# “A” Level Sociology

# Teaching Notes for Students

## Theory and Methods:
### Sampling Techniques

© Chris.Livesey: www.sociology.org.uk 

## Sampling Techniques

- **Simple Random sample**
- **Systematic sample**
- **Stratified Random sample**
- **Stratified Quota sample**
- **Snowball sample**
- **Multi-stage (Cluster) sample**
- **Multi-phase sample.**

**The key ideas here are:**

- Target Population
- Sample Size
- Representative Sample
- Sampling Frame
- Sampling Types
- Sampling Errors

Whenever a sociologist starts a piece of research, they always have in mind a group of people they want to study. These people, whoever they may be, are known as the target population and are, in effect, everyone in a particular group you would like to research.

Your target population might, for example, be a small group (perhaps 10 or 12 people in all) who seem to meet regularly in your local park.

On the other hand, your target population might be the 70,000 football fans who attend Manchester United's home games at Old Trafford.

If you think about these examples, it might be fairly easy to do some research on the first group, since the target population is small. Whether this research involves observing the group from a distance, asking them questions, participating in their behaviour or whatever, the size of the group makes it relatively easy to manage your research.

With the second example, however, things might be more difficult, since its size is going to make it very hard for you to personally observe or question everyone in it. This, therefore, is where the concept of sampling comes into its own...

When you, as a sociological researcher, take a sample of a population to study all you are doing is looking at a relatively small number of people who belong to that group.

In the second case, for example, you might decide to choose 100 Manchester United fans to research and, by studying their behaviour, hope to say something about the characteristics or behaviour of all of the fans in your target population.

Whenever you take a sample, therefore, you are immediately choosing a certain number of people from a much larger group, which brings us to the question of sample size; how large or how small, for example, should your sample be?

Although this may, at first sight, seem to be an important question (is a 90% sample too large or a 10% sample too small?) it's one that actually has no correct answer.

Not only is there no correct answer, the question itself can actually be considered largely irrelevant in the context of social research. Of much greater significance than the question of how many - or indeed how few - people you have in your sample is that of representativeness.

This is an important idea which, in basic terms, means:
“Are the people you’ve chosen for your sample representative of your target population?”.

This is important because if your sample group are representative then anything you discover about them can also be applied to your target population.

For example, if you have selected a **representative sample** of Manchester United fans and discovered that 90% of your **sample** are male, then it follows that 90% of your **target population** will also be male.

The concept used by sociologists to describe this idea is that of **generalisation**; in simple terms, can the things we say about the people in our **sample** also be said about the people in our **target population**?
If the answer is yes, then we are able to make **general statements** about a group of people we ***haven’t*** studied (our **target population**) based on the behaviour of a group of people we ***have*** studied (our **sample** of the target population).

As a general "rule of thumb" when doing sociological research you should try to make your sample representative of your target population. However, there are times when you might deliberately choose not to study a representative sample.

For example, in some types of social research you might **not** want to make **generalisations** about a very large group of people based only on a very small group.

You might, for example, simply be interested in the sociological characteristics of the group itself, rather than what they may or may not represent.

This is particularly true of a research technique called **case studies**.

In a **case study** the objective is to study, in great detail, the characteristics of a particular group (or “case”). This might involve, for example, joining a gang of young males, living amongst a group of monks in a monastery or studying the prescribing practices of doctors in a particular part of the country. In this respect the researcher is **not particularly concerned** about whether or not the group being studied is **representative** of all other, similar groups. In effect, therefore, the **sample** in this type of research is the **target population**.

This is a perfectly acceptable form of research - just as long as the researcher doesn't try to claim that the findings about one group are applicable to all other, similar, groups (in reality, of course, there's a good chance that if the groups are very similar, the findings in relation to one group probably will be broadly applicable to another group).
A second type of non-representative sample can be characterised as a "best opportunity" sample.

What this means is that, when testing an idea you deliberately choose a sample that will provide the best possible opportunity to show that the idea is true.

