

MODULE ONE: PRESENTING AND DESCRIBING INFORMATION

TOPIC 1: DATA COLLECTION





Famous Quotes

- There are three kinds of lies: lies, damned lies and statistics.
— Benjamin Disraeli
- Figures don't lie; liars figure.
— Mark Twain
- Statistics can be used to support anything— especially statisticians.
— Franklin P. Jones
- There are two kinds of statistics, the kind you look up and the kind you make up.
— Rex Stout
- 58.6% of all statistics are made up on the spot
— Unknown

+ Learning Objectives

At the completion of this topic, you should be able to:

- identify how statistics is used in business
- recognise the sources of data used in business
- identify the types of data used in business
- distinguish between different survey sampling methods
- evaluate the quality of surveys

+1.1 Basic Concepts of Statistics

4

Statistics is a large discipline that comprises three broad tasks:

1. collection of data
2. processing and presentation of data
3. analysis and interpretation of data

+1.1 Basic Concepts of Statistics (cont)

Key Definitions

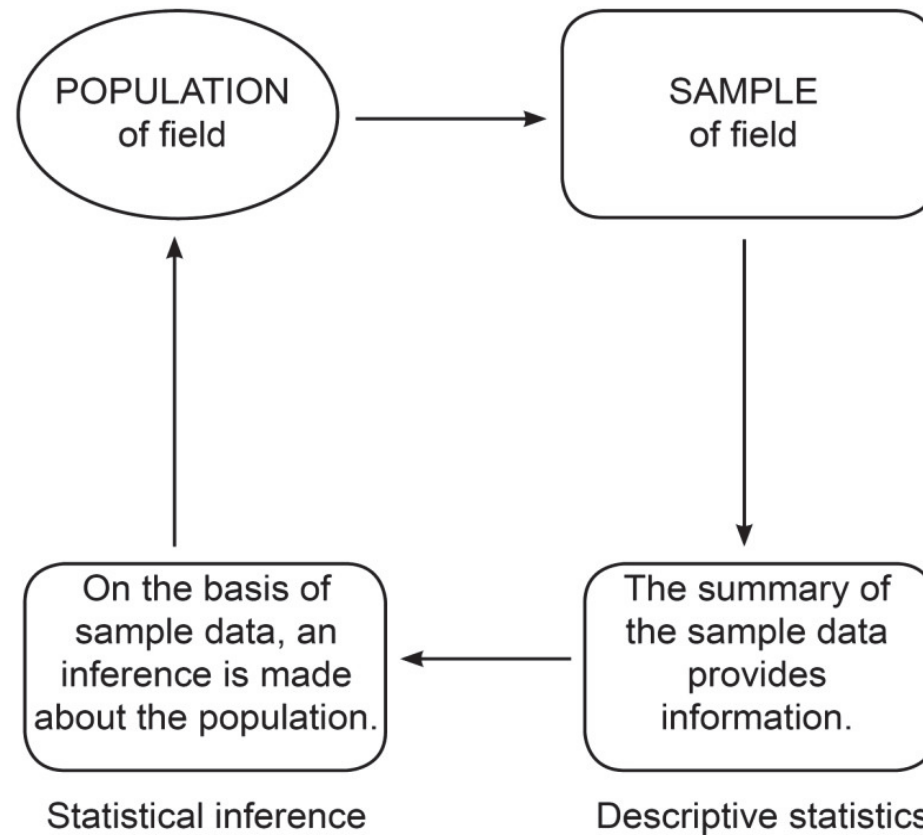
- A **population** consists of all the members of a group about which you want to draw a conclusion
- A **parameter** is a numerical measure that describes a characteristic of a population
- A **sample** is the portion of the population selected for analysis
- A **statistic** is a numerical measure that describes a characteristic of a sample

+1.1 Basic Concepts of Statistics (cont)

As a field of study, statistics can be split into two main groups:

1. *Descriptive statistics*, which relates to a set of techniques based around certain tables, graphs and calculated summary measures used for describing the important features of a given set of data
2. *Inferential statistics*, which relates to the use of sample data to draw inferences and conclusions about the whole population of individuals or items from which the sample was drawn

