

Section 3.1.2 – Representing Distributions

VCAA “Dot Points”

Investigating data distributions, including:

- review of representation, display and description of the distributions of numerical variables: dot plots, stem plots, histograms.

Most of the common methods of representing statistical distributions, such as dot plots, stem plots, histograms etc., have already been presented in Year 11 General maths. This section of the notes is designed as a quick revision of the main features of each type of representation.

Stem Plots

A **Stem and Leaf Plot** is a type of graph that is used to order and display a set of data. The Stem-and-Leaf Plot summarizes the shape of a set of data (the distribution) and provides extra detail regarding individual values.

A five number summary can also be calculated from a stem and leaf plot.

The data is arranged in **increasing order**. The digits in the largest place are referred to as the **stem** and the digits in the smallest place are referred to as the **leaf** (leaves). The leaves are displayed to the right of the stem. Stem and Leaf Plots are great organizers for large amounts of information.

Example.1

The following daily temperatures (°C) were recorded for the month of April:

12, 22, 32, 18, 24, 28, 31, 14, 8, 20, 16, 15, 26, 24, 17, 30, 19, 23, 18, 19, 9, 21, 26, 11, 9, 13, 20, 27, 24, & 27

Task.1 Construct a stem and leaf plot for the above data

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

Key: 1|5 = 15°C

NB: Data is entered into the stem and leaf in **increasing order**.
The shape formed by the stem and leaf provides a vertically orientated distribution.

Task.2 Calculate a 5-number summary upon the data.

Min = 8°C

$Q_1 = 15^{\circ}\text{C}$

Med = 20°C

$Q_3 = 26^{\circ}\text{C}$

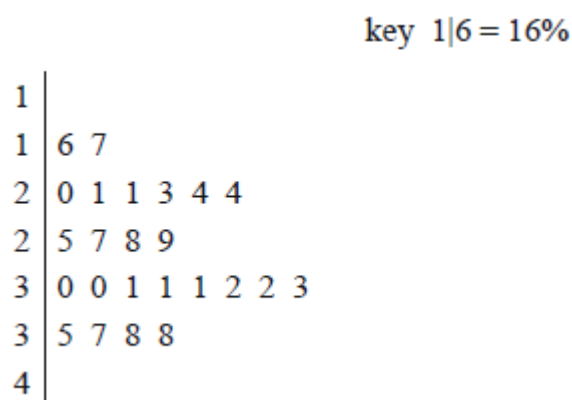
Max = 32°C

Task.3 Describe the shape/type of distribution

The distribution of the temperature is symmetrical. There is no obvious apparent tail on either side of the distribution.

Use the following information to answer Examples 2 and 3.

The following ordered stem plot shows the percentage of homes connected to broadband internet for 24 countries in 2007.



Example 2

The number of these countries with more than 22% of homes connected to broadband internet in 2007 is

- A. 4
- B. 5
- C. 19
- D. 20
- E. 22

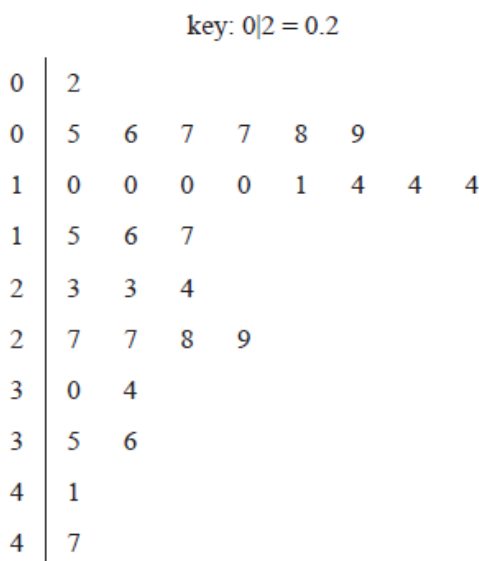
C

Reading from off the stem plot, there are 19 “scores” entered which have a percentage greater than 22%

\therefore **Option C**

Use the following information to answer Examples 5 and 6.

The stem plot below displays the average number of decayed teeth in 12-year-old children from 31 countries.



