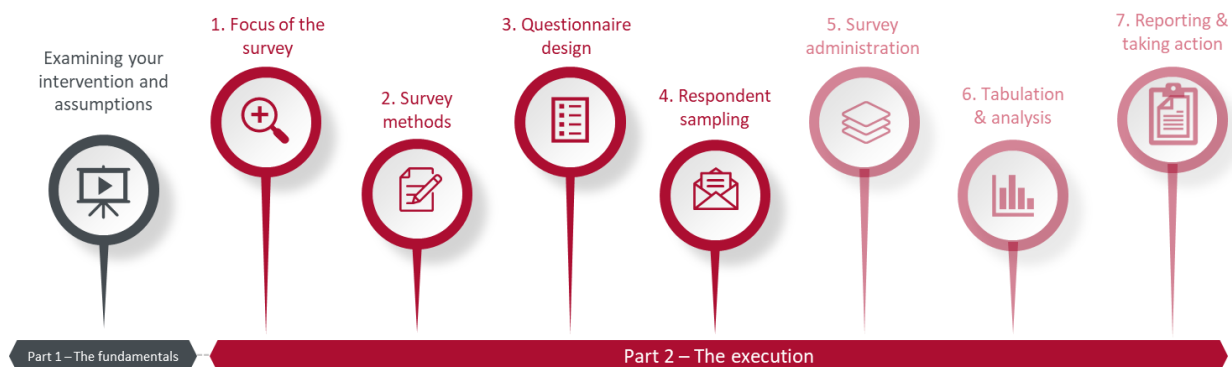


MODULE 4. DETERMINE THE RESPONDENT SAMPLE



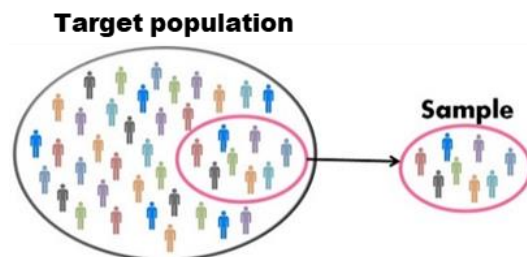
Now that you have your survey ready, you need to sample your pool of respondents. Of course, the ideal would be to collect answers from all of them, but in most cases this is not feasible, practical nor cheap, unless your target group is very small. *Module 4* helps you select a sample without affecting the quality of the results.

Essentially, you need to make sure that your selection of respondents is representative, i.e., reflects the characteristics of the entire population you would want to survey. If, for example, you want to survey your clients, who are mainly female, you should seek to collect answers primarily from women. If instead you happen to select more male respondents (for example, because you chose an online survey and men are more likely to have access to a computer or smartphone), this could yield biased results.

Before starting to select your sample, you need to understand some key concepts:

The **target population** represents the entire group of individuals you are interested in. They are the people from whom you want answers. For example, they can be your suppliers, customers or employees.

A **representative sample** is a subset of the target population that bear the same characteristics as the larger group. If you have a well-selected representative sample, you do not have to survey the whole population, but only the representative sample, and you will get similar results. This will save you time and money.



Sampling is the process of determining the right number and profile of respondents in a sample.

The **sample size** is the number of respondents. There are a few important concepts related to sample size:

- **Maximum reachable sample size:** the estimated number of answers you would likely collect if you tried to survey everyone in your target population, based on the quality of your database and the expected response rate. This represents the upper limit for your sample size
- **Desired sample size:** the number of answers you would like to collect. This is set by you
- **Actual sample size:** the actual number of people who complete your survey. This will only be known after implementing the survey, and ideally will coincide with your desired sample size.

The **database quality rate** is the number of correct contacts in your database over the total number of entries. Most contact databases have at least some entries which are either not correct (someone made a mistake when entering the contact information), or not up to date. How accurate

do you believe your database is?

The **response rate** is the number of people that answer your survey divided by the number of people you contacted for the survey. It can vary depending on the quality of your database, on your mode of contact (e.g. SMS typically have a much lower response rate than phone calls), and your logistics (e.g., the time of the day may affect respondents' ability and willingness to answer your survey). On average, you can assume the following:

Method	Response rate
Online	30%
Phone	60%
SMS	40%
Face-to-face	70%

In general, a higher response rate will give you more representative and reliable results. To increase the response rate, you will need to attempt to reach each contact multiple times (e.g., repeated calls, email/SMS reminders etc.). You could also choose to offer a small incentive for people who complete the survey (e.g., chance to win a small item or voucher, access to a digital resource, etc).

Now that definitions are clear, let us dive into the process. First, you need to identify exactly who and how large your target population is. Then you need to set your desired sample size. Choosing the "right" sample size is a trade-off between reliability and feasibility: the bigger your sample, the more reliable, representative, and rich the results of your survey will be, but also the more time and resources you will need to achieve that size. While there are rigorous statistical methodologies to determine the sample size, a good rule of thumb is that your sample should not be smaller than 100 contacts (unless your target population is very small¹), and not cover less than 30% of your target population (unless your target population is very large²).

Nevertheless, you must consider that not all the people in your contact database will answer. The response rate as well as the quality of your database must be factored in to determine the number of people to contact to achieve your desired sample size. Indeed, even if you tried to survey the whole target population, you would end up with only a fraction of responses (i.e. with your maximum reachable sample). For this reason, the number of contacts you should try to reach should be higher than your desired sample size.

Here is a practical example: suppose that your whole target population consists of 1000 clients. You estimate that your database is 90% accurate, and your response rate will likely be around 60% (with a phone survey). This means that your maximum reachable sample size is $1000 * 90\% * 60\% = 540$. However, you decide that a sample of 300, i.e. 30% of your target population, could be sufficient. To achieve that, you will then need to target $300 \text{ divided by } 60\% \text{ divided by } 90\% = 556$ contacts. With 556 contacts, an estimated 56 will fail because of poor entries into your database, and another 200 will fail because people will not answer your survey, leaving you with your desired sample size of 300.

Next, you need to select the right type of respondents to ensure representativeness. To do this, you will need to think about which characteristics of your target population are relevant and might affect the answers you get. General attributes such as gender and age are typical examples: in many cases women answer differently from men, and younger people differently from older people. But you need to think through what makes sense for your survey, and identify characteristics that are relevant for your questions in particular. For example, if you want to interview the farmers supplying you produce to ask them about their satisfaction with your company, a relevant variable could be how long they

¹ If your target population is 200 or smaller, it is preferable to target the full population for the survey.

² If your target population is 2000 or larger, a sample covering less than 30% of the population is also acceptable.

have been supplying produce to for. Or if you want to interview individuals who completed your training program to see if they got a job afterwards, relevant variables might be the type of training they took and the date of completion. Two words of caution though: you need to select characteristics that you track already in your database, so you can check what proportions would be representative; you cannot select too many characteristics either, as this would make the sampling very difficult to calibrate. You essentially need to zoom into the 2-5 characteristics that could significantly influence the answers to your survey.

Finally, you need to see how the characteristics you identified are distributed in your target population, and select a sample that is calibrated to reproduce roughly the same proportions. To continue with the example above: after determining that you need to select 556 contacts, you identify 2 relevant sampling variables: gender and region. Among your 1000 clients, 300 are men and 700 are women, and they are located in 3 regions: 500 in region A, 250 in region B, and 250 in region C. Therefore, the 556 contacts will need to respect the same proportions: 30% men (167) and 70% women (389); 50% from region A (278), 25% from region B (139) and 25% from region C (139).