+1.1 Basic Concepts of Statistics (cont)



+1.3 Collecting Data

8

Identifying Sources of Data:

- External sources
 - Data collected by others (Use if acceptable)

or

- Collect your own
 - Census
 - Sample

+1.3 Collecting Data (cont)

9

Existing Sources

- Within a firm – almost any department
- Business database services – Australian Stock Exchange
- Government agencies – Australian Bureau of Statistics
- Industry associations – Real Estate Institute of Australia
- Special-interest organizations – Graduate Management Admission Council
- Internet – more and more firms and government departments/authorities

+1.3 Collecting Data (cont)

10

There are many issues and potential traps when collecting your own data, including:

- Do we take a Census or a Sample?
- How big a sample size?
- Collection methodology
- Time and cost issue

+1.3 Collecting Data (cont)

Sample surveys: Purpose, Purpose, Purpose

- The broad **purpose** of a cross-sectional survey is:
 - To draw conclusions or make inferences about the **whole** group (**population**) of items or individuals at a given point of time
- For example, a University wishes to determine if student grades are being adversely affected by work at part-time jobs
- We must ALWAYS keep the **purpose** in mind

+1.3 Collecting Data (cont)

12

Given a purpose and desired outcome, we first need to **collect** relevant data

Collecting **high quality** data is arguably the most difficult part of a statistical exercise

Two options:

- Census: We investigate the whole population
- Sample: Investigate some of the population

But the purpose is **the same**

+1.3 Collecting Data (cont)

Census vs Sample:

- Census
 - In theory, more accurate
 - But time consuming and expensive
- Sample
 - Saves time and money
 - Only option if items have to be destroyed, and for some types of experiments
 - Can provide a very high level of accuracy

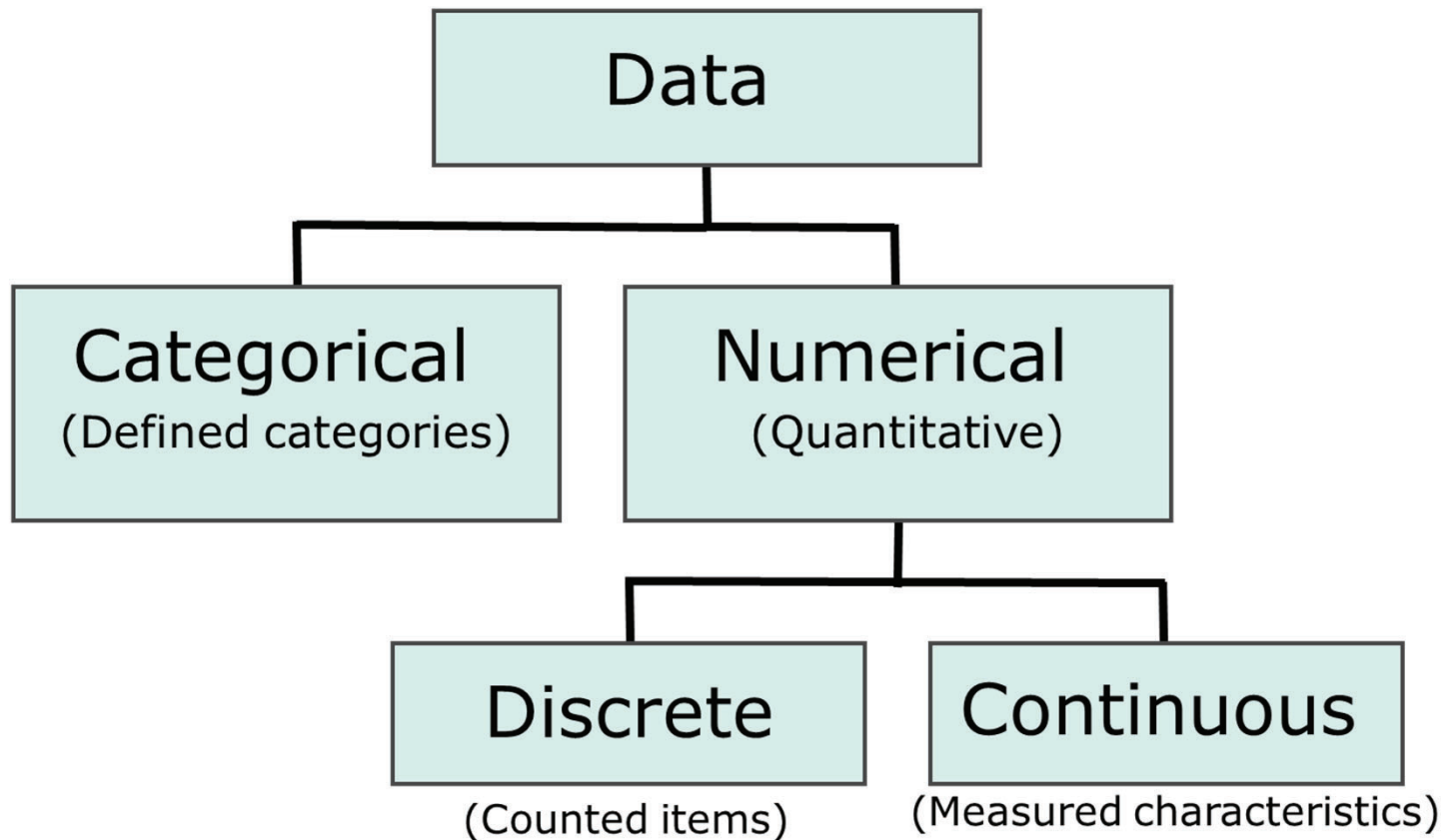
+1.3 Collecting Data (cont)

