**Chapter 1**

**Introduction and data collection**

After studying this chapter you should be able to:

1. identify how statistics is used in business

2. recognise the sources of data used in business

3. identify the types of data used in business

1.1 (a) The type of beverage sold yields categorical or ‘qualitative’ responses.

(b) The type of beverage sold yields distinct categories in which no ordering is implied.

1.2 Three sizes of soft drink are classified into distinct categories — small, medium and large — in which order is implied.

1.3 (a) The time it takes to download a video from the Internet yields numerical or ‘quantitative’ responses.

(b) The download time is a ratio scaled variable because the true zero point in the measurement is zero units of time.

1.4 (a) Numerical, discrete, ratio scale

(b) Numerical, continuous, ratio scale

(c) Categorical, nominal scale

(d) Categorical, nominal scale

1.5 (a) Numerical, continuous, ratio scale

(b) Numerical, discrete, ratio scale

(c) Categorical, nominal scale

(d) Categorical, nominal scale

1.6 (a) Categorical, nominal scale

(b) Numerical, continuous, ratio scale

(c) Numerical, discrete, ratio scale

(d) Numerical, discrete, ratio scale

1.7 (a) Numerical, continuous, ratio scale\*

(b) Numerical, discrete, ratio scale

(c) Numerical, continuous, ratio scale\*

(d) Categorical, nominal

\*Some researchers consider money as a discrete numerical variable because it can be ‘counted’.

1.8 (a) Income may be considered discrete if we ‘count’ our money. It may be considered continuous if we ‘measure’ our money; we are only limited by the way a country’s monetary system treats its currency.

(b) The first format is preferred because the responses represent data measured on a higher scale.

(c) The first format would bring a greater response because it has a large number of possible income levels, as opposed to a limited number of income groups in the second format.

1.9 (a) The population is ‘all Australian working women’. A systematic or random sample could be taken of Australian working women. The director might wish to collect both numerical and categorical data.

(b) Three categorical questions might be occupation, marital status and type of clothing purchased online.

Numerical questions might be age, average monthly hours shopping online for clothing and income.

1.10 An experiment will be carried out in artificial conditions and will try to assess people’s attitudes to making charitable donations. These may differ in real-life situations. A survey of the same subjects will rely on self-reporting. Respondents may not always report accurately if they are trying to appear more generous.

1.11 The company is likely to have more data relating to its own sales. It can use invoices to retailers, if it is a wholesaler, or records of cash and credit card purchases, if it is a retailer. For competitors, it will have access to advertised prices from the media or online advertisements but may not know the sales volumes.

1.12 A population contains all the items of interest whereas a sample contains only a portion of the items in the population.

1.13 A statistic is a summary measure describing a sample whereas a parameter is a summary measure describing an entire population.

1.14 Descriptive statistical methods deal with the collection, presentation, summarisation and analysis of data whereas inferential statistical methods deal with decisions arising from the projection of sample information to the characteristics of a population.

1.15 Categorical random variables yield categorical responses, such as yes or no answers. Numerical random variables yield numerical responses, such as height in centimetres.

1.16 Discrete random variables produce numerical responses that arise from a counting process. Continuous random variablesproduce numerical responses that arise from a measuring process.

1.17 An operational definition is a universally accepted meaning that is clear to all associated with an analysis. Without an operational definition, confusion can occur.

1.18 The four types of measurement scales are (i) nominal scale, (ii) ordinal scale, (iii) interval scale and (iv) ratio scale.

Answers for 1.19 to 1.24 provided below are just some of the many different possible answers.

1.19 The answers to this question depend on which story is being selected.

1.20 The answers to

1.21 (a) This publication presents information about the Australian labour force, showing monthly data from 1978 onwards. It examines employment by gender and age group showing full time, part time, total employment and unemployment rates, which can be viewed by state. It is based on a survey sample of approximately 26,000 homes and non-private dwellings, such as hotels, with interviews carried out each month over an eight-month period. More recently, online electronic data collection methods have been used as well.

(b) Categorical variables: gender, employment status, industry, school or tertiary education attendance, state of residence etc.

(c) Numerical variables: number of hours worked, age

(d) Both hours worked and age are continuous numerical variables. Sometimes the way the data is collected results in integer values only, especially age in whole years.

1.22 (a) Categorical variables: gender, marital status, education, country of birth, religious affiliation, languages spoken other than English.

(b) Numerical variables: weekly income, age, weekly rent, monthly mortgage payment.

(c) Discrete numerical (weekly income, age in years, weekly rent, monthly mortgage payment).

1.23 (a) Numerical discrete, numerical continuous and categorical (nominal) data were used.

(b) Departing or arriving airline.

(c) Number of flights departed, arrived, cancelled.

1.24 (a) The populations of interest were all school principals and P&C organisations in NSW and the ACT, and the general public of NSW and the ACT when the survey was conducted.

(b) Sample 1 consisted of respondents to the survey of principals and P&C organisations. Sample 2 consisted of respondents to the survey from the general public.

(c) If a greater proportion of responses were from the school community, there could be more concern about child safety and more observations made by pedestrians approaching the school. If a greater proportion of responses was from the general public, other issues such as irritation about poor signage for 40 km an hour school zones could predominate.

1.25 (a) The population of interest was all households in Tasmania.

(b) Categorical variables (where and what flavour of milk is purchased) and numerical variables (number of people living in the household and number of millilitres drunk).

(c) Answers could vary.

(d) Answers could vary.

1.26 (a) The population of interest was the residents living in a north-eastern region of Sydney when the study was conducted.

(b) Categorical (ticket type and main purpose of using the bus) and numerical (resident’s age and frequency of bus use).

(c) Numerical continuous questions (distance travelling) and numerical discrete (weekly income of passengers).

(d) Categorical questions (gender and social status such as students, full-time or part-time workers and pensioners).