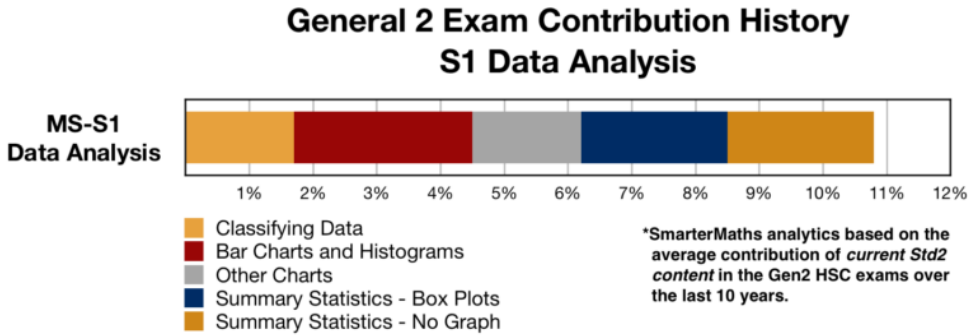
STD 1: Statistical Analysis (Std 1), S1 Data Analysis (Y11)

Summary Statistics - No Graph (Std 1)
Summary Statistics - Box Plots (Std 1)

Teacher: Kirtana Hariharan

Exam Equivalent Time: 111 minutes *(based on HSC allocation of 1.5 minutes approx. per mark)*



IMPORTANT FEATURES AND TIPS FROM 2UG EXAM HISTORY

- *MS-S1 Data Analysis* was a major contributor to the old Gen2 course, contributing an average of 10.8% per exam over the past decade (note past allocations are no guarantee of future contributions but can nonetheless cast light on a topic's likely importance).
- This analysis looks at the sub-topic *Summary Statistics - No Graph (2.3%).*

ANALYSIS - What to Expect and Common pitfalls

- *Summary Statistics - no Graph* questions require students to calculate statistics such as *median, mean and standard deviation* (by calculator) given a simple data set.
- A core competency of calculating a five number summary from a data set is a must, along with understanding mean and median changes when a dataset is adjusted.
- ***Pitfalls:*** The *2017 (Q30a)* and *2015 (Q27d)* exams exposed a lack of understanding in what constitutes an "outlier" and related calculations. This area is at the highest level of difficulty for Standard 1 exam questions in our view, but should be revised thoroughly.
- *Marker Comments* have highlighted past issues for many students in finding the mean of grouped data, where they must use the "class centres" for their calculations. This caused major issues in both 2014 and 2006 and deserves attention.

Questions

1. Data, 2UG 2005 HSC 1 MC

What is the mean of the set of scores?

$$3,\ 4,\ 5,\ 6,\ 6,\ 8,\ 8,\ 8,\ 15$$

(A) 6

(B) 7

(C) 8

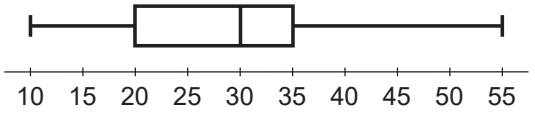
(D) 9

---

2. Data, 2UG 2017 HSC 1 MC

RAP Data - Bottom 19%: School result (100%) was 5% above state average (95%)

The box-and-whisker plot for a set of data is shown.



What is the median of this set of data?

A. 15

B. 20

C. 30

D. 35

---

3. Data, 2UG 2009 HSC 3 MC

The eye colours of a sample of children were recorded.

When analysing this data, which of the following could be found?

(A) Mean

(B) Median

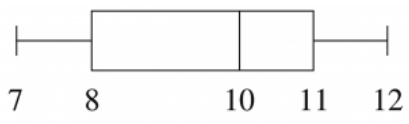
(C) Mode

(D) Range

---

4. Data, 2UG 2004 HSC 12 MC

This box-and-whisker plot represents a set of scores.



What is the interquartile range of this set of scores?

(A) 1

(B) 2

(C) 3

(D) 5

---

5. Data, 2UG 2018 HSC 1 MC

A set of scores has the following five-number summary.

lower extreme = 2
lower quartile = 5
median = 6
upper quartile = 8
upper extreme = 9

What is the range?