14

Census vs Sample

- Both census and sample require similar attention to detail in terms of:
 - survey type used
 - questionnaire design
 - training of interviewers
 - etc.
- Taking a sample has the added complication of choosing the sample

+1.4 Types of Variables - Data

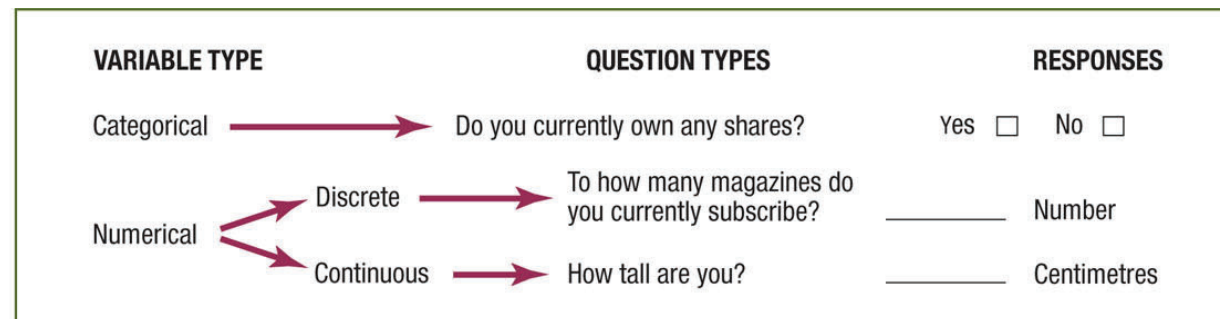


+1.4 Types of Variables – Data (cont)

16

Figure 1.1

Types of variables



+Levels of Measurement and Types of Measurement Scales

CATEGORICAL VARIABLE	CATEGORIES
Personal computer ownership	Yes <input type="checkbox"/> No <input type="checkbox"/>
Type of fuel used	Unleaded <input type="checkbox"/> Premium Unleaded <input type="checkbox"/> Diesel <input type="checkbox"/> LPG <input type="checkbox"/>
Internet connection	Cable <input type="checkbox"/> Wireless <input type="checkbox"/>

