



DCED

The Donor Committee for Enterprise Development

Practical Advice for Selecting Sample Sizes

William Fairbairn and Adam Kessler. Sample Size Calculator by Richard Tanburn. May 2015

1 Introduction

Many project managers today have to measure their results, without ready access to the specialist expertise required for rigorous studies. In this imperfect world, a question frequently asked by those using the [DCED Standard for Results Measurement](#) is: “how big a sample do I need? Budgets are tight, so I cannot include everyone.” This guidance gives you practical advice that you can use in your daily work, to be roughly right.

1.1 What does this guide aim to do?

This guide aims to introduce simple rules of thumb which will help users find an appropriate sample size. No mathematical ability is required, although users will be asked to think hard about the research that they aim to conduct. The guide focuses on quantitative surveys, although qualitative information is also essential for strong results measurement systems.

This guide is aimed at organisations who prioritise **cheap, rapid** surveys over **expensive, rigorous** ones. Consequently, they permit greater levels of inaccuracy and uncertainty than would be the case in scientific research, or in more specialised sectors. The aim is to get estimates that are good enough to inform management decision making – not to prove results beyond reasonable doubt. If you require more accurate sample size calculations, it would be best to consult a statistician.

Sampling is only one aspect of a successful research strategy. If the data quality is poor, the attribution strategy badly designed, or the sampling strategy poorly implemented, then the results will not be credible. For more information on dealing with attribution, qualitative data and research, [please consult other DCED guidance](#).

1.2 How to use this guide.

The rest of this guide is split in four parts. Chapter two describes **initial considerations affecting sample design**. Chapter three describes **types of sample design**. This shows how to select the sample, which is just as important as how many people are in it. Chapter four then outlines a **sample size calculator** which can be used to select a sample size, and describes how to use it.

2 Initial Steps

2.1 Define the population

The **population** is the target group of your research – the group that you want to ultimately find out about. This can be a big or small group of people, depending on the question that you are trying to answer. It is important to specify the population as clearly as possible, since it will affect who you draw your sample from. For example, the population could be:

- Everyone in Bangladesh.

- All carpenters over 18 within a particular province of Ghana.
- Every customer of a business that your organisation supported.

2.2 Understand the purpose of the research

A smaller sample size doesn't necessarily mean a useless survey, but does restrict the amount of analysis possible. For example, you might be able to analyse fewer subgroups (see section 2.3), be less confident in your findings, or have less precise results (see section 4).

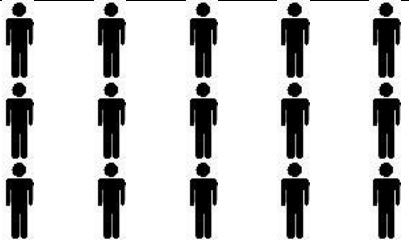
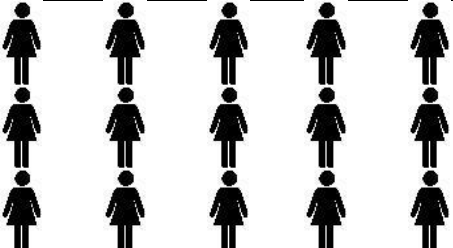
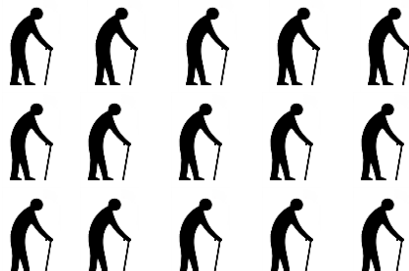
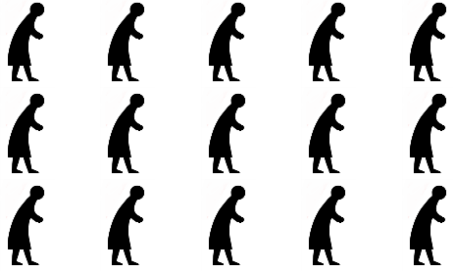
Of course, larger sample sizes also cost more money. Consequently, you need to think about **what the results will be used for**, in order to understand whether a large sample size is necessary. For example:

Factors leading to higher sample size	Factors leading to lower sample size
Information from the study is used for advocacy to external stakeholders.	Information from the study is used internally.
Intervention is expensive, important, or has a large expected impact.	Intervention is cheap, relatively less important, or has a low (or unpredictable) impact.
Intervention is an untested pilot, and the study will help decide whether to scale it up.	There is already significant secondary evidence supporting the intervention logic.

2.3 Subgroup analysis

Development programmes often need to analyse sub-groups in a survey separately. This is known as 'disaggregating' data. Disaggregation can be a valuable part of a results measurement strategy, as it shows how different groups are affected. For example, you may wish to disaggregate the data in order to find out whether men and women are benefitting equally.

This can create problems in analysis, however, as the disaggregated data has a smaller sample size than the original. For example, assume that a study required a sample of 60 people to carry out satisfactory analysis. The results measurement team then disaggregated their data by gender and age. They end up with the following samples:

	Male	Female
Young		
Old		

If the organisation wants to analyse the overall results of the study, they can do this based on a reasonable sample size of 60. However, if they want to give results for sub-groups, such as young men

or old women, they can only use a sample size of 15 – which may not be large enough to enable you to be confident your estimate is correct.

Consequently, the sample calculations below should be used to estimate the sample size needed for **one sub-group**. If you want to analyse multiple sub-groups, you need to either increase the total sample size by the number of sub-groups, or accept that your analysis of the sub-groups will be much less precise.

3 Types of sample design

Key message:

There are a number of ways to select a sample. The most important consideration is to make the sample as **representative as possible of the population**. The best way to do this is to randomly choose the sample, and be aware of the possible sources of bias. Consequently, when selecting your sample, continually ask yourself whether the sample is unrepresentative of the population, and how this bias could be removed.

3.1 Introduction to Sample Design

The aim of sample design is to pick a sample which is **as representative as possible** of the population, given the available resources and information.

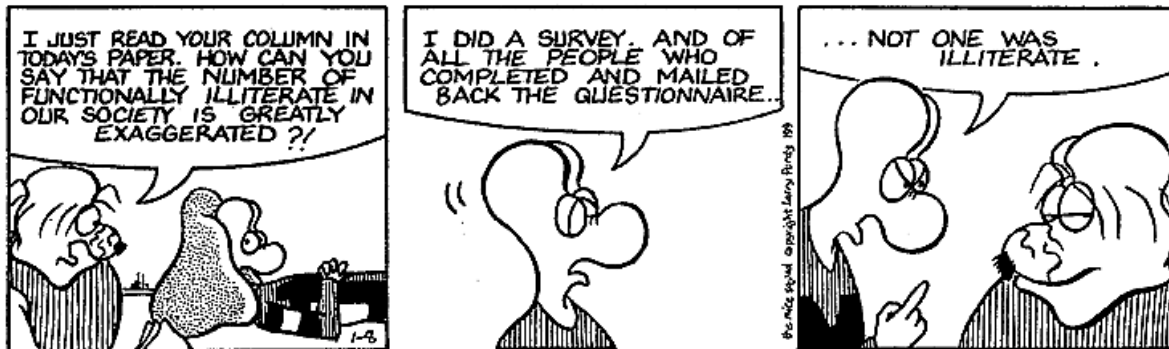
If the sample is **not** representative, then you risk being misled by the survey. For example, imagine that a survey team only interviewed farmers who lived near to a main road. These farmers might be wealthier than those who live in remote areas, and so the sample would not be representative of the whole population of farmers.

