Frequency Distributions

The frequency is the number of times each score, event or measurement occurs. A **frequency distribution** is shown in a table in which information or data is arranged in order. From the frequency distribution, the average and range can be found, and the data can be shown on graphs such as *histograms, frequency curves and cumulative frequency graphs.*

Definitions	Averages	Histogram	Frequency Curve
Stem and Leaf Diagram	Cumulative Frequency Curve	Measures of Spread	Box and Whisker Graph

Definitions

- Discrete data can be counted.

    e.g. Number of people in a theatre, number of leaves on a tree.

- Continuous data is obtained by measuring.

    e.g. Heights of people, lengths of leaves on a tree.

- Ungrouped data is used with small amounts of data where individual scores are listed.

    e.g.Test results of five people: 15, 20, 18, 17, 16

- Grouped data is used with larger amounts of data where scores are placed in groups or classes.

    e.g.Test results of 30 people

Score	Frequency
0 - 5	3
6 - 10	8
11 - 15	12
16 - 20	7

Averages

An **average** is a number that typifies a set of numbers or statistics.

There are three commonly used types of average.

- **Median.** The median of a set of scores is the middle score when the scores have been arranged in order from smallest to largest.
- **Mode.** The mode is the most frequent score, that is, the one that occurs most often.
- **Mean.**

The **mean** of a set of scores is the sum of all of the scores divided by the number of scores.

The **mean** is often known as the **average**.

The symbol for the mean is $\bar{x}$

For ungrouped data, $\bar{x} = \dfrac{\text{sum of scores}}{\text{number of scores}}$

For grouped data, $\bar{x} = \dfrac{\text{sum of (middle score x frequency)}}{\text{number of scores}}$

Example	Answer
For the following set of test results of 10 people: 3, 8, 5, 6, 8, 5, 4, 1, 10, 5 Find: (a) The mean (b) The mode (c) The median	(a) The mean $= \dfrac{3 + 8 + 5 + 6 + 8 + 5 + 4 + 1 + 10 + 5}{10}$ $= \dfrac{55}{10} = 5.5$ (b) The mode = 5 (occurs 3 times) (c) Arrange in order: 1, 3, 4, 5, 5, 5, 6, 8, 8, 10 Median is 5

Findng the mean on a calculator

The procedure for finding statistical values will vary slightly from calculator to calculator.

For a scientific calculator such as the Texas Instruments TI-30 IIB, to find the mean of 5, 7, and 12:

Task	Keys	Select
Select the statistical mode STAT	2nd STAT =	Choose 1-VAR
Enter the data	DATA 5 ▼ ▼ for each item of data	$x_1 = \_,\dots$ and FRQ = 1 for each item of data.
Check number of items entered	STATVAR n = 3	Select n
	STATVAR ▼	▼

| Clear old data | EXIT STAT | Always do this before entering a new set of data. |

Finding the mean on a spreadsheet

Enter the data:

	A	B
1	Values	
2	5	
3	7	
4	12	
5	Average	8

The function entered in cell B5 to find the mean is **=AVERAGE(A2..A4)** this gives a mean of 8.

Histogram

A histogram is a graph that is used to show the information from a frequency distribution.

It is similar to a column graph and the area of each column is proportional to the frequency of the score that it represents.
In a histogram the bars always touch, whereas in a bar graph the bars do not need to touch.

e.g Ages of workers in a company.

Note that this is grouped data.

Ages	Frequency
15 < 25	14
25 < 35	20
35 < 45	24
45 < 55	18
55 < 65	10
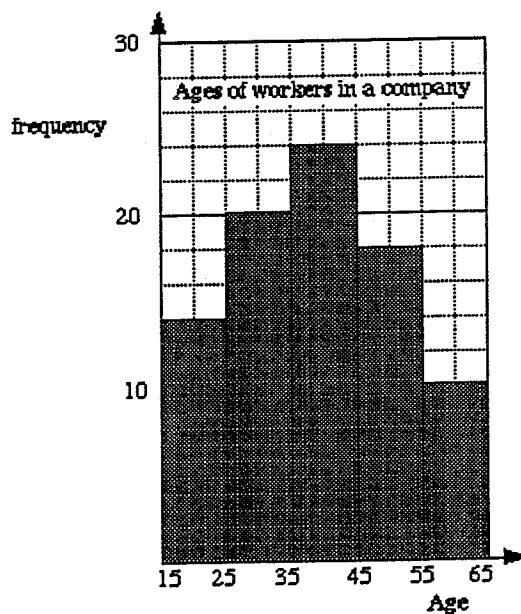
Note: 15 < 25   means $15 \le height < 25$