Example 5

Based on this stem plot, the distribution of the average number of decayed teeth for these countries is best described as

- A. negatively skewed with a median of 15 decayed teeth and a range of 45
- B. positively skewed with a median of 15 decayed teeth and a range of 45
- C. approximately symmetric with a median of 1.5 decayed teeth and a range of 4.5
- D. negatively skewed with a median of 1.5 decayed teeth and a range of 4.5
- E. positively skewed with a median of 1.5 decayed teeth and a range of 4.5

E

Based on the key, the data range is between 0.2 – 4.7. So a median value of 12 or range of 45 are impossible.

The stem plot has a tail towards the higher/larger numbers and so it is a positively skewed distribution.

∴ **Option E**

Example 6

For an ordered set of data containing an odd number of values, the middle value is always

- A. the mean.
- B. the median.
- C. the mode.
- D. the mean and the median.
- E. the mean, the median and the mode.

B

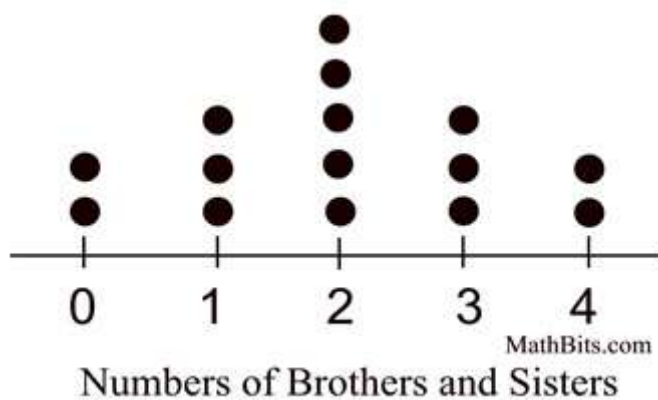
The median is always the middle value. The mode and mean are data dependent.

∴ **Option B**

Dot Plots

Another way to display statistical data is the Dot Plot. Consider the following data, where a class was asked to construct a dot plot for the number of siblings.

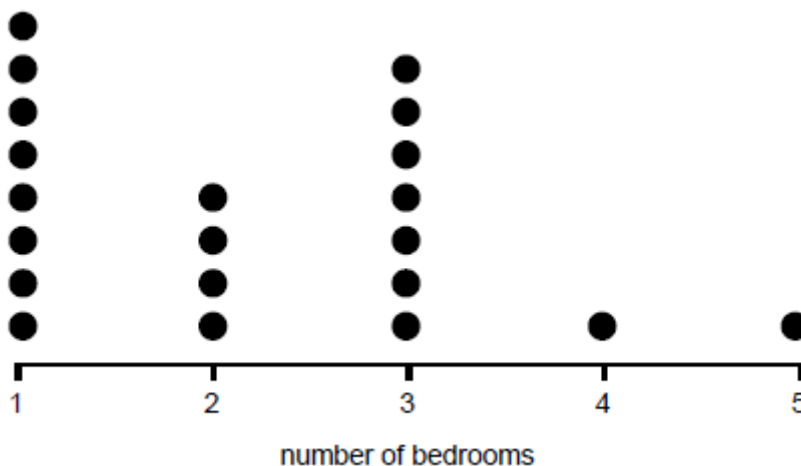
Student	Number of Brothers/Sisters
Allison	2
Bernard	4
Carlos	3
Catherine	2
Delia	2
Dion	1
Emma	0
Fiona	2
Harley	3
Ian	2
Justin	1
Paul	1
Rhianna	3
Stanley	0
Vincent	4



NB: Each dot represents a single data value.
The axis must be labelled correctly.

The following information relates to Examples 7 and 8.

The dot plot below shows the distribution of the number of bedrooms in each of 21 apartments advertised for sale in a new high-rise apartment block.