In practice, it is often challenging to get a perfectly representative sample. Consequently, carefully to think about all the factors which might bias your sample, design your study to eliminate as many as possible, and acknowledge clearly in your report what factors remain.

Common type of sample bias	Description	Ways to improve it
Geographic bias	It is easier to conduct interviews in areas which are easily accessible, for example, because they are close, have good road access. However, people living in these areas may be different (typically wealthier) than those living in more remote areas.	Use the below sampling designs to ensure that you cover a representative set of areas. If you still suffer from geographic bias, for example because it is too costly to visit really remote area, consider conducting additional secondary analysis or interviews to test whether the area you missed is significantly different.
Response bias	Some cultural/ethnic/socioeconomic groups might be more likely to answer a survey than others. This can be for many reasons. For example, women may be less willing to speak to strangers in some societies, leading to fewer women answering a survey.	If this is a big concern, stratify the survey (see below) to ensure that you have a representative sample. You can also plan to address this issue. For example, if you know that farmers will not answer during the day because they are out in the field, conduct the survey during the evenings.

<p>Self-selection bias</p>	<p>If participants in your survey have a choice about whether they will respond or not, then participants who respond may be significantly different to those who don't. For example, if you send a written survey out to 1000 customers of a business, it may be that only those who are really happy (or unhappy) with the service can be bothered to reply.</p>	<p>As far as possible, deliberately select your respondents using the below techniques, rather than relying on them to respond to your survey. Monitor how many people fail to respond to your survey, and consider self-selection bias in your analysis.</p>
----------------------------	--	---

THE MICE SQUAD



3.2 Simple Random Sampling

Simple random sampling is a technique which requires individuals to be picked at random from a list of all people in the population.

For example, imagine that you are interested in researching the impact of an intervention among the customers of a particular company. If the company is able to give you a list of all their customers, then you can pick your sample at random from this list. This is an effective way to ensure that the sample is likely to be representative.

It is relatively rare for development organisations to have a complete list of their population. If you are trying to conduct simple random sampling in a small population (for example, a village), but don't have a complete list of the population, you could randomly select houses to interview. For example, you could interview one person from every third house in the village. This method is potentially biased, as some households may not be available and willing to answer a survey. It is, however, cheap and often a practical sampling technique.

3.3 Stratified Sampling

Simple random sampling can be misleading when there are many subgroups in a population. Using simple random sampling, you may by chance survey more of one group than another. This makes the sample less representative of the population, and so less accurate. It also means you can't analyse sub-groups, as you may not have enough data from one sub-group for detailed analysis.

Stratified sampling overcomes this problem by splitting the sample into sub-groups, and then randomly selecting respondents within each sub-group. The sub-groups could be based on gender, ethnicity, age, or any other significant difference in the population. For example, suppose you were

sampling from the Rwandan population, which has three ethnic groups. The Hutu make up 85% of the population, the Tutsis make up 14%, and the Twa make up 1%. In order to ensure that your sample is representative of the population, you could ensure that 85% of your sample were Hutu, 14% Tutsi, and 1% Twa.

Of course, this would mean that you had very few Twa respondents. If you were particularly interested in understanding whether the Twa differed from other ethnic groups, you may wish to deliberately choose more respondents from the Twa community, in order to have enough samples for a reasonable analysis. This would mean that your sample was unrepresentative of the total population, but allowed you to examine the characteristics of the Twa community.

Remember that, if you analyse sub-groups, you need to increase the sample size accordingly, as discussed in section 2.3.

3.4 Cluster Sampling:

Often it is not possible to use simple random or stratified sampling. This can be for two reasons:

- **There is no complete list of the population.** For example, imagine that you are conducting research into fish-farmers in Ethiopia. It is likely that nobody has a full list of all the fish-farmers in Ethiopia, and so it would be difficult to use simple random sampling. Similarly, if the aim of the study is to interview listeners to a radio show, there may be no way of knowing in advance who these listeners are.
- **It is too expensive to visit every location.** Simple random sampling can require a lot of travel. Even if you could find a complete list of all the fish-farmers in Ethiopia, a random sample from this list might send you to multiple locations, thousands of kilometres apart. This imposes significant additional costs.

