Student Name

Mathematics Stage 4

DATA ANALYSIS

Book 3 Analyse datasets presented in various ways and draw conclusions

Version: 260207

Report mistakes & suggest improvements:
https://MrDingMaths.com

CONTENTS

OVERVIEW

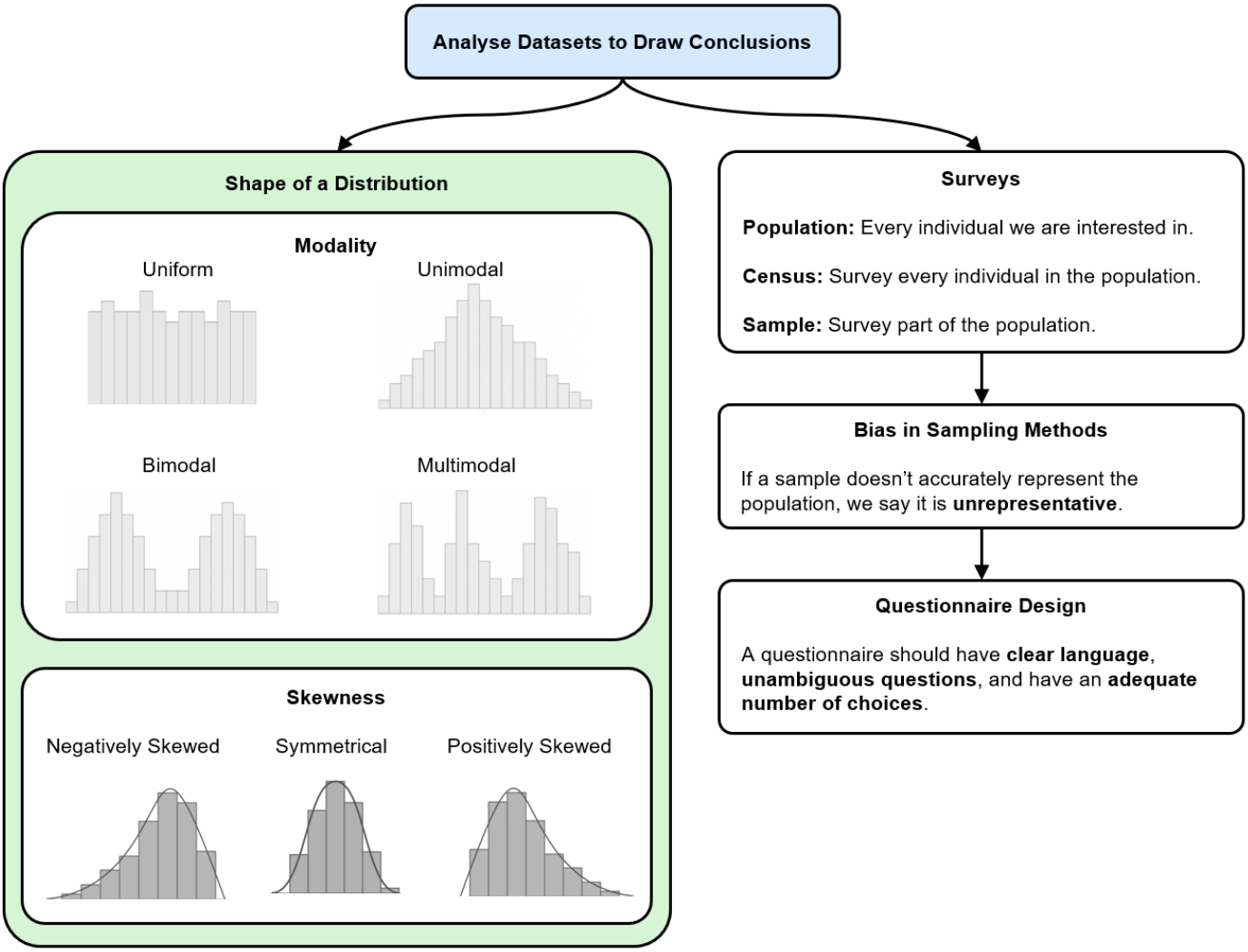
SYLLABUS CONTENT

MA4-DAT-C-02 analyses simple datasets using measures of centre, range and shape of the data