If your research shows the idea to be false for this "best opportunity" group then it follows that there is a high probability that it will be false for any other relevant groups.

Finally, there may be instances when it is simply not possible to construct a representative sample for a group you want to study.

An example here might be research into "secretive" organisations such as some religious groups who refuse to disclose details of their membership to "outsiders". If you want to research such groups it will not be possible to know with any degree of certainty that the group members you actually study are representative of everyone who is a member of the organisation.

The name we give to this type of sample is "snowball" or "opportunity" sampling and this is discussed in more detail in the section on "Types of Sampling".

### Examples of Non-Representative Sampling

1. "The Affluent Worker In The Class Structure": Goldthorpe, Lockwood and others.

In the early 1960's, Goldthorpe and Lockwood wanted to test the idea, (put forward by Zwieg amongst others), that Britain was gradually becoming a middle class society. That is, the idea (termed "embourgeoisement" or "becoming middle class") that class distinctions such as "upper class", "middle class" and "working class" were losing their significance in a modern, affluent ("well-off") society. In effect, they wanted to test the idea that, to all intents and purposes, working class groups were indistinguishable from middle class groups.

To test this theory they selected a "best opportunity" sample of highly-paid car assembly workers at the Vauxhall factory in Luton. They chose this group as their sample on the basis that if any working class group was likely to show attitudes, beliefs, lifestyles and so forth that were indistinguishable from middle class lifestyles it would be this group of "affluent workers".

By choosing these (unrepresentative) workers as their sample Goldthorpe and Lockwood were effectively saying that if these people did not show evidence of embourgeoisement then it was highly unlikely that any other, less affluent, working class would show evidence of "becoming middle class".

Thus, Goldthorpe and Lockwood deliberately chose an unrepresentative group for their study in order to save time, money and effort. By choosing a "best opportunity" sample it meant that they did not have to sample all working class groups.

Their study concluded that the assembly-line workers were actually very different to their middle class peers who worked in administrative and managerial positions in the factory. By showing this, therefore, they claimed that the theory of embourgeoisement was not applicable to all working class groups in our society at that time.

2. Roy Wallis.

Wallis was interested in studying a religious group called The Church of Scientology ("Scientologists"). The leaders of this group, however, refused to co-operate with Wallis' requests for information. In order to study Scientology, therefore, Wallis was forced to find ex-members of the group who could put him in touch with current members. In this way Wallis was able to build-up a (non-representative) sample of Church members to study.

## Sampling Frame and Sampling Units.

If we assume, for the sake of argument, that sociologists generally try to construct representative samples, it follows that a researcher will normally need some way of identifying everyone in their target population so that an accurate, representative, sample can be taken.

This list is called a sampling frame and the individuals, groups or other phenomena (you could, for example, have a target population of magazines aimed at a particular age group that you are going to sample) are known as sampling units.

Examples of sampling frames might be:

- Electoral Roll / Register - provides a list of everyone eligible to vote.
- School Registers - provide lists of children attending school.
- Professional Membership Lists - organisations such as the British Medical Association (BMA) keep a register of all doctors in Britain.
- Company payrolls - provides a list of all employees in a company.

For most types of sampling (there are exceptions) a sampling frame is essential for two main reasons:

1. If a researcher can't identify everyone in their target population it's unlikely that their sample will be representative of that population.

2. If a researcher is to make contact with the people in their sample, they clearly need to know who they are...

However, just because a list / sampling frame exists, it doesn't mean a researcher will automatically have access to it.

For example, it is quite possible that access to a sampling frame may be denied on the basis of:

**Legal reasons.**

A school, for example, is unlikely to give an outside researcher access to their registers.

**Confidentiality.**

A business organisation is unlikely to give an outside researcher access to their payroll records.

**Secrecy**.

Some religious groups, political parties and so forth do not want outside researchers to study their activities.

