

PART D

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

# D CONTINUOUS DATA

**Continuous data** is numerical data which has values within a continuous range.

For example, if we consider the weights of students in a rugby training squad we might find that all weights lie between 50 kg and 100 kg.

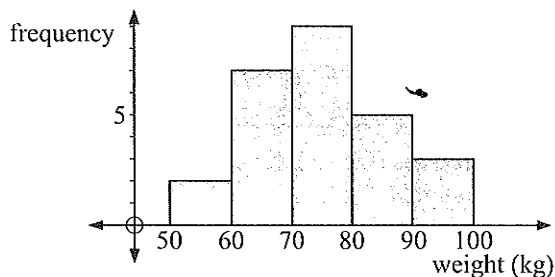
Suppose

- 2 students lie in the 50 kg up to but not including 60 kg,
- 7 students lie in the 60 kg up to but not including 70 kg,
- 9 students lie in the 70 kg up to but not including 80 kg,
- 5 students lie in the 80 kg up to but not including 90 kg,
- 3 students lie in the 90 kg up to but not including 100 kg.

The frequency table would be: We could use a histogram to represent the data graphically.

Weight interval	Frequency
50 -	2
60 -	7
70 -	9
80 -	5
90 -	3

**Weights of the students in the rugby squad**



## HISTOGRAMS

A **histogram** is a vertical column graph used to represent continuous grouped data.

There are no gaps between the columns in a histogram as the data is continuous.

The bar widths must be equal and each bar height must reflect the frequency.

### Example 2

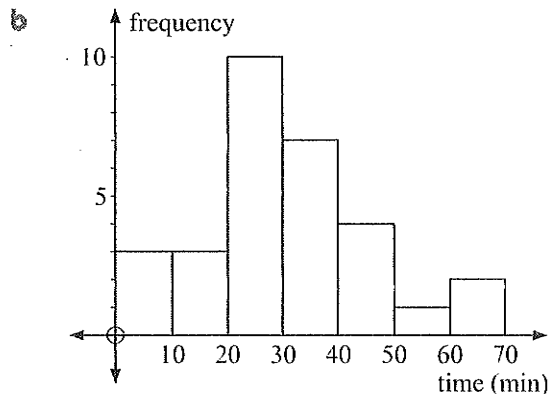
The time, in minutes (ignoring any seconds) for students to get to school on a given day is as follows:

5 12 32 26 41 37 22 27 41 52 49 38 41 62 69  
 37 21 4 7 12 32 36 39 14 24 27 29 22 21 25

- a Organise this data on a frequency table. Use time intervals of 0-, 10-, 20-, etc.
- b Draw a histogram to represent the data.

a

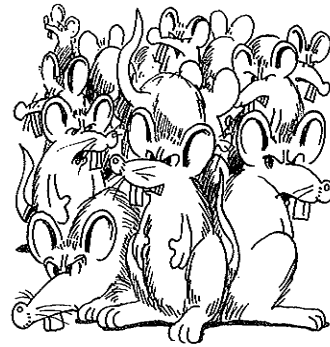
Time int.	Tally	Freq.
0 -		3
10 -		3
20 -		10
30 -		7
40 -		4
50 -		1
60 -		2



## COLLATING CONTINUOUS DATA

The following data are results (in grams) of weighing 50 laboratory rats:

261	133	173	295	265	142	140	271	185
251	166	100	292	107	201	234	239	159
153	263	195	151	156	117	144	189	234
171	233	182	165	122	281	149	152	289
168	260	256	156	239	203	101	268	241
217	254	240	214	221				



Continuous data can also be grouped.

Each group is called a **class**.

The size of the grouping is called the **class interval**. Because continuous data can be grouped it is sometimes referred to as **interval data**.

For example, from the data given above, a table may be designed so that the classes are 100 up to 120, 120 up to 140, .....

These are written as 100 -, 120 -, etc. where 100 - means all of the numbers from 100 up to but not including 120.

<i>Class interval</i>	<i>Tally</i>	<i>Frequency</i>
100 -		4
120 -		2
140 -	#	10
160 -	#	5
180 -		4
200 -		4
220 -	#	6
240 -	#	5
260 -	#	6
280 -		4
<i>Total</i>		50



### INVESTIGATION 3

### CHOOSING CLASS INTERVALS



When dividing data values into intervals, the choice of how many intervals to use, and hence the width of each class, is important.



#### What to do:

- 1 Click on the icon to experiment with various data sets. You can change the number of classes. How does the number of classes alter the way in which we can read the data?
- 2 Write a brief account of your findings.

As a rule of thumb we generally use approximately  $\sqrt{n}$  classes for a data set of  $n$  individuals. For very large sets of data we use more classes rather than less.

**EXERCISE 5D**

1 The initial weights of students (in kg) in a fitness class were:

69 90 61 58 94 68 77 80 64 87 59 81 73 56 72 66  
 62 75 78 75 87 73 81 70 73 58 85 73 76 92 82 86  
 82 71 79 80 85 70 60 75

- Using classes 50 - , 60 - , 70 - , 80 - , 90 - , tabulate the data using columns of score, tally, frequency.
- How many students are in the 70 - class?
- How many students weighed less than 80 kg?
- Find the percentage of students who weighed 70 kg or more.

2 A group of young athletes were invited to participate in a javelin throwing competition. The following results were obtained:

<i>Distance (metres)</i>	0 -	10 -	20 -	30 -	40 -
<i>No. of athletes</i>	3	26	41	14	6

- How many athletes threw less than 20 metres?
- What percentage of the athletes were able to throw at least 30 metres?

3

<i>Height (mm)</i>	<i>Frequency</i>
300 -	12
325 -	18
350 -	42
375 -	28
400 -	14
425 -	6

A plant inspector takes a random sample of two week old seedlings from a nursery and measures their height to the nearest mm.

The results are shown in the table alongside.

- How many of the seedlings are 400 mm or more?
  - What percentage of the seedlings are between 349 and 400 mm?
- c The total number of seedlings in the nursery is 1462. Estimate the number of seedlings which measure:
- less than 400 mm
  - between 374 and 425 mm.

## **E** FREQUENCY DISTRIBUTION TABLES

Frequency tables may include other columns which show **relative frequency**, **cumulative frequency**, and **cumulative relative frequencies**.

The **relative frequency** of an event is the frequency of that event expressed as a fraction (or decimal equivalent) of the total frequency.

The **cumulative frequency** of an event is the accumulation (sum) of the frequencies up to and including that event.

The **cumulative relative frequency** of an event is the accumulation (sum) of the relative frequencies up to and including that event.