i.e Greater or equal to 15 but less
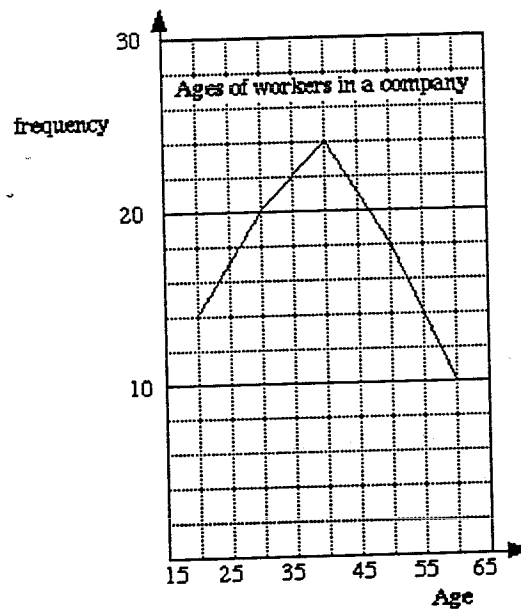
than 25.

Ages of workers in a company

Different people use different ways of presenting this type of data.

Frequency Curve

This is sometimes called a frequency polygon or line graph.

A frequency curve is obtained by joining up the mid-points of the tops of the columns of the histogram.

e.g. For the ages of the workers.

Stem and Leaf Diagram

A stem and leaf diagram or chart provides a means of ordering and displaying data.

Numbers are split into categories of the first digit(s) and the last digits(s)

e.g. Given the following test scores:

06, 24, 43, 23, 12, 34, 21, 08, 15, 47,
40, 20, 12, 05, 19, 28, 34, 23, 39, 34

Let the stem be the first digit (the tens column). Now enter each of the second digits (the units column) into the leaf part of the diagram.

```
0 | 6 8 5
1 | 2 5 2 9
2 | 4 3 1 0 8 3
3 | 4 4 9 4
4 | 3 7 0
```
Now arrange the leaf digits in order.
```
0 | 5 6 8
1 | 2 2 5 9
2 | 0 1 3 3 4 8
```

```
2 | 0 1 3 3 4 0
3 | 4 4 4 9
4 | 0 3 7
```

Both parts of a stem and leaf diagram can contain more than one digit.

e.g. For numbers such as 3476, the *stem* could be 34 and the *leaf* could be 76.

Cumulative Frequency Curve

In a cumulative frequency table, each of the scores from the frequency distribution is added to the total of the ones before it.

e.g. For the ages of the workers:

Age (x)	Cumulative frequency
< 15	0
< 25	14
< 35	34
< 45	58
< 55	76
< 65	86

e.g. There are a total of 58 workers aged less than 45 years old.

Ages of workers in a company
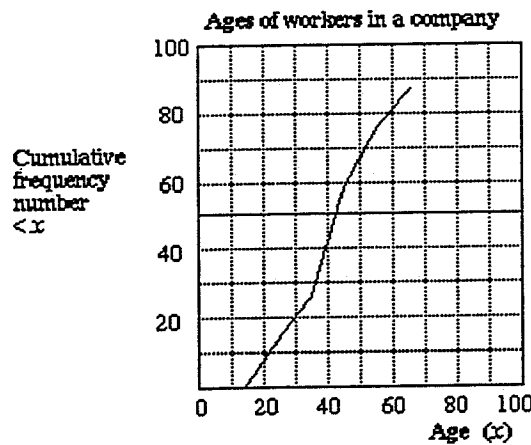
Measures of Spread

As well as averages, it is necessary to have a measure of how spread out the scores are.

The range and quartiles are ways of measuring this spread.
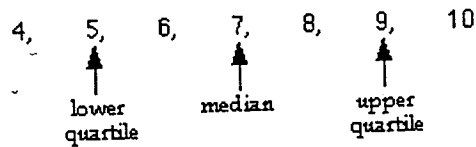
Range

The range of a set of scores is the difference between the highest and the lowest scores.

e.g. For {4, 5, 6, 7, 8, 9, 10} the range is 10 - 4 = 6

Quartiles

The **quartiles** divide a distribution into four equal parts.