**Complete the following questions**: (* Delete those words that do not apply)

1. An appropriate sampling frame for research on the voting intentions of people in Dorset would be

It would be *easy / difficult / impossible to get access to this sampling frame because:

2. An appropriate sampling frame for research into families with new-born babies in my town would be

It would be *easy / difficult / impossible to get access to this sampling frame because:

3. An appropriate sampling frame for research into registered drug-addicts in Newcastle would be

It would be *easy / difficult / impossible to get access to this sampling frame because:

4. An appropriate sampling frame for research into students in my school / college would be

It would be *easy / difficult / impossible to get access to this sampling frame because:

5. An appropriate sampling frame for research into Members of Parliament in Britain would be

It would be *easy / difficult / impossible to get access to this sampling frame because:

**In your own words, outline the difference between representative and non-representative forms of sampling. Please include an example of each in your answer.**

**Give one reason why a researcher might use a representative sample.**

**Give one reason why a researcher might use a non-representative sample.**

## Types of Sampling

### 1. Simple Random Sampling

This is one of the most basic (simple) forms of sampling, based on the probability that the random selection of names from a sampling frame will produce a sample that is representative of a target population. In this respect, a simple random sample is similar to a lottery:

- Everyone in the target population is identified on a sampling frame.
- The sample is selected by randomly choosing names from the frame until the sample is complete.

For example, a 20% sample of a target population of 100 people would involve the random selection of 20 people to be in the sample.

An example of a simple random sample you could easily construct would be to take the names of every student in your class from the register, write all the names on separate pieces of paper and put them in a box.

If you then draw out a certain percentage of names at random you will have constructed your simple random sample…

### 2. Systematic Sampling

A variation on the above is to select the names for your sample systematically rather than on a simple random basis. Thus, instead of putting all the names on your sampling frame individually into a box, it's less trouble to select your sample from the sampling frame itself.

For example, if you were constructing a 20% sample of a target population containing 100 names, a systematic sample would involve choosing every fourth name from your sampling frame.

A simple example of a systematic sample would be for you to use a class register to construct a sample of students in your class.

You could try constructing a 50% sample of your class using this sampling technique.

This type of sampling technique tends to be used when the target population is very large.

For example, if you were going to select a 10% sample from a target population of 1 million people you would either need a very large box and a lot of patience or a computer and some means of getting the names in your target population into a program that would select your sample randomly

Both of the above types of sampling are very similar, but the main difference between them is that while simple random samples are truly random (everyone in the target

population has an equal chance of being chosen for the sample), systematic samples are only near-random (or pseudo-random).

Systematic samples are near-random because if you have a list of 100 people, for example, and you start to select a 20% sample beginning with the first name on your sampling frame, the second, third and fourth names actually have no chance of being included in your sample...

When deciding which of these two types of sampling to use, their basic advantages and disadvantages are very similar and we can summarise them in the following terms:

**Uses:**

1. Both are relatively quick and easy ways of selecting samples (if the target population is reasonably small).

2. They are random / near random, which means that everyone in the target population has an equal chance of appearing in the sample (this is not quite true of systematic sampling, but such samples are “random enough” for most research purposes).

3. They are both reasonably inexpensive to construct. Both simply require a sampling frame that is accurate for the target population.

4. Other than some means of identifying people in the target population (a name and address, for example), the researcher does not require any other knowledge about this population (an idea that will become more significant when we consider some other forms of sampling)

**Limitations**

1. The fact that these types of sample always need a sampling frame means that, in some cases, it may not be possible to use these types of sampling. For example, a study into “underage drinking” could not be based on a simple random or systematic sample because no sampling frame exists for the target population.

2. In many cases a researcher will want to get the views of different categories of people within a target population and it is not always certain that these types of sampling will produce a sample that is representative of all shades of opinion.