Example 7

The mode of this distribution is

- A. 1
- B. 2
- C. 3
- D. 7
- E. 8

A

The mode is the most commonly occurring value/score

∴ **Option A**

Example 8

The median of this distribution is

- A. 1
- B. 2
- C. 3
- D. 4
- E. 5

B

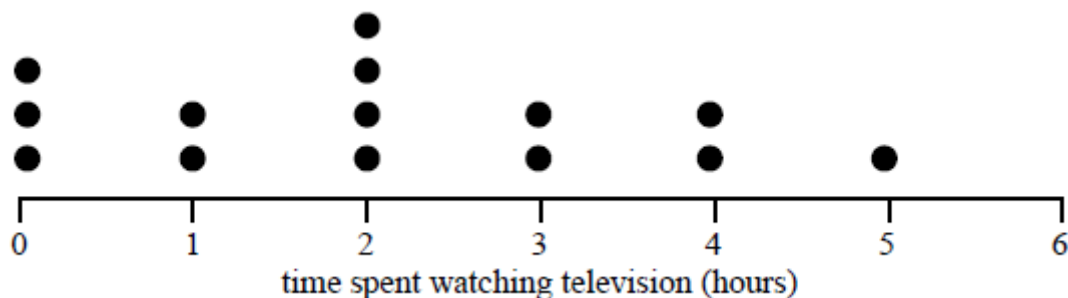
The median score = $\frac{21+1}{2} = 11^{\text{th}}$ score

11th score = 2 bedrooms

∴ **Option B**

Example 9

A sample of 14 people were asked to indicate the time (in hours) they had spent watching television on the previous night. The results are displayed in the dot plot below.



Correct to one decimal place, the mean and standard deviation of these times are respectively

- A. $\bar{x} = 2.0$ $s = 1.5$
- B. $\bar{x} = 2.1$ $s = 1.5$
- C. $\bar{x} = 2.1$ $s = 1.6$
- D. $\bar{x} = 2.6$ $s = 1.2$
- E. $\bar{x} = 2.6$ $s = 1.3$

C

Using the TI-nspire

∴ **Option C**

*Unsaved			
A	B	C	D
time			
=			=OneVar(
2	0	\bar{x}	2.07142...
3	0	Σx	29.
4	1	Σx^2	93.
5	1	$s_x := s_{n-1}$	1.59152...
6	2	$\sigma_x := \sigma_{n-1}$	1.53363...
D6	=1.5336364681131		

Frequency Tables & Histograms

Yet another way to display statistical data is via a **frequency table** or **histogram**. Consider the following data, where a class was asked to construct both a frequency table and histogram from their Year 11 exam scores in General Maths:

Class Scores – Year 11 General Maths

20 88 56 74 91 62 88 92 42 49 58 81 67
33 72 65 74 27 94 83 77 62 86 91 99 50

The **frequency table** for this data using a class interval of 10 would look like this:

Class interval describes the group/class size increments. Traditionally this is 5 or 10 or 20 etc.

In this example 20- represents values from 20 up to but not including 30.

Class interval	Tally	Frequency
20-	II	2
30-	I	1
40-	II	2
50-	III	3
60-	IIII	4
70-	IIII	4
80-	IIII	5
90-	IIII	5
	Total	26

The frequency column shows the frequency of scores from each class interval.

One simply adds up the tally for each interval and records the total in the corresponding frequency section of the table.

The frequency for the 20- interval is 2.

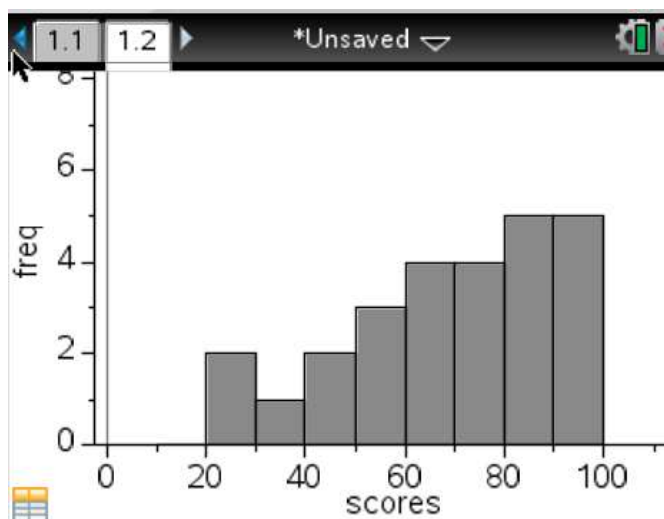
The **tally** column is used to simply tally the number of scores within the given interval.