A. 2

B. 3

C. 6

D. 7

---

6. Data, STD2 S1 SM-Bank 2 MC

A dataset has the following five-number summary.

Minimum value	?
First quartile	10
Median	12
Third quartile	13
Maximum value	15

If the range of the dataset is 8, what is the minimum value of the dataset?

A. 2

B. 3

C. 4

D. 7

---

7. Data, 2UG 2004 HSC 6-7 MC

Use the set of scores 1, 3, 3, 3, 4, 5, 7, 7, 12 to answer Questions 6 and 7.

Question 6

What is the range of the set of scores?

(A) 6

(B) 9

(C) 11

(D) 12

Question 7

What are the median and the mode of the set of scores?

(A) Median 3, mode 5

(B) Median 3, mode 3

(C) Median 4, mode 5

(D) Median 4, mode 3

---

8. Data, 2UG 2011 HSC 14 MC

A data set of nine scores has a median of 7.

The scores 6, 6, 12 and 17 are added to this data set.

What is the median of the data set now?

(A) 6

(B) 7

(C) 8

(D) 9

---

9. Data, 2UG 2011 HSC 17 MC

The heights of the players in a basketball team were recorded as 1.8 m, 1.83 m, 1.84 m, 1.86 m and 1.92 m. When a sixth player joined the team, the average height of the players increased by 1 centimetre.

What was the height of the sixth player?

(A) 1.85 m

(B) 1.86 m

(C) 1.91 m

(D) 1.93 m

---

10. Data, 2UG 2013 HSC 14 MC

The July sales prices for properties in a suburb were:

$552 000, $595 000, $607 000, $607 000, $682 000, and $685 000.

On 1 August, another property in the same suburb was sold for over one million dollars.

If the property had been sold in July, what effect would it have had on the mean and median sale prices for July?

(A) Both the mean and median would have changed.

(B) Neither the mean nor the median would have changed.

(C) The mean would have changed and the median would have stayed the same.

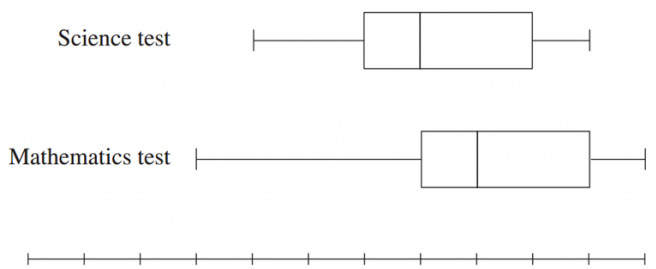
(D) The mean would have stayed the same and the median would have changed.

---

11. Data, 2UG 2008 HSC 10 MC

RAP Data - Bottom 16%: School result (68%) was 4% above state average (64%)

The marks for a Science test and a Mathematics test are presented in box-and-whisker plots.



Which measure must be the same for both tests?

(A) Mean

(B) Range

(C) Median

(D) Interquartile range

---

12. Data, 2UG 2008 HSC 13 MC

The height of each student in a class was measured and it was found that the mean height was 160 cm.

Two students were absent. When their heights were included in the data for the class, the mean height did not change.

Which of the following heights are possible for the two absent students?

(A) 155 cm and 162 cm

(B) 152 cm and 167 cm

(C) 149 cm and 171 cm

(D) 143 cm and 178 cm

---

13. Data, 2UG 2008 HSC 8 MC

What is the median of the following set of scores?

Score	*Frequency*
12	13
14	6
16	2
18	12
Total	33

(A) 12

(B) 13

(C) 14

(D) 15

---

14. Data, 2UG 2006 HSC 12 MC

The mean of a set of 5 scores is 62.

What is the new mean of the set of scores after a score of 14 is added?

(A) 38

(B) 54

(C) 62

(D) 76

---

15. Data, 2UG 2007 HSC 16 MC

Leanne copied a two-way table into her book.

	Male	*Female*	*Totals*
Full-time work	279	356	635
Part-time work	187	439	716
Totals	466	885	1351

Leanne made an error in copying one of the values in the shaded section of the table.

Which value has been incorrectly copied?