Consequently, a common sampling technique is **cluster sampling**, which splits the population up into groups, or 'clusters'. A cluster might be geographical (such as residents of a village), or based on a business (such as the customers of a particular input suppliers).

In order to use cluster sampling, first get a list of all the clusters that you are using. This might be a list of villages, for example. Pick some villages randomly from this list – the more you can visit the better. Then visit each of these clusters, and within the cluster try to randomly select households to interview. For example, if your cluster is a village, you might survey everyone within that village. If your cluster is an input supplier, then you might visit the input supplier and try to get a list of all their customers, and randomly sample from that list.

The problem with cluster sampling is that respondents from the same cluster (such as a village) tend to be quite similar to each other. Consequently, interviewing a second person from the same village doesn't tell you as much about the whole population as interviewing a person who is randomly selected from another village.

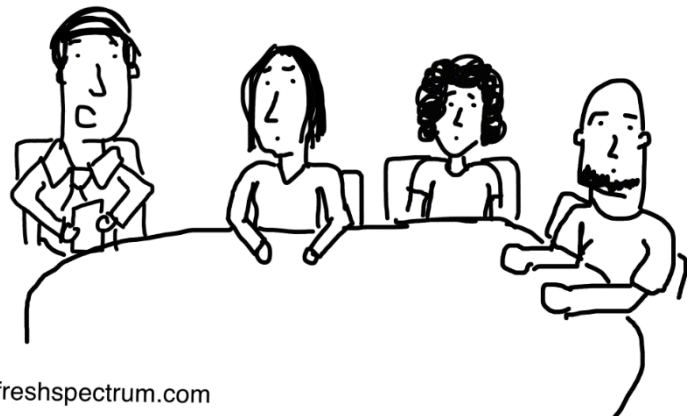
As a simple rule of thumb, when using cluster sampling try to use as many clusters as possible to increase your accuracy. To be safe, increase your sample size by about 20% to compensate for the reduced power.

3.5 Mixing sample designs

In practice, many sample designs combine several different techniques.

Consider a programme that wanted to conduct research into farming in a particular district of Nigeria. Given time and budget constraints, the researchers may decide to conduct a **cluster sampling** technique. They obtain a list of all villages within this district, and randomly select four villages to visit. In order to understand how results differ by gender, the researchers decides to use **stratified sampling**, and decides to select a sample that is 50% male and 50% female. Consequently, within each village, the researchers visit the local chiefs and obtain a list of all inhabitants of the village who work as farmers. They select 10 men and 10 women at random from that list to interview.

I invited you all to join our focus group because you were the easiest to track down and therefore should provide us with the best feedback.



The aim is to find a way of selecting samples that is practical, fits within your budget, and avoids major sources of bias.

4 Calculating your sample size

In order to calculate a sample size, you will need to use the [sample size calculator](http://enterprise-development.org/page/calculator) provided online at: <http://enterprise-development.org/page/calculator>.

The calculator asks a series of simple questions – assuming no prior mathematical ability – and produces a minimum sample size based on your answers. This section provides simple guidance on how to use the calculator.

Sample size calculator

What kind of calculation do you wish to perform?

Which type of test are you looking to perform?

What range will 95% of your data fall in? - Estimate the range 95% of data will fall in.

Your data have a mean of 50 and a standard deviation of 25.

What margin of error do you require? The margin of error is the amount of error that you can tolerate.

Your margin of error is 5, which is 10% of 50, your implied mean.