So take the 30- interval for scores from 30 up to, but not including 30. There is a tally of two scores (II) in this interval, namely scores of 20 and 27. A single mark is added each time a score is tallied, with a cross over line added for the fifth score within a tally.

The **histogram** for this data using a **class interval of 10** would look like this:

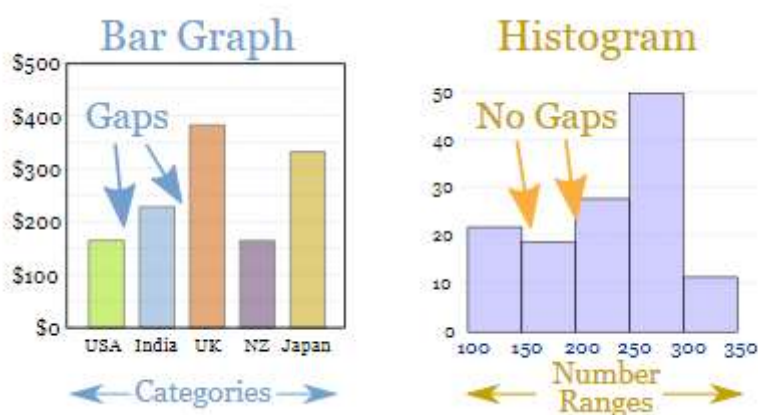
A **frequency histogram** uses the vertical columns to indicate how many times a score occurred.

The **class interval of 10** is represented by the grouping of scores upon the horizontal axes of the histogram.

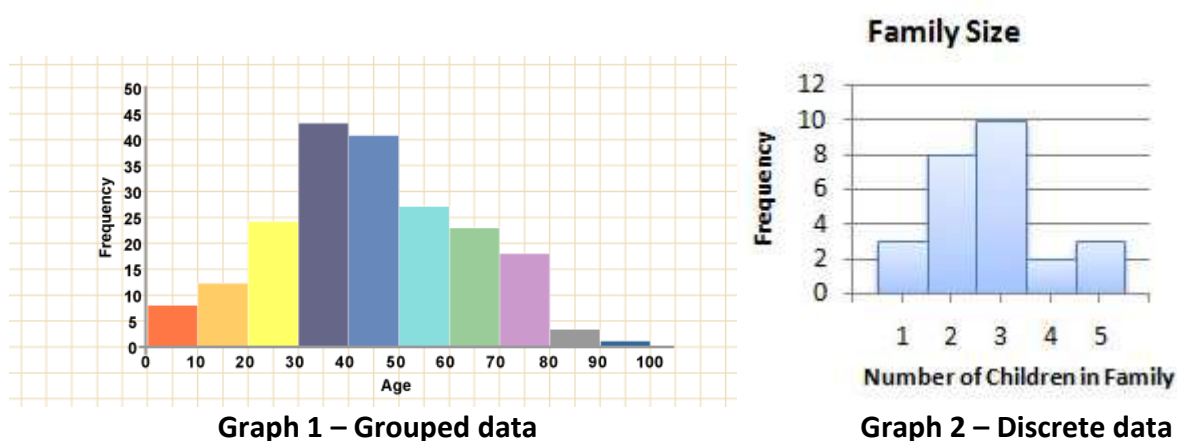


YouTube Video: <https://www.youtube.com/watch?v=vX8DOh1cTzs>

Note that **histograms** do not have a gap between the bars or columns of the graphs. **Bar Charts** on the other hand do have a gap.



The diagrams below show the difference between a histogram constructed from **grouped data** (class intervals) to that of a histogram constructed from **discrete data**.