Analyse datasets presented in various ways and draw conclusions

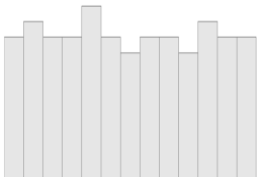
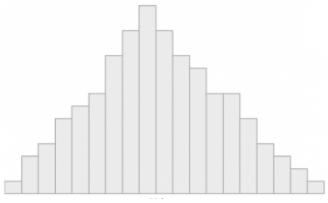
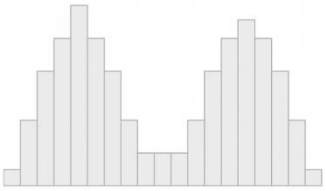
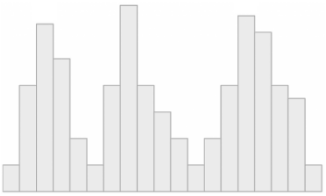
- Identify and describe the shape and distribution of a dataset using the terms symmetrical, negatively skewed and positively skewed
- Define a census as a study of every unit, everyone or everything in a population
- Define a sample as a subset of units in a population selected to represent all units in a population of interest
- Draw conclusions and make informed decisions about data gathered using data-collection techniques, including census and sampling, which is then presented in tables, graphs and charts

SUMMARY

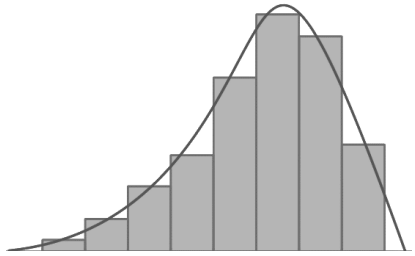
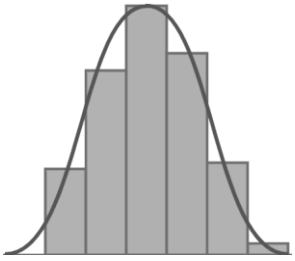
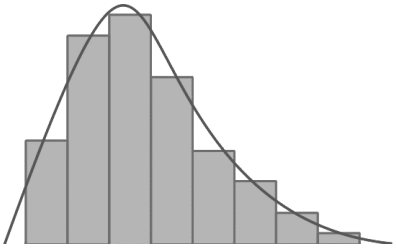


SHAPE OF A DISTRIBUTION

- In statistics, "**distribution**" refers to the way in which data values are spread out.
- A distribution reveals how often values occur, where they cluster, and the data's overall shape.
- To describe a **distribution's shape**, we discuss its **modality** and **skewness**.

Modality			
Uniform	**Unimodal**	**Bimodal**	**Multimodal**
No mode	One mode	Two modes	Many modes
			

- **Modality** describes if data has **clusters**.
- It counts how many **peaks** there are. The frequencies don't need to be the same.

Skewness		
Negatively Skewed	**Symmetrical**	**Positively Skewed**
Tail to the left	Symmetric data	Tail to the right
		

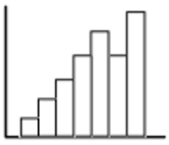
- **Skewness** describes how symmetrical a distribution is.
- We name skew based on the "**tail**" of the data.

Key ideas

1. ..... skewed data has a tail to the right.
2. ..... skewed data has a tail to the left.

1 Decide whether these graphs show **positively** or **negatively** skewed data.

a

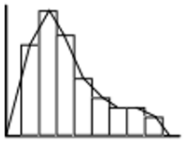


Negatively skewed

b

Negatively skewed

c

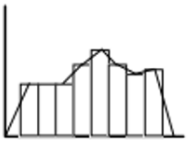


Positively skewed

d

Positively skewed

e



(very slightly) negatively skewed

f

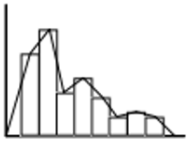
Stem	Leaf
5	0
6	5 6
7	3 8 8
8	7 7 7

Negatively skewed

g

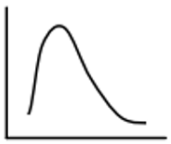
Negatively skewed

h



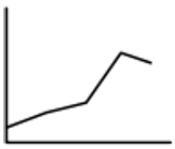
Positively skewed

i



Positively skewed

j



Negatively skewed

k

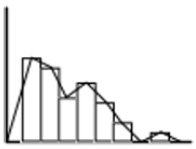
Stem	Leaf
10	1 5 6 6 8
11	2 3 3
12	4
13	1 5