Figure 1.2 Examples of nominal scaling

CATEGORICAL VARIABLE	ORDERED CATEGORIES
Product satisfaction	Very unsatisfied Fairly unsatisfied Neutral Fairly satisfied Very satisfied
Clothing size	S M L XL
Type of Olympic medal	Gold Silver Bronze
Education level	Primary Secondary Tertiary

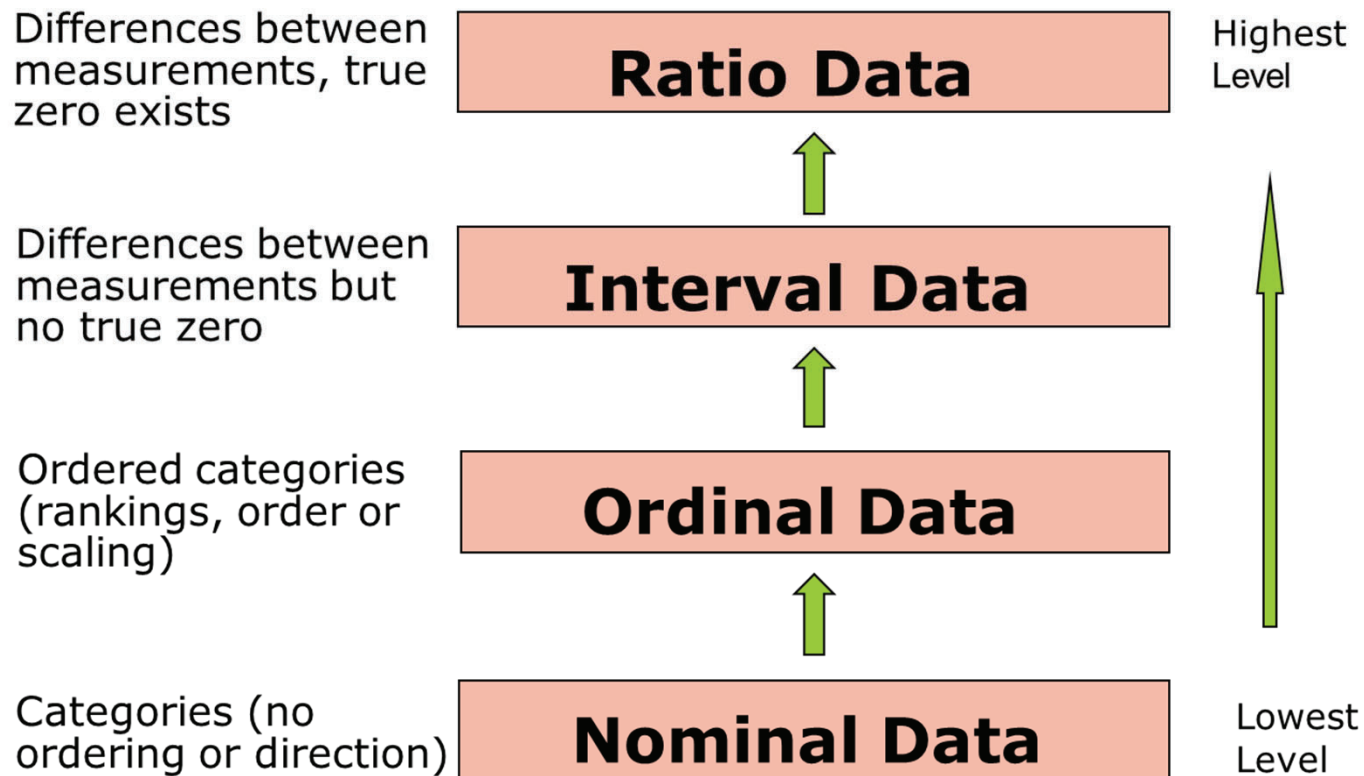
Figure 1.3 Examples of ordinal scaling

NUMERICAL VARIABLE	LEVEL OF MEASUREMENT
Shoe size (UK or US)	Interval
Height (in centimetres)	Ratio
Weight (in kilograms)	Ratio
Salary (in US dollars or Japanese yen)	Ratio