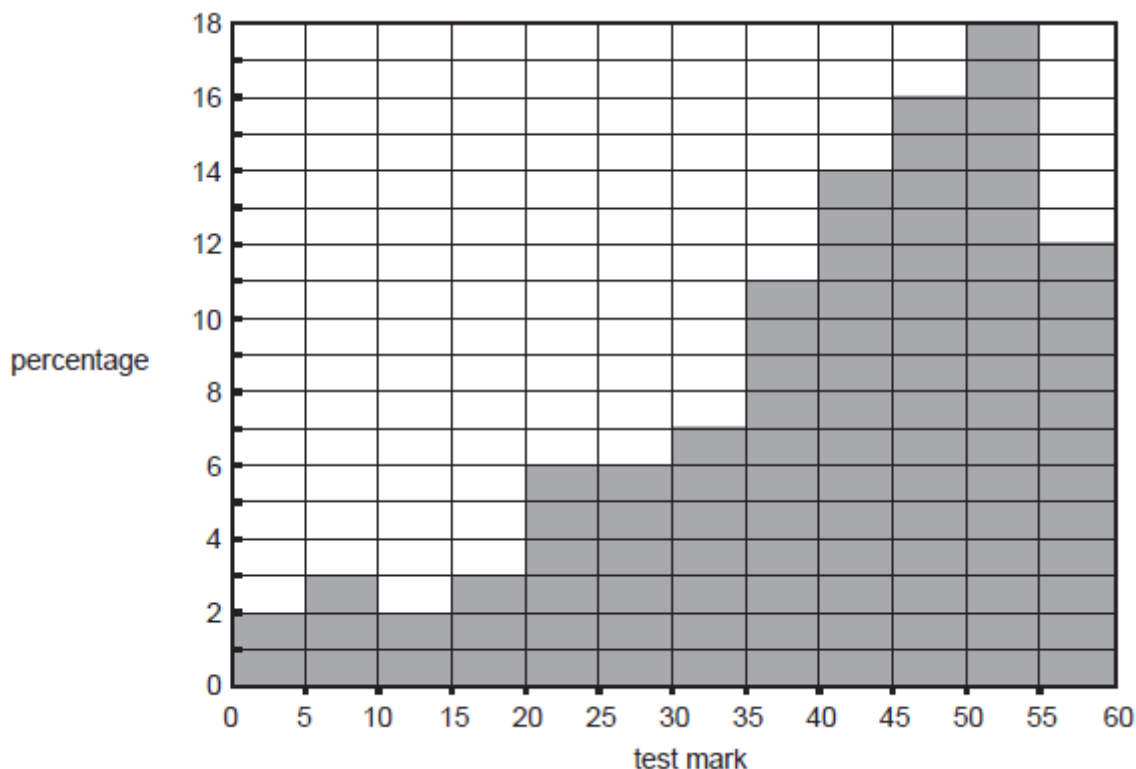


Graph 1 clearly shows how grouped data is displayed on a histogram. Each column clearly start and finishes upon an interval displayed upon the x axis. For example ages between 0-10 years are shown in the orange column.

Graph 2 clearly shows how discrete data is displayed in a histogram. Each discrete value is represented by a coloured bar or column. Each column is placed centred upon a particular value on the x axis.

The following information relates to Examples 10 and 11.

The distribution of test marks obtained by a large group of students is displayed in the percentage frequency histogram below.



Example 10

The pass mark on the test was 30 marks.

The percentage of students who passed the test is

- A. 7%
- B. 22%
- C. 50%
- D. 78%
- E. 87%

Test Marks

- 30- (7%)
- 35- (11%)
- 40- (14%)
- 45- (16%)
- 50- (18%)
- 55- (12%)

D

\therefore marks 30+ = 78%

\therefore **Option D**

Example 11

The median mark lies between

- A. 35 and 40
- B. 40 and 45
- C. 45 and 50
- D. 50 and 55
- E. 55 and 60

The histogram consist of 100 squares or scores.