- For ungrouped data {4, 5, 6, 7, 8, 9, 10}:

- The **interquartile range** is the difference between the upper and the lower quartiles

    i.e. Interquartile range = 9 - 5 = 4

- For grouped data the quartiles and median can be obtained by using the vertical axis of the cumulative frequency curve. The total number of scores is divided by four and then read off using the scale.

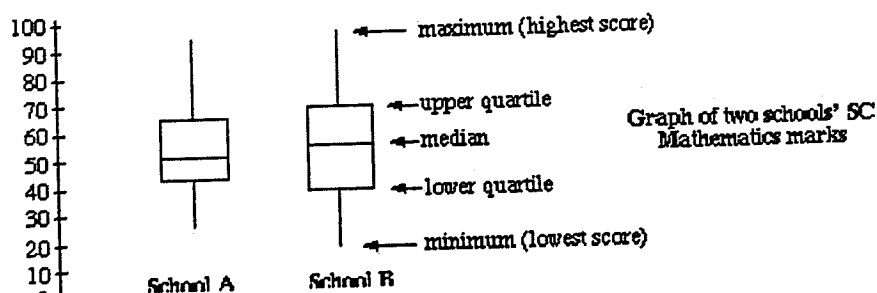---

Box and Whisker Graphs

Box and whisker graphs are useful for comparing the data in different frequency distributions.

The diagrams consist of a rectangle which indicates the quartiles and the median, and a line on each end to show the maximum and minimum values of the distribution.

e.g. Draw a box and whisker diagram to compare the following frequency distributions of two schools' School Certificate Mathematics results.

School A: Median, 51; Lower quartile, 42; Upper quartile, 65; Maximum, 95; Minimum, 23

School B: Median, 54; Lower quartile, 39; Upper quartile, 68; Maximum, 99; Minimum, 15

Graph of two schools' SC Mathematics marks

Box and Whisker Plots

Box and whisker graphs are useful for **comparing** the data in different frequency distributions.

The diagrams consist of a rectangle which indicates the quartiles and the median, and a line on each end to show the maximum and minimum values of the distribution.

Example

Draw a box and whisker diagram to compare the following frequency distributions of two schools' School Certificate Mathematics results.

School A: Median, 51; Lower quartile, 42; Upper quartile, 65; Maximum, 95; Minimum, 23

School B: Median, 54; Lower quartile, 39; Upper quartile, 68; Maximum, 99; Minimum, 15
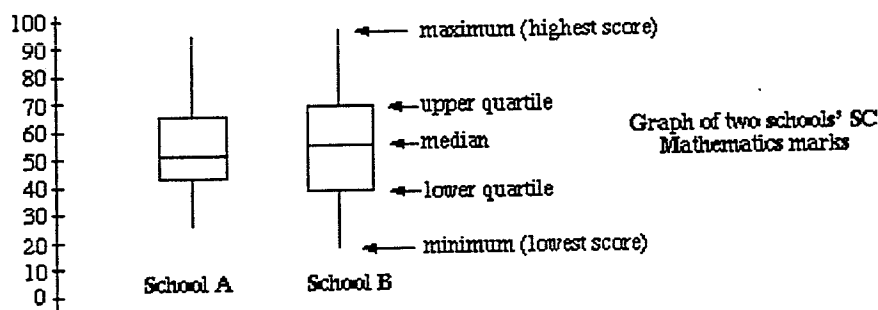
From the diagram above, the two sets of marks can now be compared.

School B would appear to have the higher marks with a higher median, upper quartile and maximum.

School A's results are less spread, with an interquartile range of 23, compared to **School B**'s 29.

Comparing Data

This experiment compares two sets of data created by you. The results are shown on histograms, a box plots and in tables.

(The table shows one or two features which are outside the scope of this course.)

u †   - ---- ----   - ---- -- -

From the diagram above, the two sets of marks can now be compared.

School B would appear to have the higher marks with a higher median, upper quartile and maximum.

School A's results are less spread, with an interquartile range of 23, compared to School B's 29.