Positively skewed

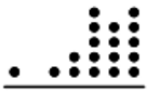
l

Negatively skewed

m

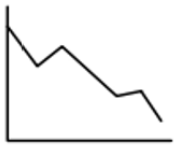


n



Negatively skewed

o



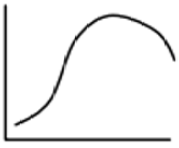
Positively skewed

p

Tens	Units
4	1
5	0 3
6	1 8 8

Negatively skewed

q



Negatively skewed

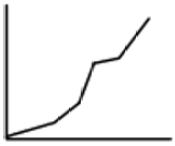
r

Positively skewed

s

Negatively skewed

t



Negatively skewed

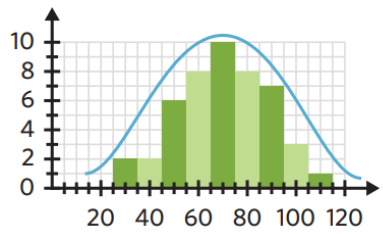
u

Tens	Units
9	2 3 3 7
10	0 4 9
11	3

Positively skewed

2 Circle the most appropriate term that describes the distribution of data.

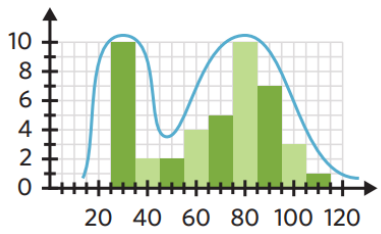
a



Symmetric
Bi-modal
Negatively skewed
Positively skewed

symmetric

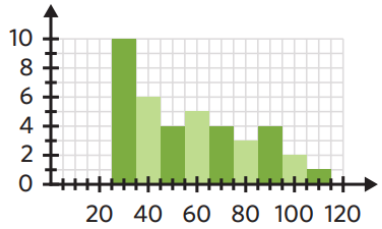
b



Symmetric
Bi-modal
Negatively skewed
Positively skewed

bimondal

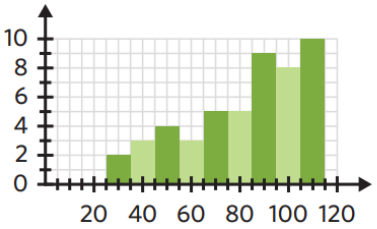
c



Symmetric
Bi-modal
Negatively skewed
Positively skewed

Positively skewed

d



Symmetric
Bi-modal
Negatively skewed
Positively skewed

Negatively skewed

3 Match each description with one of the curves shown below

a Unimodal, positively skewed

A

b Unimodal, negatively skewed

D

c Symmetrical, bell-shaped, unimodal

E

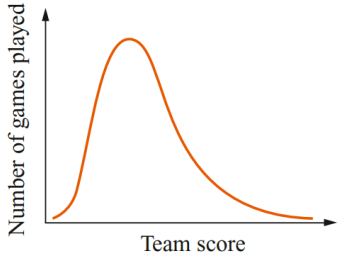
d Symmetrical, bimodal

B

e Aymmetrical, multimodal

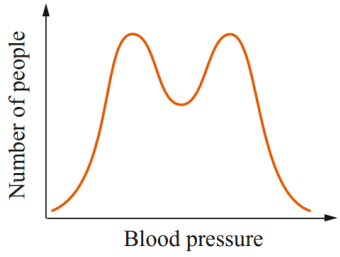
C

A



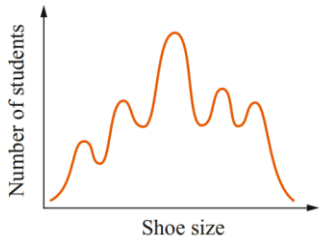
M

B



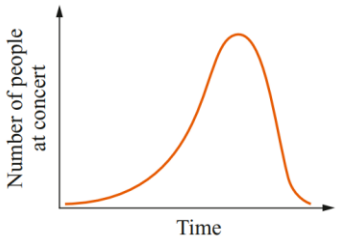
M M

C



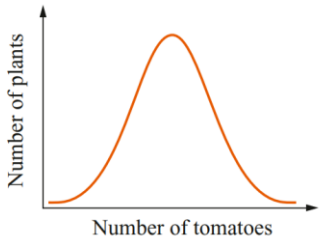
M

D



M

E

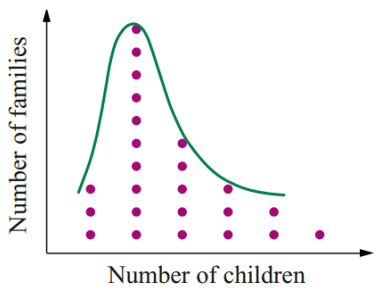


M

f For each of the graphs, place an M on the x axis to show the mode(s).