\therefore the median score = $\frac{100+1}{2} = 50.5^{\text{th}}$ score

The 50th score occurs between 40 & 45 interval

The 50th score occurs between 40 & 45 interval

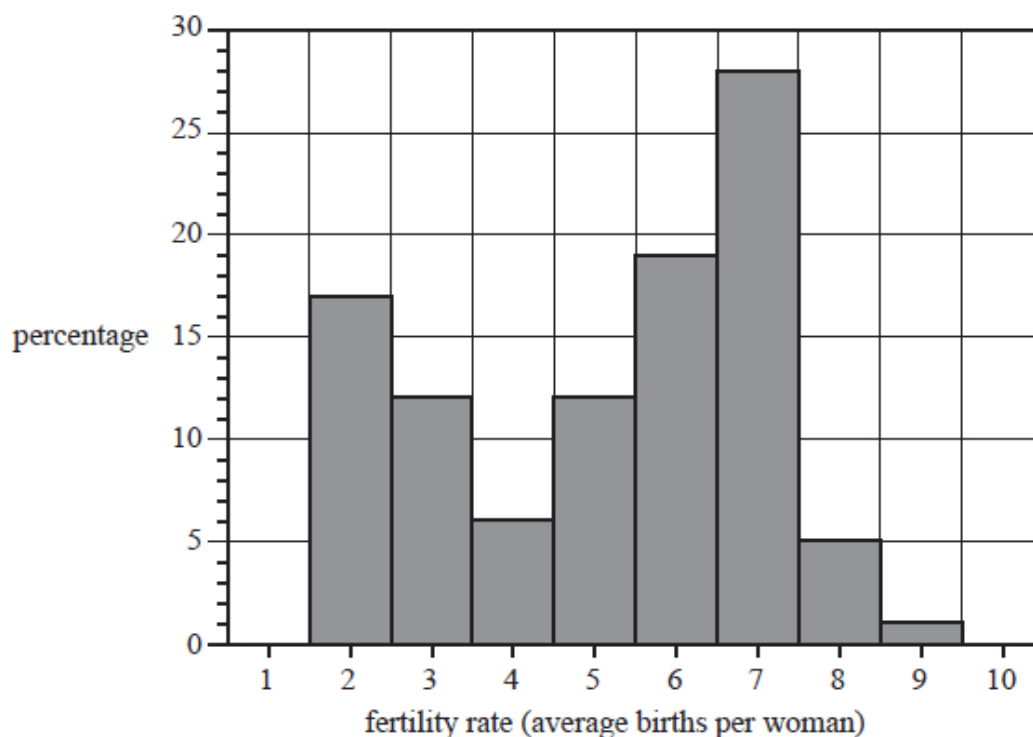
\therefore the 50.5th score occurs between 40 & 45 interval

\therefore **Option B**

B

The following information relates to Examples 12 to 13.

The percentage histogram below shows the distribution of the fertility rates (in average births per woman) for 173 countries in 1975.



Example 12

In 1975, the percentage of these 173 countries with fertility rates of 4.5 or greater was closest to

- | | |
|--------|-----------------|
| A. 12% | Fertility rates |
| B. 35% | 5 (12%) |
| C. 47% | 6 (19%) |
| D. 53% | 7 (28%) |
| E. 65% | 8 (5%) |
| | 9 (1%) |