(A) The number of males in full-time work

(B) The number of males in part-time work

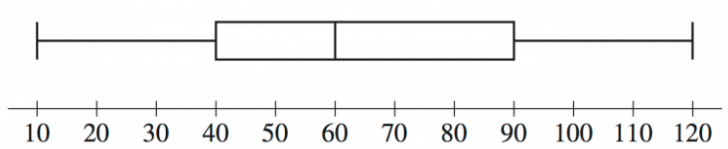
(C) The number of females in full-time work

(D) The number of females in part-time work

---

16. Data, 2UG 2015 HSC 6 MC

The times, in minutes, that a large group of students spend on exercise per day are presented in the box-and-whisker plot.



What percentage of these students spend between 40 minutes and 60 minutes per day on exercise?

(A) 17%

(B) 20%

(C) 25%

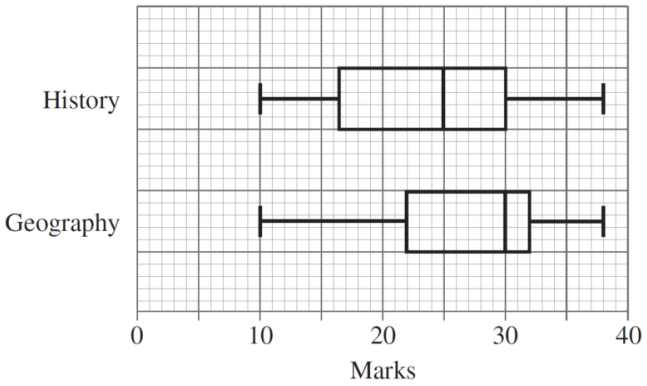
(D) 50%

---

17. Data, 2UG 2016 HSC 22 MC

RAP Data - Bottom 11%: School result (64%) was 2% above state average (62%)

The box-and-whisker plots show the results of a History test and a Geography test.



In History, 112 students completed the test. The number of students who scored above 30 marks was the same for the History test and the Geography test.

How many students completed the Geography test?

(A) 8

(B) 50

(C) 56

(D) 112

---

18. Data, 2UG 2018 HSC 11 MC

A set of data is summarised in this frequency distribution table.

Score	*Frequency*
3	1
4	2
5	6
6	7
7	9
8	5
	Total = 30

Which of the following is true about the data?

A. Mode = 7, median = 5.5

B. Mode = 7, median = 6

C. Mode = 9, median = 5.5

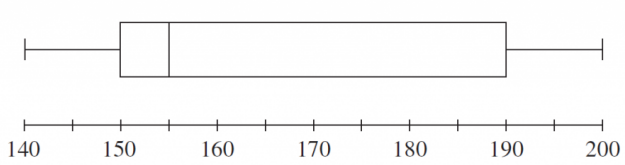
D. Mode = 9, median = 6

---

19. Data, 2UG 2011 HSC 7 MC

RAP Data - Bottom 23%: School result (54%) was 7% above state average (47%)

A set of data is displayed in this box-and-whisker plot.



Which of the following best describes this set of data?

(A) Symmetrical

(B) Positively skewed

(C) Negatively skewed

(D) Normally distributed

---

20. Data, 2UG 2007 HSC 21 MC

This set of data is arranged in order from smallest to largest.

$$5,\ 6,\ 11,\ x,\ 13,\ 18,\ 25$$

The range is six less than twice the value of x.

Which one of the following is true?

(A) The median is 12 and the interquartile range is 7.

(B) The median is 12 and the interquartile range is 12.

(C) The median is 13 and the interquartile range is 7.

(D) The median is 13 and the interquartile range is 12.

---

21. Data, 2UG 2014 HSC 14 MC

Twenty Year 12 students were surveyed. These students were asked how many hours of sport they play per week, to the nearest hour.

The results are shown in the frequency table.

Hours per week	*Frequency*
0–2	5
3–5	10
6–8	3
9–11	2

What is the mean number of hours of sport played by the students per week?

(A) 3.3

(B) 4.3

(C) 5.0

(D) 5.3

---

22. Data, 2UG 2016 HSC 19 MC

RAP Data - Bottom 7%: School result (46%) was equal to state average (46%)

A soccer referee wrote down the number of goals scored in 9 different games during the season.