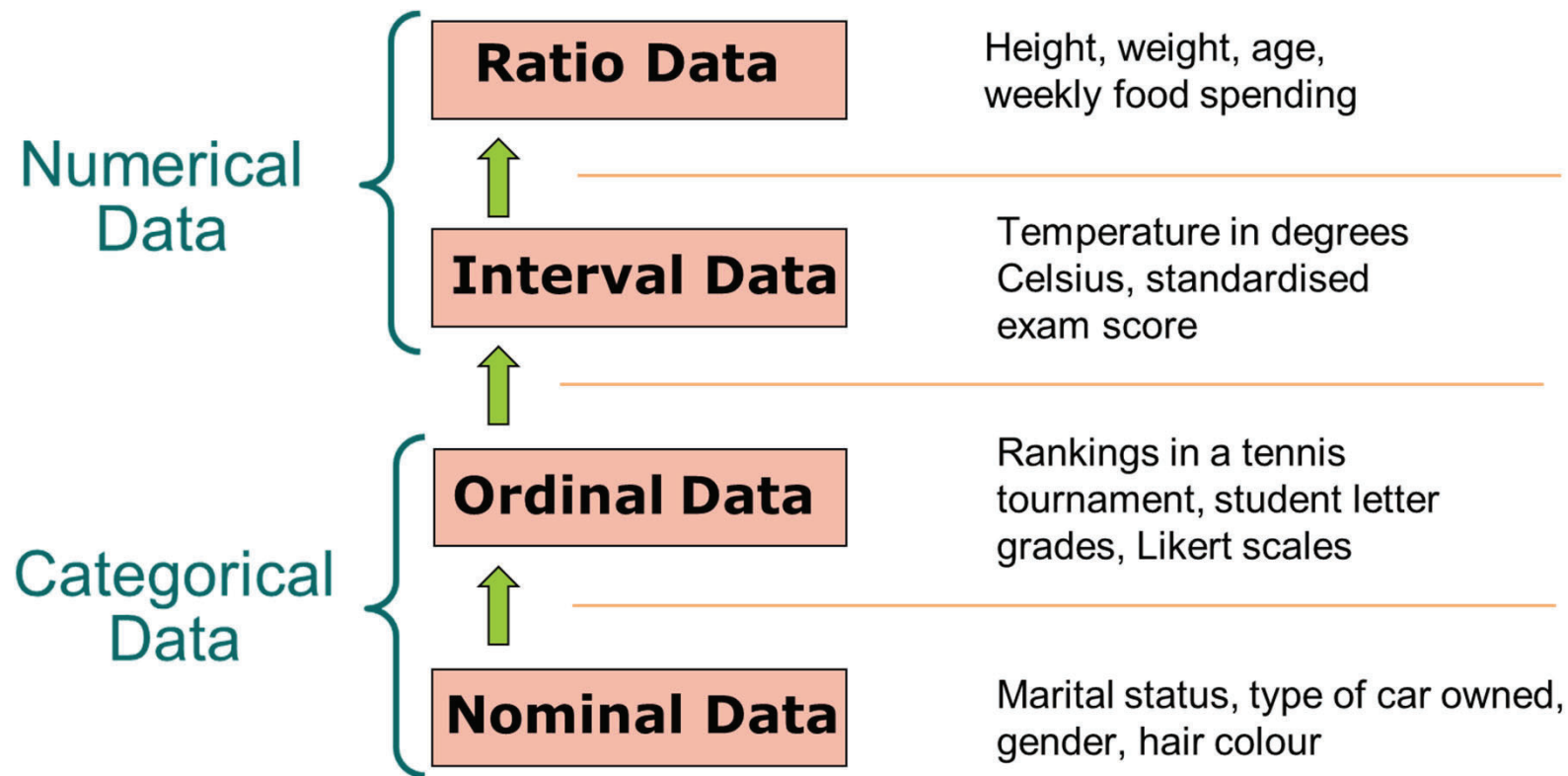
Figure 1.4

Examples of interval and ratio scales

+Levels of Measurement and Types of Measurement Scales



+Levels of Measurement and Types of Measurement Scales



+Levels of Measurement and Types of Measurement Scales

Nominal

- Data has **labels** or **names** used to identify an attribute of the element
- A nonnumeric label or numeric code may be used
 - Eg: Nominal data that would relate to employees