E

\therefore fertility rates $\geq 4.5 = 65\%$

\therefore **Option E**

Example 13

In 1975, for these 173 countries, fertility rates were most frequently

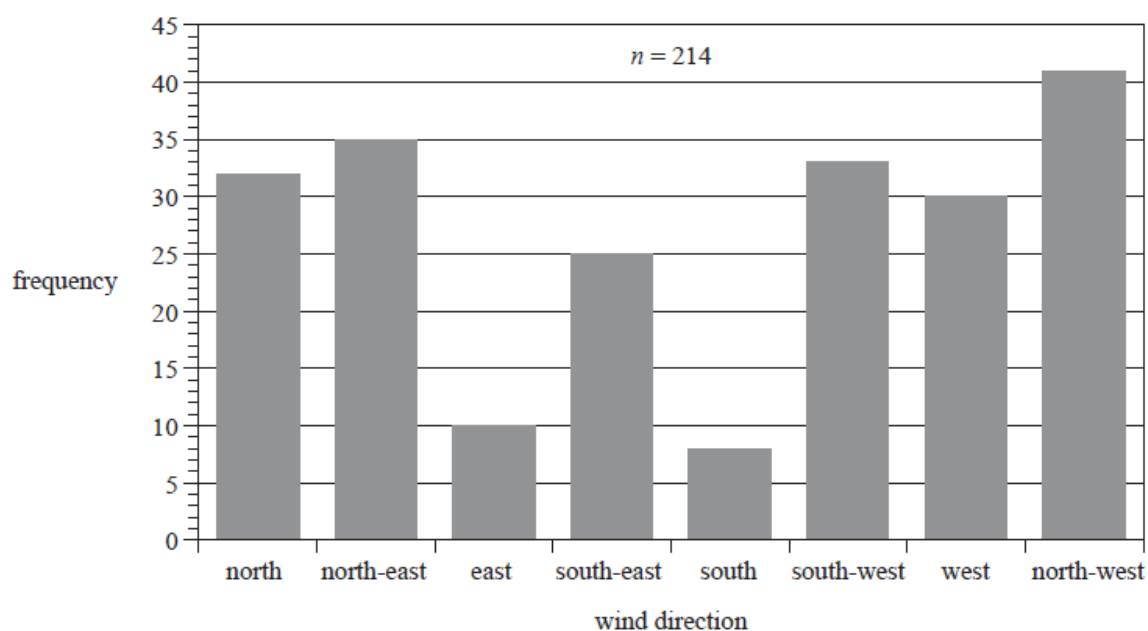
- A. less than 2.5
- B. between 1.5 and 2.5
- C. between 2.5 and 4.5
- D. between 6.5 and 7.5
- E. greater than 7.5

D

The most frequent class or interval is between 6.5 & 7.5 births per woman. \therefore **Option D**

Use the following information to answer Examples 14 and 15.

The following **bar chart** shows the distribution of wind directions recorded at a weather station at 9.00 am on each of 214 days in 2011.



Example 14

According to the bar chart, the most frequently observed wind direction was

- A. south-east.
- B. south.
- C. south-west.
- D. west.
- E. north-west.

E

The most frequent wind direction is north-west \therefore **Option E**

Question 15

According to the bar chart, the percentage of the 214 days on which the wind direction was observed to be east or south-east is closest to

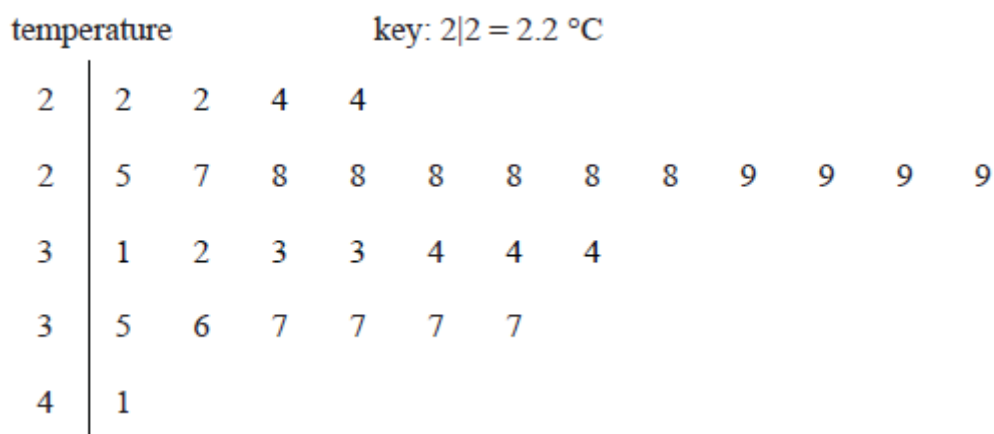
- | | |
|--------|---|
| A. 10% | Wind direction |
| B. 16% | East (10) |
| C. 25% | South East (25) |
| D. 33% | |
| E. 35% | \therefore wind directions of East or South East = 35 occasions |
| | \therefore percentage occurrence = $\frac{35}{214} \times 100 = 16.4\%$ |
| B | \therefore Option B |

Exam Styled Questions (current study design) – Multiple Choice

Question 1

(2016 Exam 1 Section A – Qn 3)

The stem plot below displays 30 temperatures recorded at a weather station.