4.1 Overall advice for using the calculator:

- **Sample size choices include trade-offs.** A small sample size can, in some cases, be good enough. It does, however, restrict the scope of your analysis. In particular, with a small sample size you can typically only detect larger changes, and can be less confident in your results.
- **If you want to do sub-group analysis, then you will need to increase the sample size.** As discussed in section 2.3, the sample sizes given here are for **one sub-group**. If you want to analyse multiple sub-groups, you need to either increase the total sample size by the number of sub-groups, or accept that your analysis of the sub-groups will be much less accurate.
- **The calculator gives the minimum sample size.** Given likely inaccuracies in the assumptions, it may be wise to sample more than the minimum. To be safe, add 10-15% to your sample.
- **In all cases, the minimum sample size refers to people who actually respond to your survey.** If you survey 100 people, of whom only 50 respond correctly, then your sample size is 50. The easiest way to deal with this is just to keep on sampling until you get enough respondents.
- **Beyond a point, sample sizes will be too small to be credible for quantitative analysis.** 30 people is often considered the minimum sample size for representative data. The sample calculator is designed to not give answers smaller than that.

4.2 Choosing your study design

This calculator supports three different research options, described below. In order to use this guidance, decide which research option you want to use, and then go to the relevant section below.

Section	Research option	Description	Example question
4.3	Descriptive	In a descriptive study, the aim is to estimate the characteristics of a population.	How many hectares of wheat do farmers plant?
4.4	Comparative	In a comparative study, the aim is to establish whether there are significant differences between two groups.	Do farmers who have participated in our intervention have a higher income than farmers who didn't?
4.5	Comparative (paired)	In a paired comparative study, the aim is to compare the same people at two points in time.	Do farmers who have participated in our intervention have a higher income than they did a year ago?

4.3 Descriptive

4.3.1 Information required

After you select this option, the calculator will ask you the following questions:

What range will 95% of your data fall in?

The 'range' is simply the difference between the highest value and the lowest value in your sample. Obviously, before you have done your survey, you will not know the range exactly. You will need to make the best guess you can, based on the evidence that you have available. Ask yourself the following two questions:

- **What is the lowest value that would typically occur?** This will be the lowest value of the range.
- **What is the highest value that would typically occur?** This will be the highest value of the range.

Note that you are not being asked for the highest value that would **ever** occur. You are being asked for the range of values that 95% of your data is likely to fall into. If you don't know the answer, try asking a few specialists in the sector, such as practitioners or research institutes. The sample size calculator will assume that the **average value** of the sample is half way between the highest and the lowest.

This question asks about just one range, although a survey would typically assess multiple variables, each with different expected ranges. For example, a single survey might assess yields, land size, and costs. Consequently, you should select the most important variable in your study, (often an impact indicator, such as income) and use the range for that variable in the calculator. Alternatively, you could include ranges from several different variables and use the one which suggests the largest sample size.

What margin of error do you require?¹

¹ Sometimes people say 'confidence interval'. A confidence interval is just twice the margin of error. So if the margin of error is +/-10, then the confidence interval is 20 wide. The margin of error means that your real value is plus or minus 10, while the confidence interval means the whole range is 20.

The margin of error is a measure of how **precisely** your sample can estimate the true average of the population.

Imagine that a survey found that respondents, on average, used 70 kilogrammes of improved seed per hectare, with a margin of error of 10. This means that the precision of the survey is 70 kilogrammes, plus or minus ten. In other words, the true population average for the amount of seed used is likely to be between 60 and 80 kilogrammes. If the margin of error had been 20, then the survey would have been much less accurate, even if the average finding was the same. With a margin of error of 20, all you would know is that the true population average is likely to be between 50 and 90 kilogrammes of seed per hectare. (Exactly how likely this is depends on the confidence level; see below.)

In general, a smaller margin of error requires a larger sample size. Select your margin of error by considering the purpose of the study. If the aim is to do a rapid scan of a population which will give a quick idea of their characteristics, a high margin of error is acceptable. If the aim is to get a more accurate understanding of a population, then a low margin of error would be better. A commonly used value is a margin of error of 10% of the expected average value.

