



REVIEW

Chapter

5

# Descriptive statistics

**Contents:**

- A Describing data  
*Investigation 1: Statistics from the internet*
- B Presenting and interpreting data
- C Grouped discrete data  
*Investigation 2: Taxi Sir?*
- D Continuous data  
*Investigation 3: Choosing class intervals*
- E Frequency distribution tables
- F Summarising the data  
*Investigation 4: Effects of outliers*
- G Measuring the spread of data
- H Box-and-whisker plots
- I The standard deviation  
*Investigation 5: Heart stopper*
- J Statistics using technology
- K Parallel boxplots  
*Investigation 6: How do you like your eggs?*

Review set 5A

Review set 5B

Prepare a report of your work. You may choose how you present your work. You may present it as:

- a newspaper or magazine article
- a video
- a powerpoint slide presentation
- a wordprocessed document.

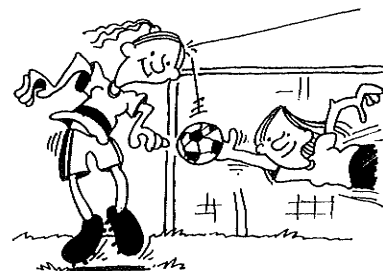
In your report include:

- A description of the problem or issue that you are investigating.
- A simple account of the method you have employed to carry out the investigation.
- The analysis you carried out. This includes a copy of your data, any graphics and summary statistics you produced and the argument that you wrote to support your conclusion.
- Your conclusion.
- A discussion of any weaknesses in your method that may cause your conclusion to be suspect.

### REVIEW SET 5A

1 Classify the following data as categorical, discrete numerical or continuous numerical:

- a the number of pages in a daily newspaper
- b the maximum daily temperature in the city
- c the manufacturer of a car
- d the preferred football code
- e the position taken by a player on a soccer field
- f the time it takes 15-year-olds to run one kilometre
- g the length of feet
- h the number of goals shot by a soccer player
- i the amount spent weekly, by an individual, at the supermarket.



2 The following marks were scored for a test where the maximum score was 50:

47 32 32 29 36 39 40 46 43 39 44 18 38 45 35 46 7 44 27 48

- a Construct an ordered stemplot for the data.
- b What percentage of the students scored 40 or more marks?
- c What percentage of the students scored less than 30 marks?
- d If a score of 25 or more is a pass, what percentage of the students passed?
- e Describe the distribution of the data.

3 A frequency table for the heights of a basketball squad is given below.

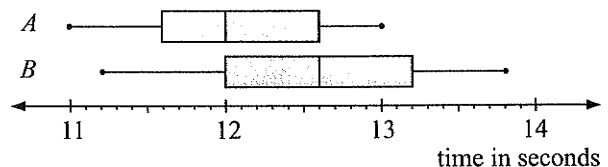
Height (cm)	Frequency
170 -	1
175 -	8
180 -	9
185 -	11
190 -	9
195 -	3
200 - < 205	3

- a Explain why 'height' is a continuous variable.
- b Construct a histogram for the data. The axes should be carefully marked and labelled and include a heading for the graph.
- c What is the modal class? Explain what this means.
- d Describe the distribution of the data.

- 4 Six scores have mean 8. What must the seventh score be to increase the mean by 1?
- 5 The data below shows the distance, in metres, Kapil was able to throw a cricket ball.
- |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|
| 71.2 | 65.1 | 68.0 | 71.1 | 74.6 | 68.8 | 83.2 | 85.0 | 74.5 | 87.4 |
| 84.3 | 77.0 | 82.8 | 84.4 | 80.6 | 75.9 | 89.7 | 83.2 | 97.5 | 82.9 |
| 90.5 | 85.5 | 90.7 | 92.9 | 95.6 | 85.5 | 64.6 | 73.9 | 80.0 | 86.5 |

- a Determine the highest and lowest value for the data set.
- b Produce between 6 and 12 groups in which to place all the data values.
- c Prepare a frequency distribution table.
- d For this data, draw:
- i a frequency histogram
  - ii a relative frequency histogram
  - iii a cumulative frequency graph.
- e Determine: i the mean ii the median.

- 6 The given parallel boxplots represent the 100-metre sprint times for the members of two athletics squads.



- a Determine the 5-number summaries for both *A* and *B*.
- b Determine the i range ii interquartile range for each group.
- c Copy and complete:
- i The members of squad ..... generally ran faster times.
  - ii The times in squad ..... were more varied.
- 7 Find, using your calculator, the mean and standard deviation of these sets of data:
- a 117, 129, 105, 124, 123, 128, 131, 124, 123, 125, 108
  - b 6.1, 5.6, 7.2, 8.3, 6.6, 8.4, 7.7, 6.2

- 8 Find *a*, given that 3, 0, *a*, *a*, 4, *a*, 6, *a* and 3 have a mean of 4.

- 9 The back to back stemplot alongside represents the times for the 100 metre freestyle recorded by members of a swimming squad.

<i>Girls</i>		<i>Boys</i>
	32	1
4	33	0227
763	34	13448
87430	35	024799
8833	36	788
7666	37	0
6	38	
0	39	
	40	
	41	

leaf unit: 0.1 sec

- a Copy and complete the following table:

	<i>Girls</i>	<i>Boys</i>
outliers		
shape		
centre (median)		
spread (range)		

- b Write an argument that supports the conclusion you have drawn about the girls' and boys' swimming times.