For example, in a classroom it might be important to get the views of both the teacher and their students about some aspect of education. A simple random / systematic sample may not include a teacher because this category is likely to be a very small percentage of the overall class; there is a high level of probability that the teacher would not be chosen for by any sample that is simply based on chance…

One way of trying to overcome some of the potential limitations of simple random / systematic sampling is to use an alternative sampling technique that avoids the problem of possible under-representation, while retaining the idea of selection based on chance.

This type of sampling is called Stratified sampling and we can look briefly at this idea next.

### 3. Stratified Random Sampling.

As a simple way of illustrating the idea of a stratified random sample, consider the following research problem.

A researcher wants to study the views of a group of sixth-form students about their education.

The school they choose to study has 60 students in its first year (40 males and 20 females) and 60 students in its second year (10 males and 50 females).

A 20% sample will be drawn from this target population (24 students) and the views of 1st year females (20) are just as important to the researcher as the views of 2nd year males (10).

The problem, therefore, is how can the researcher be sure that all possible views will be represented in the sample?

If the researcher used a simple random or systematic sample it is possible that some minority views (for example, 2nd year males) might, at best, be under-represented and, at worst, not represented at all, in the sample. To overcome this problem, therefore, we need to look at the idea of "a sample" in a different way.

Instead of simply drawing one sample from the target population, why not break the target population down into the different groups we want represented in the final sample?

The obvious thing to do in this situation is to decide which groups we want represented and then draw a number of smaller, representative, samples that can be combined to give us our final sample. For example, the views we want represented are year 1 males and females and year 2 males and females.

This gives us four sampling frames and four samples to take.

To ensure these samples are representative, we have to choose names for our sample from each in direct proportion to their representation in the target population. Thus:

**1. First and second year students.**

To achieve this, we create **two sampling frames**.

**Frame 1** consists of the names of **60 first year** students.

**Frame 2** consists of the names of **60 second year** students.

**2. Males and females.**

The **first sampling frame** (**first year** students) is split into **two** further sampling frames (**males and females**).

The **second sampling frame** (**second year** students) is also split into **two** further sampling frames (**males and females**).

We then choose the names for each sample on a random basis and, when this has been done, we simply combine the four samples to give us the overall sample we require that will be representative of the target population.

**Year One**

**Sample 1** consists of **8 first year males**

**Sample 2** consists of **4 first year females**

**Year Two**

**Sample 3** consists of **2 second year males**

**Sample 4** consists of **10 second year females.**

**Briefly explain why stratified sampling depends on the researcher knowing the characteristics of the target population sampled.**

### 4. Stratified Quota

This sampling technique is a variation on the stratified random type of sampling we've just discussed. As such, the basic principles involved in the sampling process are exactly the same (the division of the main sample into smaller samples). There is, however, one major difference and that comes at the stage when the people in the sample are actually selected. Rather than selecting people randomly, quota samples allow the researcher to select people non-randomly.
A brief example should make this idea clear.

Imagine you want to question people about their voting behaviour. For the sake of illustration, you decide your sample will be stratified in the following way:

1. By Gender:

Males and Females.

2. By Age:

Males 18 - 64 and males 65+. ]
Females 18 - 64 and females 65+.

In effect, therefore, you will initially be conducting four small samples that can then be combined into a sample that is representative of your target population.

You are going include 100 people in your sample and, based on their proportions in the target population, you find you need to question:

40 males aged 18 - 64
5 males aged 65+
40 females aged 18 - 64
15 females aged 65+.

These age and gender stratified groups are your four quota samples and you decide to select the sample by going down to your local shopping centre and asking people if they would answer questions about their voting behaviour.

- The first 40 males aged 18 - 64 that you get to agree to answer your questions are then included in your sample, as are the other quota groups in the proportions you've decided upon.

- Once you've found 40 males aged 18 - 64 who agree to be in the sample you have filled your quota for this group and do not need to include any more men of this age.

- When every quota has been filled, a sample will have been drawn that should be representative of the target population.

