Terminology

In this video, we're going to be talking about the design terminology that we use whenever we're surveying populations by sampling them. The first definition that we're going to use is an element. And an element is an individual observation of interest. And we make reference to elements based on the population and the sample, as these may not be the same. So a population element is one from your wider population, and your sampling element is that from the sample that you've actually been able to take from.

Our target population, then, is this wider group that we would like to make inferences about. And our sample is the group that we actually got to observe. In an ideal world, these two will be the same. However, for many reasons, for example, the size of the population, the cost of sampling the population, or the difficulty of accessing the population, this might not always be the same in practice.

So the target population is the wider group about which we are making inferences. And our survey population, then, our second definition for populations, is those that we actually have access to. So think about this as every element of the population that we can actually observe and actually sample from or survey. This might be the same as our target population.

But, usually, it won't be. It might be a more limited but more accessible population, whose properties we hope we can extrapolate or generalise towards the target population. An example of this would be looking at the voting behaviours of over 65s in Scotland. That is a very large population. And there's no way that we're going to be able to sample every single person.

We won't be able to contact them. And even if we could, they're not all going to give us the data that we need. So the survey population at Strathclyde could potentially be those people in Glasgow and the surrounding area. And we would hope that that sample is representative enough to reflect the wider opinions of those in Scotland.

However, care must be taken whenever we are generalising because we have to make sure

that this is actually the case. And in this example where we've got Glasgow attempting to be generalised to the entirety of Scotland, we probably cannot be sure of that due to the very specific characteristics of the Glasgow population, which probably isn't going to be mirrored in every other part of Scotland. The next definition that we're going to be describing is a characteristic. And that is a variable of interest of our elements, so any quality that could vary. It'll be usually made with reference to the population or the sample. So a characteristic, for example, could be debt.

A population parameter then is mathematical function applied to these characteristics across the whole population. So, for example, the average debt of UK graduates. The debt is the characteristic that we're studying, the feature of the population. And our population parameter that we've calculated about the characteristic is the average, be it mean, median or mode.

Another example could be the proportion of the population who are homeless. So being homeless is your characteristic that you're studying of the population, and the proportion is the parameter that you've calculated in respect to this statistic. We then need to look at the sampling frame.

And this is the complete and up-to-date list of all the accessible elements of the population from which we are looking to sample, so who we are sampling from. Usually this will be a subset of the survey population. For example, in our voting intentions of over 65s that we discussed, we decided to sample the population of Glasgow.

Our sampling frame would be our voter registration list, the complete and up-to-date list of information about all of the elements of our sample. So from our sample, we will look to make some sample statistics. These are analogous to the population parameters. So just like for population parameters how we used a mathematical function to define something in relation to a characteristic of interest, we're doing the same here using our statistical methods to calculate those with respect to our sample. So for example, the average debt of UK graduates in the city of Glasgow is our sample. The proportion of homeless people in Glasgow would be our sample statistic compared to the proportion of homeless population in the UK, which is the population parameter.

So the key thing to remember is that with populations we reference parameters. With samples, you reference statistics. Now let's delve deeper into the characteristics and definitions related to our sample. So the sampling units are the observations that form the sampling frame, so your individuals who make up the population that you will sample from your sample. Your sample itself then is those sampling units that you actually drew from the sampling frame and actually collected data and information about.

The data then is the values of these variables that we've collected from our sampling units. And the data is also sometimes referred to as the sample. So, here, we have a diagram to

visualise the differences between our definitions of our samples and our populations. So our target population is in the purple square, and that contains everybody. That's our entire population about which we would ultimately like to make inference through designing our survey, distributing it, and analysing it.

Our blue box then is our survey population, which is the population of people who we will realistically be able to work with. Ideally, these will be the same. But as previously discussed, there are instances where this will not be possible. The red square then is our sampling frame, which is a complete and up-to-date list of all of the sampling units, which is a subset of the survey population. There may be some non-overlap between the blue and red boxes if not all of the survey population is contained in the sampling frame that you choose to use.

Finally, our green box is the elements of this sampling frame who we actually sample and actually obtain data on.