In this calculator, enter the margin of error in the same units that you have given the range. (In other words, if your average is 40, and you want a margin of error of 10% of the average, then you need to enter '4' into the sample size calculator.

What confidence level do you need?

The confidence level is a measure of how confident you can be in your answer. If the confidence level is 90%, then you can be 90% confident that the true population average lies within your margins of error. We suggest using a confidence level of 90%, although 95% is often used in scientific research.

What is the population size?

Enter in the size of the population that you are sampling from. Bigger population sizes typically require a bigger sample size. However, the sample size as a percentage of the population to be sampled reduces dramatically. The sample size doesn't change much for populations larger than 20,000, so if in doubt use that figure.

4.3.2 Worked example:

A programme wants to know the average crop yield of farmers in the area that they are working with.

After asking a few businesses working with smallholder farmers, they decide that the lowest likely crop yield that a farmer will have is 50 kg/hectare, and the highest is 400.

The programme decides that it is not necessary for the survey to be too precise, and so select a margin of error of 25. They are also happy with a confidence level of 90%. This means that the programme can be 90% confident that the true population average will be 25 more, or 25 less, than the average, if their assumptions hold true.

This suggests a minimum sample of 34 people.

4.4 Comparative

4.4.1 Information required

A comparative study means that you are looking at two groups. The sample size calculator will tell you the recommended minimum sample size per group, and the total recommended minimum.

After you select this option, the calculator will ask you the following questions:

What range will 95% of your data fall in?

The 'range' is simply the difference between the highest value and the lowest value in your sample. Obviously, before you have done your survey, you will not know the range! You will need to make the best guess you can, based on the evidence that you have available. Ask yourself the following two questions:

- **What is the lowest value that would typically occur?** This will be the lowest value of the range.
- **What is the highest value that would typically occur?** This will be the highest value of the range.

Note that you are not being asked for the highest value that would ever occur. You are being asked for the range of values that 95% of your data is likely to fall into. If you don't know the answer, try asking a few specialists in the sector, such as practitioners or research institutes. The sample size calculator will assume that the **average value** of the sample is half way between the highest and the lowest.

This question asks about just one range, although a survey would typically assess multiple variables, each with different expected ranges. For example, a single survey might assess yields, land size, and costs. Consequently, you should select the most important variable in your study, (often an impact indicator, such as income) and use the range for that variable in the calculator. Alternatively, you could include ranges from several different variables and use the one which suggests the largest sample size.

Although you are doing a survey of two groups, you only need to enter the range for the baseline group, which is the group that has not received any intervention.

What minimum detectable difference do you require?

The minimum detectable difference is the smallest difference between two groups that you want to detect in your study. The smaller the difference you want to detect, the larger your sample size would need to be.

In order to work out what the minimum detectable difference should be, you need to estimate what level of change you would be happy to see between the groups.

Imagine that a business environment reform programme wanted to compare two groups of business. One group of businesses used an online portal to lodge applications for importing and exporting goods, while the other group lodged offline applications. The programme wanted to know whether the businesses using an online portal had their applications approved faster than the other group. They hypothesise, based on stakeholder interviews, that the speed of approval might change from 36 hours (using offline applications) to 24 hours (using online applications.) This is a reduction in time taken of 12 hours. The programme decides, however, that it is not really interested in any change less than 6 hours. That is the **minimum detectable difference**.

In this calculator, enter the minimum detectable difference in the same units that you have given the range.

What confidence level do you need?

The confidence level is a measure of how confident you can be in your answer. If the confidence level is 90%, then you can be 90% confident that the true population average lies within your margins of error. We suggest using a confidence level of 90%, although 95% is often used in scientific research.

What is the population size?

Enter in the size of the population that you are sampling from. Bigger population sizes typically require a bigger sample size. However, the sample size as a percentage of the population to be sampled reduces dramatically. The sample size doesn't change much for populations larger than 20,000, so if in doubt use that figure.