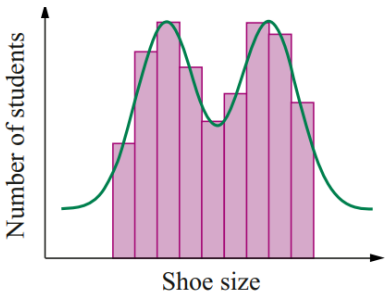
4 Describe the shape of each distribution as symmetric, positively skewed, negatively skewed, or bimodal.

a



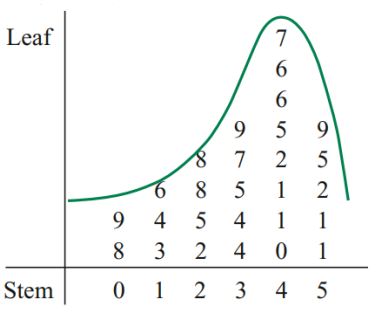
Positively skewed

b



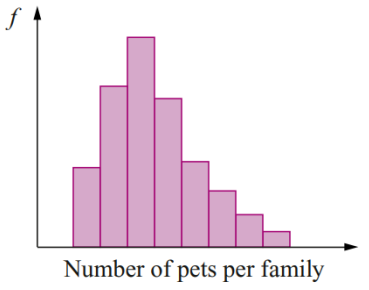
Bimodal

c



Negatively skewed

d



Positively skewed

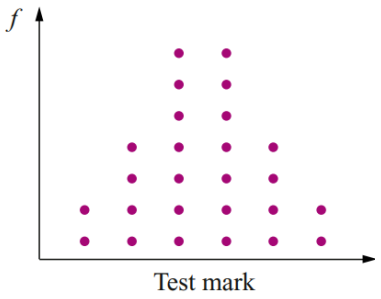
e

Marks on test

Stem	Leaf
1	5
2	3 4
3	5 5 7
4	2 3 3 9
5	1 4 4 6 8 9
6	0 1 3

Negatively skewed

f



symmetrical

5 Match each of the following descriptions to one of the curves, A to D below.

a the heights of 1000 randomly selected adult women

B

b the results in a difficult examination

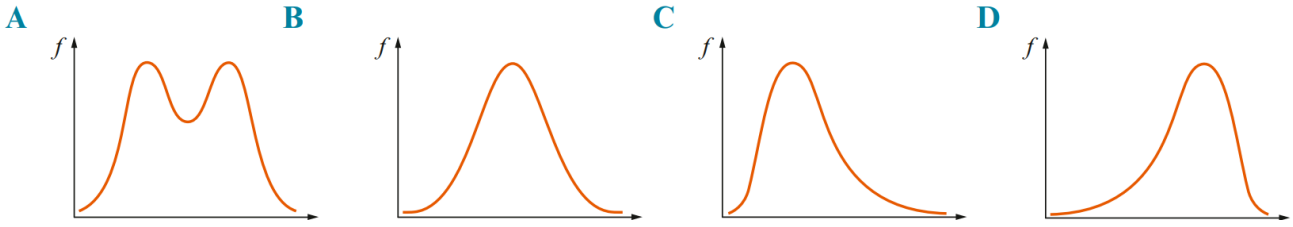
C

c the weights of jockeys and wrestlers

A

d the results in an easy examination

D



6 Describe the shape of each data set in the following back-to-back stem-and-leaf plot as symmetric, positively skewed, negatively skewed or bimodal.

