

1.1 Populations and samples

- **In statistics, a population is the whole set of items that are of interest.**

For example, the population could be the items manufactured by a factory or all the people in a town. Information can be obtained from a population. This is known as raw data.

- **A census observes or measures every member of a population.**
- **A sample is a selection of observations taken from a subset of the population which is used to find out information about the population as a whole.**

There are a number of advantages and disadvantages of both a census and a sample.

| | Advantages | Disadvantages |
|---------------|--|---|
| Census | <ul style="list-style-type: none"> • It should give a completely accurate result | <ul style="list-style-type: none"> • Time consuming and expensive • Cannot be used when the testing process destroys the item • Hard to process large quantity of data |
| Sample | <ul style="list-style-type: none"> • Less time consuming and expensive than a census • Fewer people have to respond • Less data to process than in a census | <ul style="list-style-type: none"> • The data may not be as accurate • The sample may not be large enough to give information about small sub-groups of the population |

The size of the sample can affect the validity of any conclusions drawn.

- The size of the sample depends on the required accuracy and available resources.
- Generally, the larger the sample, the more accurate it is, but you will need greater resources.
- If the population is very varied, you need a larger sample than if the population were uniform.
- Different samples can lead to different conclusions due to the natural variation in a population.
- **Individual units of a population are known as sampling units.**
- **Often sampling units of a population are individually named or numbered to form a list called a sampling frame.**

Each method of random sampling has advantages and disadvantages.

| Simple random sampling | |
|---|---|
| Advantages | Disadvantages |
| <ul style="list-style-type: none"> • Free of bias • Easy and cheap to implement for small populations and small samples • Each sampling unit has a known and equal chance of selection | <ul style="list-style-type: none"> • Not suitable when the population size or the sample size is large • A sampling frame is needed |

| Systematic sampling | |
|---|---|
| Advantages | Disadvantages |
| <ul style="list-style-type: none"> • Simple and quick to use • Suitable for large samples and large populations | <ul style="list-style-type: none"> • A sampling frame is needed • It can introduce bias if the sampling frame is not random |

| Stratified sampling | |
|---|---|
| Advantages | Disadvantages |
| <ul style="list-style-type: none"> • Sample accurately reflects the population structure • Guarantees proportional representation of groups within a population | <ul style="list-style-type: none"> • Population must be clearly classified into distinct strata • Selection within each stratum suffers from the same disadvantages as simple random sampling |

1.3 Non-random sampling

There are two types of non-random sampling that you need to know:

- Quota sampling
- Opportunity sampling

■ **In quota sampling, an interviewer or researcher selects a sample that reflects the characteristics of the whole population.**

The population is divided into groups according to a given characteristic. The size of each group determines the proportion of the sample that should have that characteristic.

As an interviewer, you would meet people, assess their group and then, after interview, allocate them into the appropriate quota.

This continues until all quotas have been filled. If a person refuses to be interviewed or the quota into which they fit is full, then you simply ignore them and move on to the next person.

■ **Opportunity sampling consists of taking the sample from people who are available at the time the study is carried out and who fit the criteria you are looking for.**

Notation Opportunity sampling is sometimes called **convenience sampling**.

This could be the first 20 people you meet outside a supermarket on a Monday morning who are carrying shopping bags, for example.

There are advantages and disadvantages of each type of sampling.

| Quota sampling | |
|--|--|
| Advantages | Disadvantages |
| <ul style="list-style-type: none"> • Allows a small sample to still be representative of the population • No sampling frame required • Quick, easy and inexpensive • Allows for easy comparison between different groups within a population | <ul style="list-style-type: none"> • Non-random sampling can introduce bias • Population must be divided into groups, which can be costly or inaccurate • Increasing scope of study increases number of groups, which adds time and expense • Non-responses are not recorded as such |

| Opportunity sampling | |
|--|--|
| Advantages | Disadvantages |
| <ul style="list-style-type: none"> • Easy to carry out • Inexpensive | <ul style="list-style-type: none"> • Unlikely to provide a representative sample • Highly dependent on individual researcher |

Summary of key points

- In statistics, a **population** is the whole set of items that are of interest.
 - A **census** observes or measures every member of a population.
- A sample is a selection of observations taken from a subset of the population which is used to find out information about the population as a whole.
 - Individual units of a population are known as **sampling units**.
 - Often sampling units of a population are individually named or numbered to form a list called a **sampling frame**.
- A **simple random sample** of size n is one where every sample of size n has an equal chance of being selected.
 - In **systematic sampling**, the required elements are chosen at regular intervals from an ordered list.
 - In **stratified sampling**, the population is divided into mutually exclusive strata (males and females, for example) and a random sample is taken from each.
 - In **quota sampling**, an interviewer or researcher selects a sample that reflects the characteristics of the whole population.
 - **Opportunity sampling** consists of taking the sample from people who are available at the time the study is carried out and who fit the criteria you are looking for.
- Variables or data associated with numerical observations are called **quantitative variables** or **quantitative data**.
 - Variables or data associated with non-numerical observations are called **qualitative variables** or **qualitative data**.
- A variable that can take any value in a given range is a **continuous variable**.
 - A variable that can take only specific values in a given range is a **discrete variable**.
- When data is presented in a grouped frequency table, the specific data values are not shown. The groups are more commonly known as **classes**.
 - Class boundaries tell you the maximum and minimum values that belong in each class.
 - The midpoint is the average of the class boundaries.
 - The class width is the difference between the upper and lower class boundaries.
- If you need to do calculations on the large data set in your exam, the relevant extract from the data set will be provided.