4.4.2 Worked example:

A programme wants to know whether businesses who use an online portal to lodge applications for importing and exporting goods have their applications approved faster than those that use offline applications.

They first need to estimate the range for the group using offline applications. After asking a few businesses, they decide that the fastest an application is likely to be approved is 12 hours, and the slowest is 60.

As discussed above, the programme has decided that the minimum detectable difference is 6 hours. They are happy with a confidence level of 90%. There are a total of 500 businesses in the sector, which means that the population is 500.

This suggests a sample of 30 businesses per group, or 60 in total.

4.5 Comparative (paired)

4.5.1 Information required

This uses exactly the same information as section 4.3, 'comparative'. Consequently, read through the above section first to understand what information that you need.

The only difference is that the range does not refer to the range of values in the group. Instead, it refers to the range of differences between people. For example, consider the below table, which interviews households twice to see how many chickens they have.

Households	Number of chickens at baseline	Number of chickens at endline	Difference (endline minus baseline)
1	0	5	5
2	5	15	10
3	22	20	-2
4	52	61	9
5	49	55	6

If the baseline and endline groups were different, then you would put the range of the baseline group into the sample size calculator. The lowest value would be 0, and the highest value would be 49. If you are doing a paired sample, where the same groups are interviewed twice, then you would enter the range of the **difference**. In the above example, the lowest value would be -2, and the highest would be 10. All other values will be exactly the same as a normal comparison of two groups.

As always, you will not be sure what the range will be before you have done the study. It will be necessary to guess, based on the opinion of experienced team members or external experts.

4.5.2 Worked example

An organisation runs an advertising campaign encouraging households to buy and keep more chickens, alongside training and advice on chicken-farming. They want to know if this intervention increases the number of chickens that households keep.

They have been told that the current average number of chickens per household is around 20. They expect some farmers to sell or lose some chickens during the intervention – for example, their chickens might die. However, they expect most farmers to significantly increase the chickens that they own. Consequently, they decide to use a range of between -10 and 50. This means that they expect the worst-off farmer to lose 10 chickens between the first and second interview, while the best-off farmer will gain an additional 40.

They decide that the minimum detectable difference is 5 – this is the minimum number of additional chickens that they want to be able to reliably detect. They select a confidence level of 95%. Since there are 5000 households in the districts where they work, they put '5000' in the population. This gives an overall sample size of 35 households. These 35 households will be interviewed twice, once before and once after the intervention.

Remember that the programme can normally not assume that any change they observe is due to their intervention. They typically need to gather additional data to check this, as discussed in the DCED's [guidance on attribution](#).

4.6 Advanced options and assumptions

4.6.1 Advanced options

This calculator currently gives you the ability to enter in a different sample size. Based on this, it will tell you what the minimum detectable difference or margin of error (depending on the test) would be.

The calculator works when you want to work out average values in a population. If you want to estimate the **proportion** of a population who does something (such as the % of the population who owns cows), simply put the range as 0 to 200. The other options will be filled in as normal.

A common study design in private sector development is called [difference in difference](#). You can use this calculator to assess the necessary sample sizes for a difference in difference study. To do this, first select the option for a comparative study (above). In the range, enter in the amount that the most successful person in the treatment group is likely to benefit by, and the amount that the least successful person is likely to benefit by. In the minimum detectable difference box, enter in the minimum difference that you would be satisfied detecting between the control and treatment group. For more guidance, feel free to contact us directly.

We are thinking of putting in more advanced options in future. If you would find anything particularly helpful, please email us at Results@Enterprise-Development.org.

4.6.2 Assumptions

This calculator assumes a [normal distribution](#). This means that the values are equally distributed around the average, with the majority closer to the mean and a smaller number some distance away. Based on the normal distribution and the range entered, the calculator guesses what the likely mean and [standard deviation](#) of the sample are.