Scores on Mathematics test

9 Blue	Stem	9 Red
9 9 8	4	7
8 7 2 1 1	5	3 4 7 8
4 3 2	6	1 5
5 4	7	0 5 5 6 8 9
3 0	8	3 4
1	9	2

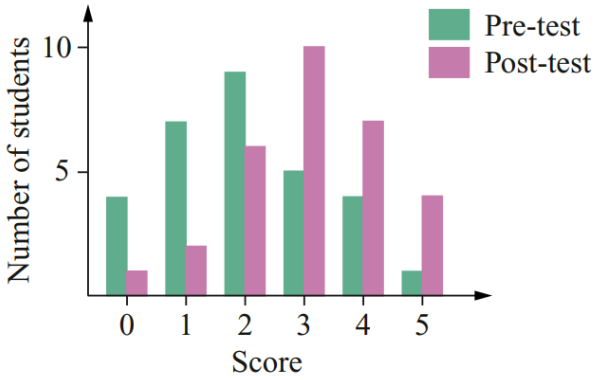
9B positively skewed

9R bimodal

7 The distribution shows the scores for a group of students on a mathematics topic pre-test and post-test.

a Describe the distribution for the pre-test.

Positively skewed



b Describe the distribution for the post-test.

Negatively skewed

c Comment on how the data has changed.

Students performed better post-test

8 **2022 HSC Standard 2 Band 5**

The ages of members of a dance class are shown in the back-to-back stem-and-leaf plot.

Women		*Men*
2	3	4 6
4 2	4	2 2 5 6 8
8 8 5 4 0 0	5	3
9 4 3 3	6	3

Pat claims that the women who attend the dance class are generally older than the men.

Is Pat correct? Justify your answer by referring to the clustering and skewness of each dataset.

Pat is correct. Women are clustered towards 50s and 60s and negatively skewed.

Men are clustered in 40s and positively skewed.

9 **2008 HSC Standard 2 Band 5**

The retirement ages of two million people are displayed in a table.

What is the relative frequency of the 51–55 year retirement age?

$\frac{35000}{2000000} = \frac{7}{400}$

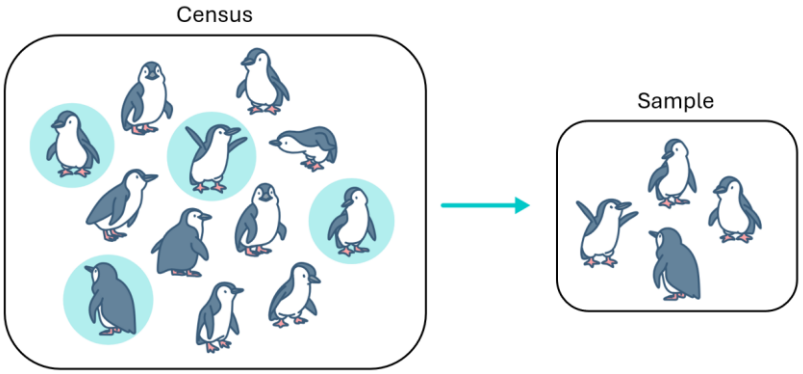
Retirement age	*Number of people* (thousands)
36–40	5
41–45	10
46–50	20
51–55	35
56–60	180
61–65	700
66–70	500
71–75	400
76–80	150

Describe the distribution.

Negatively skewed

SURVEYS

- In statistics, a **population** is every individual that we are interested in.
- A **census** collects data from the whole population.
- A **sample** collects data from part of the population.



Example Identify the population and census or survey

The Australian Government wants collect information about all adults in Australia. Every 5 years they survey every Australian resident over 18. The population is every adult in Australia. This is a census because it collects information from every adult.	Scientists are investigating migration patterns of whales in Australian waters. They track the movements of 200 whales out of 3500. The population is whales in Australian water. Since only 200 out of the total 3500 are being surveyed, it is a sample.

Determine the population of these surveys and whether it is a survey or a sample.

a Information needed to be collected about residents in Sydney. A survey collects information from all Sydney residents.

Population: .....

All residents in Sydney

Census or sample: .....

Census

b A survey was conducted to collect information about the voting intentions of people in a state election. 1,000 randomly selected voters were surveyed.

Population: .....

All voters in the state

Census or sample: .....

Sample

c Every single pen produced at a pen factory is inspected for quality.

Population: .....

All pens producted at a factory

Census or sample: .....

Census

d 1000 mL of water was collected to analyse the salt content in a river.

Population: .....

All water in a river

Census or sample: .....

Sample

1 Determine whether these surveys are a census or sample.

a Everyone in the house getting to vote on which board game to play next.

Census

b A scientist gathering 500g of soil to analyse land quality in the area.

Sample

c Every second shopper at a supermarket is asked if they like a product.