**REVIEW SET 5B**

- 1 A sample of lamp-posts were surveyed for the following data. Classify the data as categorical, discrete numerical or continuous numerical:
- the diameter of the lamp-post (in centimetres) measured 1 metre from its base
  - the material from which the lamp-post is made
  - the location of the lamp-post (inner, outer, North, South, East or West)
  - the height of the lamp-post, in metres
  - the time (in months) since the last inspection
  - the number of inspections since installation
  - the condition of the lamp-post (very good, good, fair, unsatisfactory).

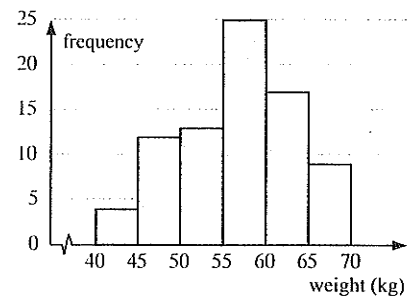
- 2 The data supplied below is the diameter (in cm) of a number of bacteria colonies as measured by a microbiologist 12 hours after seeding.

0.4	2.1	3.4	3.9	4.7	3.7	0.8	3.6	4.1	4.9
2.5	3.1	1.5	2.6	1.3	3.5	0.9	1.5	4.2	3.5
2.1	3.0	1.7	3.6	2.8	3.7	2.8	3.2	4.0	3.3

- Produce a stemplot for this data.
- Find the: i median ii range of the data.
- Comment on the skewness of the data.

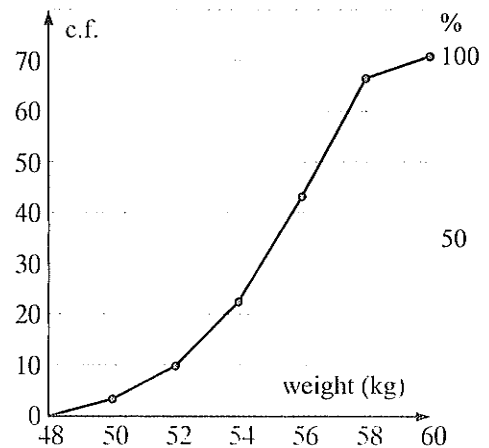
- 3 Consider the graph alongside.

- Name the type of graph.
- Is the data represented discrete or continuous?
- Find the modal class.
- Construct a table for the data showing class intervals, frequencies, midpoints and products.
- Find the approximate mean of the data.



- 4 The cumulative frequency graph shows the cumulative frequency of the weights of a herd of 12-month old female alpacas.

- How many alpacas were the the herd?
- What percentage of alpacas have weights under 52 kg?
- What percentage of alpacas have weights over 55 kg?
- Read the cumulative frequencies from the graph and set up a cumulative frequency table.
- Hence set up a frequency table for classes 48-, 50-, 52-, 54-
- Draw the histogram. Find the modal class.
- Is the histogram symmetrical about the modal class?



- 5 12 of 29 measurements are below 20 cm and 13 measurements are above 21 cm. Find the median if the other 4 measurements are 20.1 cm, 20.4 cm, 20.7 cm and 20.9 cm.
- 6 Eight sample values are: 6,  $a$ , 7,  $a$ , 4,  $b$ , 6 and 8 where  $a$  and  $b$  are single digit numbers and the mean is 7.
- Show that  $a$  and  $b$  have two possible solutions.
  - If there is a single mode, what is the median?
- 7 Determine the five number summary and the interquartile range for each of the following data sets that have already been placed in rank order. Then draw a boxplot for each data set:
- 4.0, 10.1, 13.4, 14.2, 15.0, 16.5, 22.2, 22.4, 23.1, 30.0
  - 11, 15, 17, 21, 23, 25, 25, 27, 47, 49, 49
- 8 Anneka's golf scores for her last 20 rounds were:
- 90, 106, 84, 103, 112, 100, 105, 81, 104, 98,  
107, 95, 104, 108, 99, 101, 106, 102, 98, 101
- Find the: i median ii lower quartile iii upper quartile
  - Find the interquartile range of the data set.
  - Find the mean and standard deviation of her scores.
- 9 Find, using technology, the i mean ii standard deviation iii variance of the following data sets:
- 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1
  - 12, 9, 3, 0, 1, 3, 6
- 10 The number of peanuts in a jar varies slightly from jar to jar. A sample of 30 jars for two brands X and Y was taken and the number of peanuts in each jar was recorded.

<i>Brand X</i>						<i>Brand Y</i>					
871	885	878	882	889	885	909	906	913	891	898	901
916	913	886	905	907	898	894	894	928	893	924	892
874	904	901	894	897	899	927	907	901	900	907	913
908	901	898	894	895	895	921	904	903	896	901	895
910	904	896	893	903	888	917	903	910	903	909	904

- Produce a back-to-back stemplot for the data for each brand.
- Complete the table below:

	<i>Brand X</i>	<i>Brand Y</i>
outliers		
shape		
centre (median)		
spread (range)		

- Use the above information to compare Brand X and Brand Y.