The main point to note here is that selection for the sample will not have been on a truly random basis (since to be included you had to be in the shopping centre on a particular day and at a particular time), but it should be a sample that is "random enough" for research purposes.

The following outlines some of the uses and limitations of stratified random and stratified non-random sampling.

**Uses**

1. This type of sampling ensures that **known differences** in the target population will be accurately reflected in the sample.

In basic terms, therefore, we can be sure that in terms of the characteristics of our target population our sample will be **broadly representative**.

2. Stratified samples do not have to be very big, since it is possible, (using **small samples** that are carefully stratified), to make certain that we have accurately reflected the make-up of our target population.

3. **Stratified Quota** samples, in particular, are usually relatively cheap and quick to construct accurately.

**Limitations**

1. In order to stratify a sample the researcher must have **accurate** and **up-to-date information** about the target population. This is not always available.

2. Even in situations when a researcher has accurate information about the different groups that make-up the target population it is possible that this information may be **out-of-date** by the time the research based on the sample is actually conducted. This is especially true where the sample is **large and complex** and in situations where the composition of the target population may **change rapidly** and consistently.

**For example**, where age-groups are used, these will actually change on a daily basis.

3. If you are employing a team of researchers to help you construct a quota sample you **cannot be certain** that they correctly placed everyone in the right quota category. If, for example, your research assistant cannot find “100 men over the age of 65” to fill their quota, there may be a temptation to fill it using men under that age…

4. The fact that stratified quota sample selection is **not truly random** may mean it is **not representative** of a **target population.**

**Briefly outline at least two major advantages of a stratified sample compared to a simple random or systematic sample:**

### 5. Snowball ("Opportunity") Sampling.

We looked earlier at the idea of non-representative sampling and mentioned briefly the idea of "snowball" or "opportunity" sampling.

Just as a snowball rolling downhill gets larger and larger as it picks-up more snow, a "snowball sample" picks-up more and more to be in the sample over time.

We can now briefly develop this idea.

It is not always possible for a researcher to get hold of a sampling frame for a target population. This may be because such a list doesn't exist or because someone who controls access to the list will not release it to a researcher. Whatever the reason, it may still be possible to construct a sample in a ad hoc (unsystematic) way.

As the name suggests, a snowball or opportunity sample involves the researcher identifying someone in the target population who is willing to be researched.

This person may then suggests another 2 or 3 people (perhaps more) who will help. These people, in turn, suggest another 2 or 3 people until, in a relatively short space of time, the researcher has a sample they feel they can use in their research.

Clearly, this type of sampling is not going to produce a sample that is truly representative of a target population, but it may be the best that can be achieved in certain situations.

If you use this type of sampling in your (project) work, please pay careful attention to the limitations of this technique...

We can note the following uses and limitations of this sampling technique.

**Uses**

1. This type of sampling enables a researcher to construct a sample in situations where it would not be possible to do so using any other sampling technique.

2. It can be a relatively cheap and quick method of sampling.

**Limitations**

1. The sample is unlikely to be representative of a target population.

2. There is no way of checking whether or not your sample is representative.

3. There is a high likelihood of a self-selected sample being constructed (see below: Sampling Errors).

### 6. Multi-stage (Cluster) Sampling.

This form of sampling is usually done when a target population is spread over a wide geographic area.

- For example, an opinion poll into voting behaviour may involve a sample of 1000 people to represent the 35 million people eligible to vote in a General Election. If a simple random sample were taken it's possible that the researcher might have to poll 10 people in Newcastle, 15 people in Cardiff, 3 people in Bournemouth and so forth. In other words, it would be a time-consuming and very expensive process and the results from the poll would probably be out-of-date before the poll could be finished.

- To avoid these problems, a researcher can use a multi-stage / cluster sample that firstly, divides the country into smaller units (in this example, electoral constituencies) and then into small units within constituencies (for example, local boroughs). Boroughs could then be selected which, based on past research, show a representative cross-section of voters and a sample of electors could be taken from a relatively small number of boroughs across the country.