Gender	Code
Female	0
Male	1
Female	0

+Levels of Measurement and Types of Measurement Scales

Ordinal

- The data have the properties of nominal data and the **order** or **rank** of the data is meaningful
- A nonnumeric label or numeric code may be used

+Levels of Measurement and Types of Measurement Scales

Interval

- The data have the properties of ordinal data, and the interval between observations is expressed in terms of a **fixed unit of measure**
- Interval data are always numeric
 - Eg: Temperature – today is 20C and yesterday was 24C. The difference between them is 4C (Celsius)

+Levels of Measurement and Types of Measurement Scales

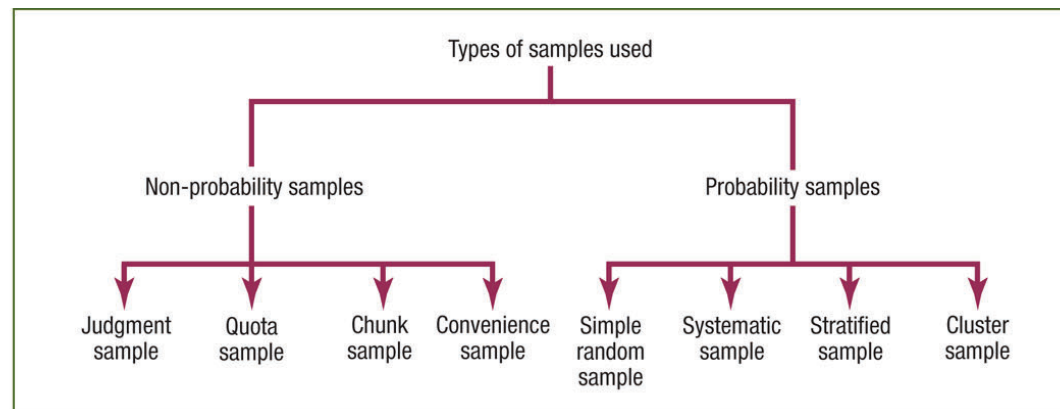
Ratio

- The data have all the properties of interval data and the **ratio of two values** is meaningful.
- This scale must contain **a zero value that indicates that nothing exists** for the variable at the zero point.
- Eg: Ratio data that would relate to employees

Salary	Years Emp
\$43,000	2
\$72,000	3.5
\$48,500	12

+7.4 Types of Survey Sampling Methods

.....
Figure 7.5 Types of samples



Non-probability sample

- Items included are chosen without regard to their probability of occurrence

Probability sample

- Items in the sample are chosen on the basis of known probabilities

+Simple Random Sample

- Every individual or item from the frame (N) has an equal chance of being selected ($1/N$)
- Selection may be with replacement or without replacement
- Samples can be obtained from table of random numbers or computer random number generators
- Simple to use but may not be a good representation of the population's underlying characteristics

+Systematic Sample

26

- Divide frame of N individuals into n groups of k individuals:
 $k = N/n$
- Randomly select one individual from the 1st group
- Select every k th individual thereafter
- Like simple random sampling, simple to use but may not be a good representation of the population's underlying characteristics

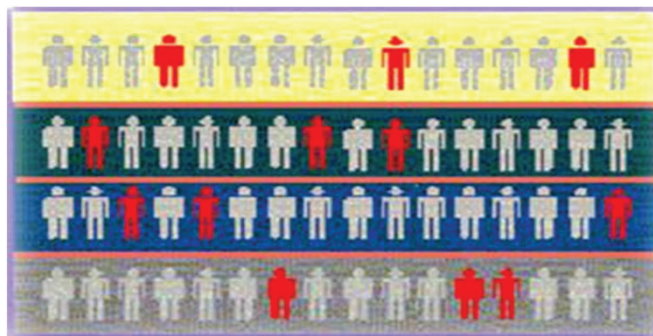
+Systematic Sample (cont)