The modal temperature is

- A.** 2.8 °C
B. 2.9 °C
C. 3.7 °C
D. 8.0 °C
E. 9.0 °C

A

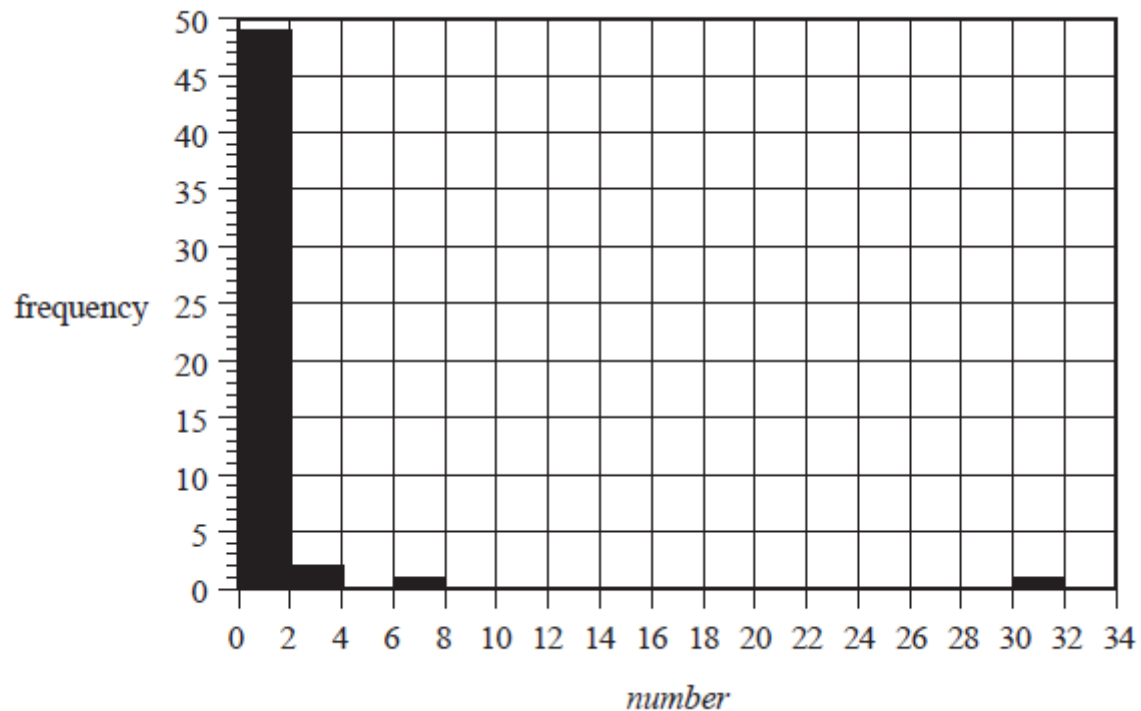
2.8 °C is the most commonly/frequently occurring.

∴ Option A

Question 2

(2016 Exam 1 Section A – Qn 6)

The histogram below shows the distribution of the *number* of billionaires per million people for 53 countries.



Using this histogram, the percentage of these 53 countries with less than two billionaires per million people is closest to

- A. 49%
- B. 53%
- C. 89%
- D. 92%
- E. 98%

D

There are 49 countries with less than two billionaires per million people.

As a percentage this is $= \frac{49}{53} \times 100 = 92.5\%$

\therefore **Option D**

(2016 Sample Exam 1 Section A – Qn 1)

The following stem plot shows the areas, in square kilometres, of 27 suburbs of a large city.

key: $1|6 = 1.6 \text{ km}^2$

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

The median area of these suburbs, in square kilometres, is

- A.** 3.0
B. 3.1
C. 3.5
D. 30.1
E. 30.5

B

The median score = $\frac{27+1}{2} = 14^{\text{th}}$ score

14th score = 3.1 km²

∴ Option B