Sample

d A referendum where everyone in a country votes on a new law change.

Census

e A shop owner using an automatic counter above the entrance to keep track of how many customers come in every day.

Census

f A school conducts a survey to gather feedback on the quality of lunches. They distribute the survey to every student in the school.

Census

g A polling agency conducts a survey to predict election outcomes. They randomly select 5,000 voters across the country and gather their voting preferences.

Sample

h A company conducts a survey to gather feedback from all its employees about their satisfaction with the workplace environment.

Census

i Researchers conduct a survey to understand the prevalence of a specific health condition in a city. They randomly select households to participate in the survey.

Sample

j A retail store sends out a survey to everyone who made a purchase in the past month to gauge their satisfaction with the store's services.

Census

BIAS IN SAMPLING METHODS

- If a sample doesn't accurately represent the population, we say it is **unrepresentative**.
- Sources of sampling bias:

Sampling Bias	Description	Example
Selection bias	The selected group misrepresents the population.	Measuring average height of 15-year-olds by only surveying the basketball team.
Small Sample Size	A small sample may miss population variety.	Surveying only three people about a new highway.
Volunteer Bias	The sampling method relies on people volunteering to participate.	A study on dietary habits attracting only health-conscious volunteers.
Survivorship Bias	Considers only successful cases, ignoring failures.	Evaluating the success of businesses only by looking at those still in operation, ignoring those that have failed.

Example Describe sampling bias.

a	A survey was conducted to determine the average income of residents in a particular city. The survey only included responses from residents who lived in wealthy harbourside neighbourhoods. Selection bias	**b**	Mr Ding wants to know the percentage of school students who have maths as their favourite subject. He surveys the Year 12 Maths Extension class. Selection bias
c	The government wanted to know farmers' opinions on a new road. They asked two farmers out of a community of 500. Small sample size	**d**	Investigating the religious beliefs of Australians by surveying people at shopping centres on Sunday mornings. Selection bias

1 Discuss why the following sampling methods may result in an unrepresentative sample:

a Responses to a survey conducted by email.

only those with internet access respond; excludes people without email.

b A survey on household spending is needed.
You take a random sample of a Sydney harbour-side suburb.

Harbour-side suburbs have higher incomes than typical Australian households.

c A survey of the number of people who can touch their toes is done with students from a gymnastics class.

Gymnasts are more flexible than the general population.

d Researching favourite music genres by interviewing people outside a classical music concert.

People attending classical concerts prefer classical music.

e Finding opinions on new cycle lanes by interviewing members of the local cycling club.

Cycling club members support cycle lanes; not representative of all residents.

f Looking at the state of the economy by interviewing wealthy landowners.

Wealthy landowners don't represent the economic situation of all citizens.

g Researching favourite films of teenagers by conducting interviews at an anime convention.

Anime convention attendees prefer anime films; not representative of all teenagers.

h A peak hour survey of the people on a train about whether they are employed.

Peak hour train passengers are more likely to be employed than the general population.

i A survey at the train station of how Australians get to work.

Excludes people who drive, work from home, or don't use trains.

j A survey of the people in your year level on Australia's most popular television shows

Your year level's preferences don't represent all Australians' viewing habits.

k Year 7 students are interviewed about whether they like changes to the school uniform.

Year 7 students may not know how it was before to make a comparison.

l Motorists stopped in peak hour are interviewed about traffic problems

Peak hour motorists experience more traffic than typical drivers.

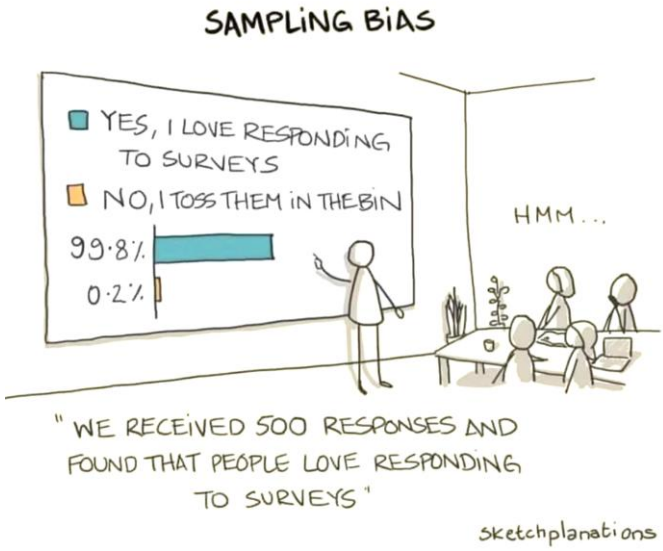
m A YouTuber asks their subscribers for feedback on their videos.