- Divide our frame of 64 into 8 groups with 8 people in each group
- Randomly select one individual from the 1st group, say the third person and then select every 8th person after that



+Stratified Sample

- Divide population into two or more subgroups (called strata) according to some common characteristic
- A simple random sample is selected from each subgroup, with **sample sizes proportional to strata sizes** (called proportionate stratified sampling)
- Samples from subgroups are **combined into one**



Strata 1 (size 16) → 3

Strata 2 (size 16) → 3

Strata 3 (size 16) → 3

Strata 4 (size 16) → 3

Sample size = 12

$$16:16:16:16 = 3:3:3:3$$

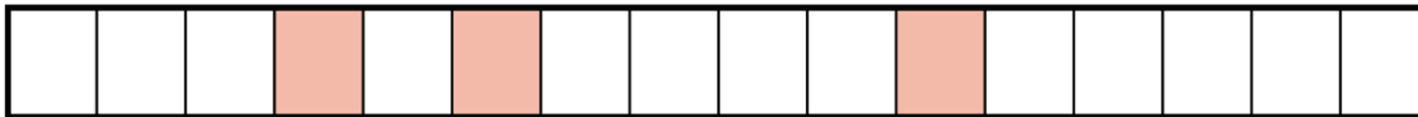
+Stratified Sample (cont)

- More **efficient** than simple random sampling or systematic sampling because of **assured representation** of items across entire population
- **Homogeneity of items** within each stratum provides greater precision in the estimates of underlying population parameters

+Cluster Sample

30

- Population is **divided into several 'clusters'**, each representative of the population; e.g. postcode areas, electorates etc.
- A **simple random sample of clusters** is selected
- All items in the selected clusters can be used, **or** items can be chosen from a cluster using another probability sampling technique



+Cluster Sample (cont)

31

- More **cost-effective** than random sampling, especially if population is geographically widespread
- Often requires a **larger sample size** compared to simple random sampling or stratified sampling for same level of precision

+7.5 Evaluating Survey Worthiness

What is the **purpose** of the survey?

Is the survey based on a **probability or non-probability sample**?

Survey errors

- **Coverage error** – appropriate or adequate frame?
- **Non-response error** – results in non-response bias
- **Sampling error** – always exists and is the difference between sample statistic and population parameter
- **Measurement error** – ambiguous wording, halo effect or respondent error

+Survey Errors

33

Coverage error



Excluded from frame

Non-response error



Follow-up non-responses

Sampling error



Random differences
from sample to sample

Measurement error



Bad or leading question

+ Summary

At the end of topic 1 you should:

Know the difference between descriptive statistics and inferential statistics

Appreciate some of the uses of statistics

Appreciate the process of a statistical exercise and the importance of data collection

+ Question

True or False:

A statistic is usually used to provide an estimate for a usually unobserved parameter.

A) True

B) False

+

Question

The evening host of a dinner dance reached into a bowl, mixed all the tickets around, and selected the ticket to award the grand door prize. What sampling method was used?

A) Simple random sample.

B) Systematic sample.

C) Stratified sample.

D) Cluster sample.

+ Question

True or False:

The amount of time a student spent studying for an exam will be measured on a ratio scale.

A) True

B) False

+TTD Week One

By the end of the week 1 make sure you ...

- Familiarised yourself with the MIS770 CloudDeakin site. In particular, the resources and discussion areas.
- Read the MIS770 UnitGuide and Your rights and responsibilities as a student in this Unit
- Read the suggested sections of chapters of the text
- Complete the first tutorial
- Summarise the key terms introduced this week
- Undertake the suggested exercises from the text

(pl. refer to the [MIS770 T2 2021 Content Overview and Unit Planner](#) for more details)