AS Statistics – Chapter 1 – Data Collection

Types of Data

Data is **qualitative** if it is associated with non-numerical observations (e.g. colour, flavour – think "quality")

Data is **quantitative** if it is associated with numerical observations or measurements (think "quantity")

A **continuous** variable can take any value in a given range (usually a measurement such as length or time)

A **discrete** variable can take only specific values in a given range

Sampling - Key Definitions

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 a population, used to find out information about the population as a whole.

A **sampling unit** is an individual unit of a population.

A **sampling frame** is a list of individually named or numbered sampling units.

Methods of Random Sampling

Simple Random Sampling – every possible sample of the required size has an equal chance of being selected.

For this, you need a sampling frame, usually a list of people or things. Each is allocated a unique number, and a selection of these numbers is chosen at random. There are two methods of choosing numbers:

- generating random numbers (using a computer, calculator or random number table)
- lottery sampling (drawing numbers out of a hat, for example)

Systematic Sampling – the required elements are chosen at regular intervals from an ordered list.

Stratified Sampling – the population is divided into mutually exclusive subsets and a random sample is taken from each. The size of each sample is proportional to the size of the subset.

Methods of Non-Random Sampling

Quota Sampling – an interviewer or researcher selects a sample that reflects the characteristics of the population.

Opportunity Sampling – taking a sample from people who are available at the time and meet your criteria.

Census or Sample?

	Advantages	Disadvantages
Census	• It should give a completely accurate result	• Time consuming and expensive • Cannot be used when the testing process destroys the item • Hard to process large quantity of data
Sample	• Less time consuming and expensive than a census • Fewer people have to respond • Less data to process than in a census	• The data may not be as accurate • The sample may not be large enough to give information about small sub-groups of the population

Advantages and disadvantages of different sampling methods

Simple random sampling	
Advantages	**Disadvantages**
• 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	• Not suitable when the population size or the sample size is large as it is potentially time consuming, disruptive and expensive. • A sampling frame is needed

Systematic sampling	
Advantages	**Disadvantages**
• Simple and quick to use • Suitable for large samples and large populations	• A sampling frame is needed • It can introduce bias if the sampling frame is not random

Stratified sampling	
Advantages	**Disadvantages**
• Sample accurately reflects the population structure • Guarantees proportional representation of groups within a population	• Population must be clearly classified into distinct strata • Selection within each stratum suffers from the same disadvantages as simple random sampling

Quota sampling	
Advantages	**Disadvantages**
• 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	• 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**
• Easy to carry out • Inexpensive	• Unlikely to provide a representative sample • Highly dependent on individual researcher