Subscribers already like the content; excludes critics and non-subscribers.

n Triple J (pop rock station) radio listeners were asked to vote for the best song of the year.

Triple J listeners prefer pop/rock; excludes fans of other genres.

2 Consider the cartoon on sampling bias.
Explain why the conclusion drawn from this dataset is not valid.

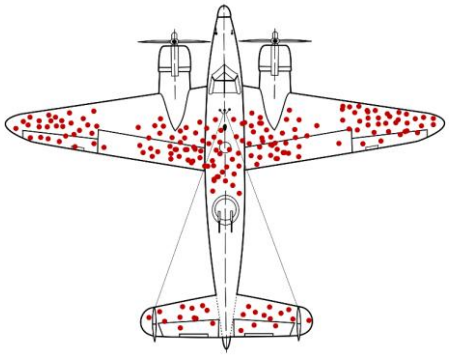


The sample only includes people who responded to the survey. Those who don't like surveys (and would toss them in the bin) are excluded from the sample because they don't respond. The sample is self-selecting and unrepresentative of the population.

3 In WW2, the US military found that returning planes were hit in these locations.

A statistician suggested that extra armour be placed at the places where the plane was *not* hit.

Explain why, with reference to survivorship bias.



The data only shows planes that survived and returned. Planes hit in unaffected areas (engines, cockpit) likely crashed and didn't return. This is survivorship bias: the sample excludes failures.
Areas without damage on returning planes are actually the critical areas where hits are fatal, so armour should go there.

QUESTIONNAIRE DESIGN

- A questionnaire should have clear language, unambiguous questions, and have an adequate number of choices.

Questioning Bias	Example
Leading question	Implying a preferred answer: "most people think that…" Using emotional language: "do you agree that this policy is brilliant?" Giving one-sided contextual information: "given the new metro will cost $10 billion, should we build it?"
Ambiguous questions	A question that asks how often someone goes to the gym but does not specify the time frame. Questions that are too open ended generally lead to bias because participants can interpret it in different ways.
Not providing enough choices	A question that asks how many hours of homework per week, with multiple choice response that ranges from 1 to 4 hours.

Example Identify bias in questions.

a A questionnaire asks: Polls show that 90% of customers prefer the new Strawberry Dream yogurt.
What do you think?

Leading question

b A questionnaire asks: Considering that many students dislike mathematics, should mathematics be made compulsory?

Leading question

c A questionnaire asks: Considering that almost all future jobs will require some use of mathematics, should mathematics be made compulsory?

Leading question

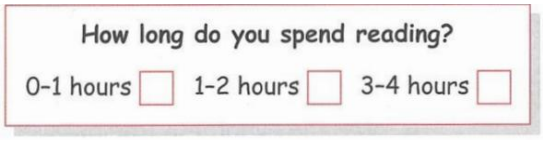
d A survey was conducted to estimate how much homework students do.
From "a little" to "a lot."

Ambiguous question

e A survey asks "what do you think about the government's evil plan to build new apartments in Croydon?"

Leading question

f



Ambiguous and not enough choices

1 Explain why the following survey questions may result in a biased statistical investigation.

a Taking into account the budget deficit, should the government cut military expenditure?

Leading question - mentions budget deficit first, priming respondents to consider financial constraints before deciding on military cuts

b Taking into account the threat of Russia, should the government cut military expenditure?

Leading question - mentions Russian threat first, priming respondents to prioritise security before considering cuts

c Taking into account that the Sydney Metro is well over-budget, do you agree that we should continue to fund it?

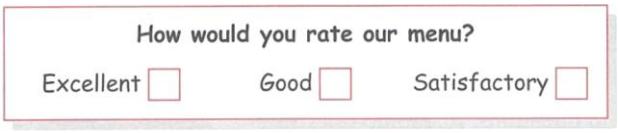
Leading question - mentions overbudget first to make residents consider the budget vs its benefits. “do you agree” encourages yes responses.

d The aging infrastructure of Sydney Trains has led to many major delays over the years. What’s your opinion on the Metro project?