Frequency Distributions

1. For the following scores, find:

(a) The median
(b) The mode
(c) The mean

(i) 8, 6, 4, 9, 3, 3, 2

(ii) 101, 99, 100, 98, 101, 104, 96, 101

(iii) 10, 12, 12, 12, 20

(iv) 50, 30, 20, 20, 10, 5

2. A golfer plays seven rounds of golf in a week. His scores are: 70, 62, 62, 69, 71, 72, 80

(a) Calculate the mean, median and mode of these seven scores.

(b) Which of these averages would not be very representative of his week's golf.

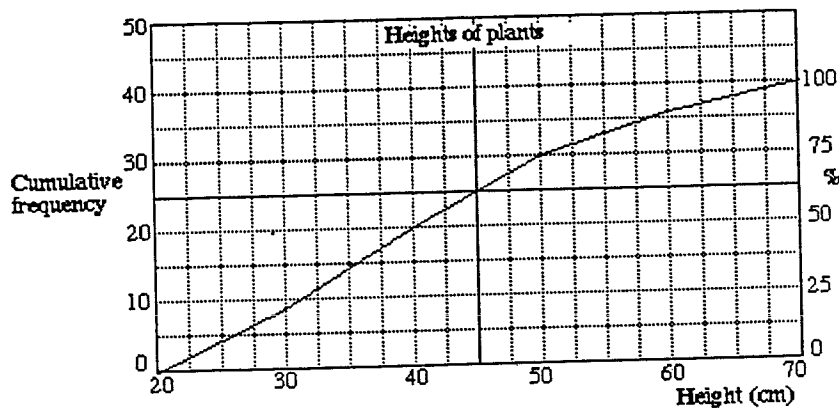3. A girl had a newspaper delivery round. Her weekly pay for 2 month's work was:

$12.50, $11.80, $10.70, $13.10, $4.50, $12.00, $9.20, $9.60

(a) What was her mean weekly pay?

(b) What was her median pay?

(c) What was the difference between her highest and lowest pays?

4. The diagram shows the height of plants, measured in centimetres.

Heights of plants

(a) How many plants were measured?

(b) What was the median plant height?

(c) What percentage of plants were over 45 cm high?

(d) What percentage of plants were less than 30 cm high?

(e) How many plants were between 50 and 60 cm high?

5. The following set of marks were for a class of 20 pupils. The test was out of 10 marks.

5, 6, 6, 3, 2, 5, 7, 6, 4, 7, 2, 0, 1, 4, 5, 6, 6, 7, 3, 8

(a) Construct a frequency distribution table.
(b) Draw a frequency histogram.
(c) Draw a frequency curve.
(d) Calculate the mean mark.
(e) Find the median mark.
(f) What is the mode?
(g) What is the probability that a person picked at random would score

    (i) Seven? (ii) Less than 5?

6. The frequency distribution for the length, in seconds, of 100 telephone calls was:

Time (seconds)	Frequency
0 - 20	0
21 - 40	5
41 - 60	7
61 - 80	14
81 - 100	28

101 - 120	21
121 - 140	13
141 - 160	9
161 - 180	3

(a) Construct a cumulative frequency table.

(b) Draw a cumulative frequency graph.

(c) What number of calls lasted no more than 2 minutes?

(d) What is the probability that a call picked at random lasted more than 140 seconds?

(e) What percentage of calls lasted more than 1 minute?

7. State whether the following would be discrete or continuous data:

(a) The weights of people.

(b) The number of people at a concert.

(c) The heights of trees.

(d) A person's shoe size.

8. The following list is the temperatures, in °C, at 20 main centres of New Zealand for a day in January.

23, 27, 19, 28, 24, 26, 27, 24, 18, 23, 15, 24, 23, 22, 24, 29, 16, 24, 22, 25

(a) Construct a frequency table with class intervals of 3°C.

(b) Draw a histogram to illustrate this distribution.

(c) What is the modal temperature range?

(d) Calculate the mean temperature.

(e) What would the median temperature be?

(f) What is the probability that a centre picked at random would have a temperature greater than or equal to 21°C?

9. The cumulative frequency graph shows the volume of drink in cans of soft drink.

(a) How many cans of soft drink were measured?

(b) How many cans contained:

(i) Less than 350 mL?
(ii) More than 340 mL?
(iii) Between 320 and 370 mL?