We can note the following uses and limitations with this type of sampling:

**Uses**

1. This type of sample saves the researcher time and money.

2. Once a relatively reliable sample has been established, the researcher can use the same or a similar sample again and again (as with political opinion polling).

**Limitations**

1. Unless great care is taken by the researcher it is possible that the cluster samples will not be representative of the target population.

2. Even though it is a relatively cheap form of sampling, this is not necessarily the case. A sample that seeks to represent the whole of Britain, for example, is still going to be too expensive for many researchers

**7. Multi-phase Sampling.**

This type of sampling depends on a sample of some description already having been taken and simply involves the idea that the researcher takes a "sample from their sample". One reason a researcher might want to do this is to follow-up any ideas raised by research on the larger sample and, rather than re-question the whole sample a smaller, more-selective, group is chosen to represent the sample of the target population.

Having looked at the major types of sampling techniques used by sociologists, we can complete this unit by considering two of the more common forms of sampling error that, if present in our sample, may make them (accidentally or deliberately) unrepresentative.
**Sampling Errors**

### 1. The Self-Selected Sample.

One of the most common forms of sampling error is the creation of a sample that effectively "selects itself" rather than being selected by the researcher. In simple terms, rather than the researcher selecting people to appear in a representative sample, people choose, in some way, to be in the sample. This type of error is fairly common in "research" that is not very rigorous or systematic in its approach; for example, the type of opinion polls that appear in newspapers and magazines almost invariably involve a self-selected (and unrepresentative) sample. For example:

A newspaper that asks its readers to respond to the question,

"**Should people convicted of murder be given the death penalty?**"

will **always** produce a **biased**, **unrepresentative**, **sample** and, as such, we have no way of knowing whether or not the answer we get from such a poll represents "public opinion" in the whole of our society.

The reasons for this lack of representativeness are not hard to find:

> Firstly, only a small minority of the population will buy the newspaper on the day the poll appears. People who do buy the newspaper have, however unwittingly, selected themselves for the sample.
>
> Secondly, an unknown number of readers will not notice the poll (and so cannot vote in it). Those who notice the question, therefore, have again selected themselves for the sample.
> Thirdly, only a small proportion of the newspaper's readers will bother to respond to the question. This proportion is made even smaller if the respondent (the person who takes part in a piece of research) has to pay to vote (for example, by making a telephone call at their own expense).
>
> Finally, the people who do respond to such polls are invariably those who have very strong views either way on the question (in this example, people who are strongly pro- or strongly anti-Capital Punishment) - and these are unlikely to be representative of the population of Britain.

This fictional example simply illustrates the idea of a self-selected sample.

## 2. Statistically-Inadequate Samples

Earlier I noted that the question of how large or how small a sample should be is not as important as the question of how representative the sample is.

This is true up to a point, but it is evident that a sample that is too small to accurately represent a target population is going to be inadequate for research purposes.

To illustrate this idea, consider the following example.

If you flip a coin, there is always an equal chance that it will land either “heads” or “tails”. This is a random process and, because of this, the more times you flip the coin the greater is the probability that you will end up counting an equal number of “heads” and “tails”.

However, because it is a random process it is possible that the coin may on occasions land “heads” a few times before it lands “tails”. The more times the coin is flipped, however, the greater the probability that these kinds of “error” will even themselves out. If you only flip the coin twice, therefore, it may land “heads” each time. If you flip the coin 100 times, on the other hand, it is highly probable that it will land “heads” 50 times and “tails” 50 times.

Try it and see…

As a general rule, therefore, the larger your sample as a proportion of your target population the greater the probability that it will be statistically-adequate. This may improve the chances of your sample being representative of the target population, but a large sample is no guarantee of a representative sample...

**“Goodbye” magazine asked a representative sample of its readers whether or not they thought Tony Blair was doing a good job as Prime Minister. Briefly explain why the findings of this research will not be an accurate reflection of public opinion**