Leading question presents negative information (aging infrastructure, delays) before asking opinion, biasing respondents against current system and towards Metro

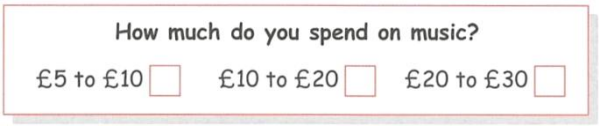
2 Describe why these questions may lead to a biased statistical investigation.

a



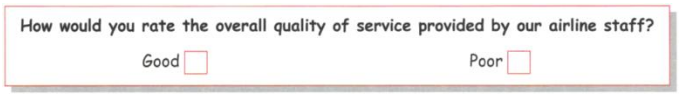
No neutral option provided; forces respondents into positive categories even if they had negative experience

b



Overlapping boundaries - unclear where exactly £10 and £20 belong
No timeframe.

c

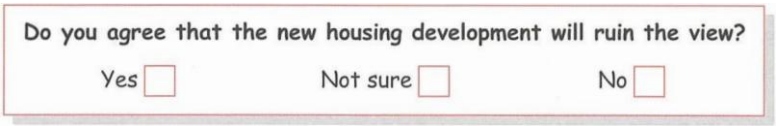


No neutral option and not enough of a range of choices.

d Write a criticism for the question and the response.

Ambiguous question has to timeframe, how often in a week?
Ambiguous answers are subjective, how many times is sometimes?

3 Flo wants to know how people feel about a new housing development. She asks this question.



a What is wrong with Flo's question?

Leading question, “do you agree that” which encourages yes responses

b Flo asks ten of her friends. Why might Flo's sample be biased?

Unrepresentative sample. Flor and her friends may think similarly.

4 Darius is conducting a survey to find out if students use the internet to help them with their homework. Here are two questions that Darius plans to ask his friends.

> Do you have the internet at home and use it to help you with your homework?

> Don't you agree that the internet is great for helping with homework?

a Write down one criticism of each question

The first question is two questions. People may only read the first part.

The second is a leading question.

b Suggest more suitable wording for the second question. Design a response section for your question.

How often do you use the internet to help you with your homework in Maths?

None of the time, less than half the time, half the time, more than half the time, all the time

c Darius also wants to find out how long his friends spend online. Write down a suitable question that Darius could ask.

How many hours do you spend online in a week?

5 There are two supermarkets in a town: Pricewise and Freshfoods. The manager at Pricewise is carrying out a survey. A member of staff stands at the entrance to their store and stops shoppers at random. The shoppers are asked the following question:

> Do you agree that our store provides better value for money than *Freshfoods*?
>
> Strongly agree ☐ Agree ☐ Don't know ☐

Write down one criticism of

a the question.

Leading question

b the response section.

Not enough choices

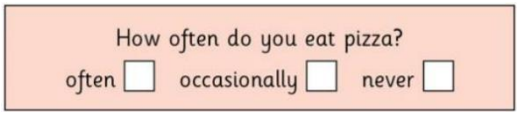
c the method of sampling.

Unrepresentative sample as people shopping at Pricewise likely already do so because they think it's good value.

CHECK YOUR UNDERSTANDING

Bias in questions

1 What's wrong with this question?



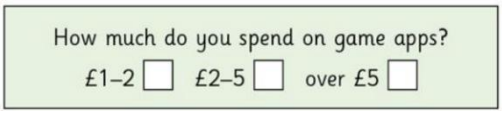
a The options are subjective.

b There is no option for someone who never eats pizza

c The options are objective

d There should only be two options.

2 What's wrong with this question?



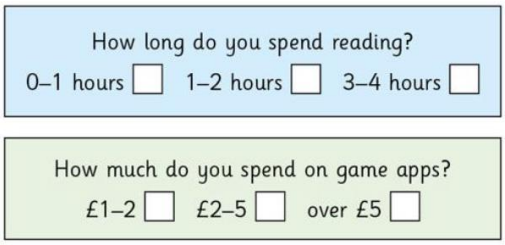
a The options are subjective.

b There is no option for someone who does not buy game apps

c Someone who spends £5 has two groups

d Someone who spends £1.50 has no group

3 What's wrong with **both** of these questions?



a Both have no option for 0.

b Both have no maximum option.

c The options are subjective

d No time frame is given.