(c) What is the probability that a randomly selected can of drink will contain less than 350 mL?

(d) If 4000 cans were produced, how many would you expect to contain less than 350 mL?

(e) What is the median amount of drink in a can?

cumulative
frequency
< x

Cumulative frequency
graph for volume of soft
drink

millilitres x
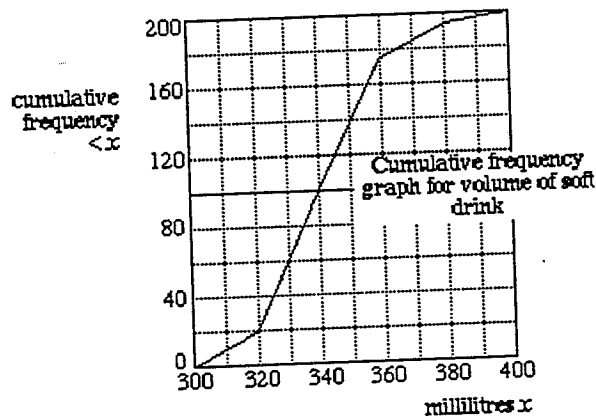
10. The table shows the lengths of 17 New Zealand rivers.

Show this information on an ordered stem and leaf diagram, using the first two digits as the stem.

Name of river	Length (in km)
Waiau	169
Waiau (Southland)	217
Waimakariri	161
Clarence	209
Waitaki	209

Rangitaiki	193
Manawatu	182
Waihou	175
Mohaka	172
Oreti	203
Wairau	169
Rakaia	145
Whangehu	161
Patea	143
Mokau	158
Ngarurora	154
Buller	177

11. The final examination results of two classes are to be compared.

The results are shown below:

First class	53	43	67	87	36	56	76	45	87	73
	45	53	74	65	38	71	40	33	59	
Second class	46	45	47	67	56	59	66	50	79	80
	35	55	77	55	30	65	42	39	39	

(a) Draw a stem and leaf diagram for each class.

(b) Arrange each set of results in order and calculate the median and the upper and lower quartiles.

(c) Draw box and whisker plots, side by side and compare the results of the two classes.

Frequency Distributions

1.
(i)	(a) 4	(b) 3 .	(c) 5
(ii)	(a) 100.5	(b) 101	(c) 100
(iii)	(a) 12	(b) 12	(c) 13.2
(iv)	(a) 20	(b) 20	(c) 22.5

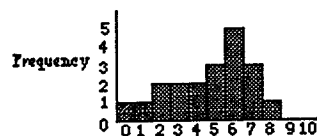2. (a) Mean = 69.4 (to 1 d.p.) Median = 70 Mode = 62

(b) The mode

3.
(a) $10.43   (b) $11.25   (c) $8.60

4. (a) 40 (b) 40 cm (c) $37 \pm 2\%$ (d) $21 \pm 2\%$ (e) 7 ± 1
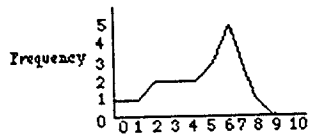
5.(a)

Score (x)	Frequency (f)	x.f
0	1	0
1	1	1
2	2	4
3	2	6
4	2	8
5	3	15
6	5	30
7	3	21
8	1	8
9	0	0
10	0	0
Totals	**20**	**93**

(b) Histogram of test results

(c) Frequency curve of test results.

Marks

(d) 4.65 (e) 5 (f) 6

(g) (i) 0.15 (ii) 0.4

6. (a)

Time (t) (seconds)	Cumulative frequency
≤ 20	0
≤ 40	5
≤ 60	12
≤ 80	26
≤ 100	54
≤ 120	75
≤ 140	88
≤ 160	97
≤ 180	100

(b)
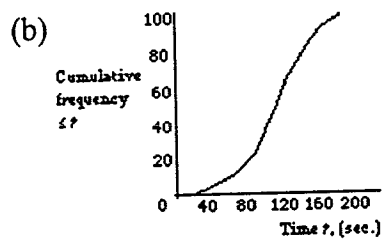
(c) 75 (d) 0.12 (e) 88%

7. (a) Continuous (b) Discrete

(c) Continuous (d) Discrete

8. (a)

