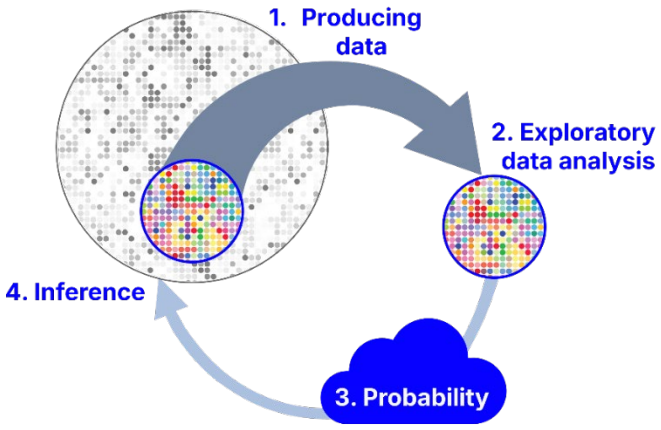
Producing Data

Why does data collection matter?

The quality of statistical conclusions depends on the quality of the data collected. Before analysing data, we must ask:

- Who was studied?
- How were they selected?
- How was the information collected?
- Does the sample represent the population?

Poorly collected data can lead to misleading conclusions, regardless of how sophisticated the analysis is.



Population vs Sample

Statistics use information from a sample to draw conclusions about a population.

	Description	Example
Population	The entire group we want to study	All Federation University students
Sample	A subset of the population	500 students studying education

Characteristics of a good sample

A good sample should be:

- Representative – it reflects the characteristics of the population.
- Unbiased – all members of the population have a fair chance of being selected.

Example:

- Representative: Students randomly selected from multiple campuses and study modes.
- Biased: Only students surveyed in the library during exam week.

Sampling errors

Different samples from the same population may produce slightly different results.

Example

- Sample A → Mean study time = 9.8 hours
- Sample B → Mean study time = 10.4 hours

This natural variation is called sampling error and occurs whenever a sample is used instead of the entire population.

CRICOS 00103D | RTO 4909 | TEQSA PRV12151 (Aust)

Sampling methods

The method used to select participants affects how representative a sample will be.

Probability sampling

Each member of the population has a known chance of being selected.

Advantages:

- Reduces bias and produces more representative samples.

Method	Description	Example
Simple random	Equal chance of selection	Randomly selecting 500 students
Systematic	Select every *n*th member	Every 20th student
Stratified	Sample from important subgroups	Students from each year level
Cluster	Sample entire groups (clusters)	Selecting several tutorial classes

Non-probability sampling

Not every member of the population has a known chance of being selected.

Limitation:

- These methods are more susceptible to bias and may not represent the wider population

Method	Description	Example
Convenience	Easily accessible people	Students in the library
Quota	Set number from each group	50 students per faculty
Purposive	People with specific characteristics	Mature-age students
Snowball	Participants recruit others	Students inviting classmates

Related helpsheets

- The Statistical Process
- Types of Studies
- Types of Data (EDA)

CRICOS 00103D | RTO 4909 | TEQSA PRV12151 (Aust)