Temp $^{\circ}$C	Frequency	Midpoint	f. x
15 - 17	2	16	32
18 - 20	2	19	38
21 - 23	5	22	110
24 - 26	7 .	25	175
27 - 29	4	28	112
	20		**467**

(b) Histogram of temperatures
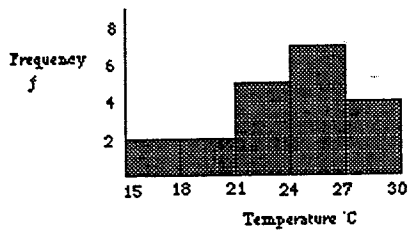
(c) 24 - 26°C (d) 23.35°C (e) 24°C (f) 0.8

9. (a) 200
(b) (i) 140 ± 5 (ii) 100 (iii) 170 ± 5

(c) 0.7

(d) 2800

(e) 340 mL

10.

```
14 | 3 5        Length of NZ rivers
15 | 4 8
16 | 1 1 9 9
17 | 2 5 7    Stem represents:
18 | 2
19 | 3        hundreds and tens
20 | 3 9 9     e.g. 17 represents 170
21 | 7
```

11. (a)

```
First class          Second class

    8 6 3 | 3 | 0 5 9 9
  5 5 3 0 | 4 | 2 5 6 7
  9 6 3 3 | 5 | 0 5 5 6 9
      7 5 | 6 | 5 6 7
  6 4 3 1 | 7 | 7 9
      7 7 | 8 | 0
                  Stem represents:  tens
                  e.g. 6 represents 60
```
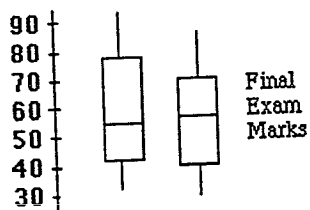
(b) There are 19 results in each class, therefore the10th score is the median, the 5th score is the lower quartile and the 15th score is the upper quartile.

From diagram:

First class:	median = 56	Second class:	median = 55
	upper quartile = 73		upper quartile = 66
	lower quartile = 43		lower quartile = 42

(c)

First class   Second class

Medians are similar, the first class has a wider spread and range of marks although the second class has the lowest mark.

Frequency Distributions

Unit Test #49

Select your answers to the following 10 questions from the pop-up menus in the right hand column. When you are satisfied with your answers, fill in your name in the space provided below the test, and click the **"Submit Test"** button. Clicking the "Begin Test Again" button will clear all the answers.

Q1:	Questions 1 to 4 refer to the following set of data: 2, 7, 7, 8, 9, 12, 15, 20 What is the **mode**?	A. 8.5 B. 7 C. 10 D. 18	**Answer 1:**	
Q2:	What is the **mean**?	A. 8.5 B. 7 C. 10 D. 18	**Answer 2:**	
Q3:	What is the **median**?	A 8.5 B. 7 C. 10 D. 18	**Answer 3:**	
Q4:	What is the **range**?	A. 8.5 B. 7 C. 10 D. 18	**Answer 4:**	
Q5:	The heights of 100 people are measured. What type of data is this?	A. continuous B. discrete	**Answer 5:**	
Q6:	In one sample of 100 pods of peas, the mean was 6 peas per pod. In another sample of 50 pods, the mean was 8 peas per pod. The combined mean for the two samples taken together was	A. exactly 6 peas per pod B. between 6 and 7 peas per pod C. exactly 7 peas per pod D. between 7 and 8 peas per pod	**Answer 6:**	
Q7:	A group of students sat a test. Their mean score was 11. Another student who scored 29 joins the group. The mean is now 13. The number of students in the	A. 5 B. 8 C. 11 D. 18	**Answer 7:**	

	original group was			
Q8:	Score: 70 71 72 73 74 Frequency: 1 3 5 4 7 David's best 20 golf scores are given in the table above. The **median** is	A. 72 B. 72.65 C. 73 D. 74	Answer 8:	▾
Q9:	In the table given above, the **mean** is	A. 72 B. 72.65 C. 73 D. 74	Answer 9:	▾
Q10:	In the table above what is the **mode** score?	A. 71 B. 72 C. 73 D. 74	Answer 10:	▾

Enter your initial and surname here: [   ]

[Submit Test] [Begin Test Again]