2, 3, 3, 3, 5, 5, 8, 9, ...

The last number has been omitted. The range of the data is 10.

What is the five-number summary for this data set?

(A) $2, 3, 5, 8.5, 12$

(B) $2, 3, 5, 8.5, 10$

(C) $2, 3, 5, 8, 12$

(D) $2, 3, 5, 8, 10$

---

23. Data, 2UG 2016 HSC 21 MC

A grouped data frequency table is shown.

Class interval	*Frequency*
1–5	3
6–10	6
11–15	8
16–20	9

What is the mean for this set of data?

(A) 6.5

(B) 10.5

(C) 11.9

(D) 12.4

---

24. Data, 2UG 2009 HSC 21 MC

RAP Data - Bottom 2%: School result (23%) was -5% below state average (28%)

The mean of a set of ten scores is 14. Another two scores are included and the new mean is 16.

What is the mean of the two additional scores?

(A) 4

(B) 16

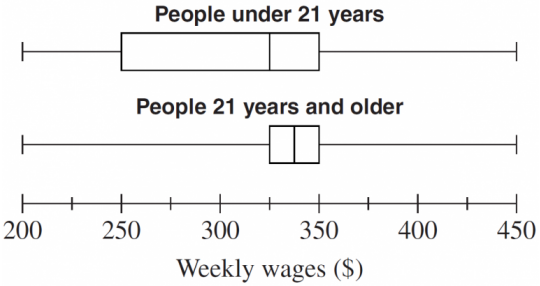
(C) 18

(D) 26

---

25. Data, 2UG 2005 HSC 22 MC

Two groups of people were surveyed about their weekly wages. The results are shown in the box-and-whisker plots.



Which of the following statements is true for the people surveyed?

(A) The same percentage of people in each group earned more than $325 per week.

(B) Approximately 75% of people under 21 years earned less than $350 per week.

(C) Approximately 75% of people 21 years and older earned more than $350 per week.

(D) Approximately 50% of people in each group earned between $325 and $350 per week.

---

26. Data, 2UG 2007 HSC 24a

Consider the following set of scores:

$$3,\ 5,\ 5,\ 6,\ 8,\ 8,\ 9,\ 10,\ 10,\ 50.$$

(i) Calculate the mean of the set of scores. *(1 mark)*

(ii) What is the effect on the mean and on the median of removing the outlier? *(2 marks)*

---

27. Data, 2UG 2013 HSC 26b

Write down a set of six data values that has a range of 12, a mode of 12 and a minimum value of 12. *(2 marks)*

---

28. Data, STD2 S1 SM-Bank 1

Write down the five-number summary for the dataset 3, 7, 8, 11, 13, 18. *(2 marks)*

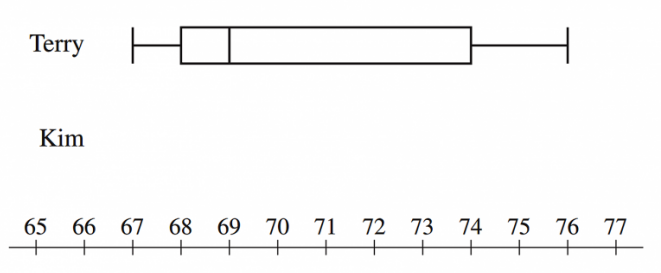
---

29. Data, 2UG 2014 HSC 29c

Terry and Kim each sat twenty class tests. Terry's results on the tests are displayed in the box-and-whisker plot shown in part (i).

(i) Kim's 5-number summary for the tests is 67, 69, 71, 73, 75.

Draw a box-and-whisker plot to display Kim's results below that of Terry's results. *(1 mark)*



(ii) What percentage of Terry's results were below 69? *(1 mark)*

(iii) Terry claims that his results were better than Kim's. Is he correct? Justify your answer by referring to the summary statistics and the skewness of the distributions. *(4 marks)*

---

30. Data, 2UG 2017 HSC 27a

Part i: RAP Data - Bottom 21%: School result (93%) was 6% above state average (87%)

Jamal surveyed eight households in his street. He asked them how many kilolitres (kL) of water they used in the last year. Here are the results.

$$220, 105, 101, 450, 37, 338, 151, 205$$

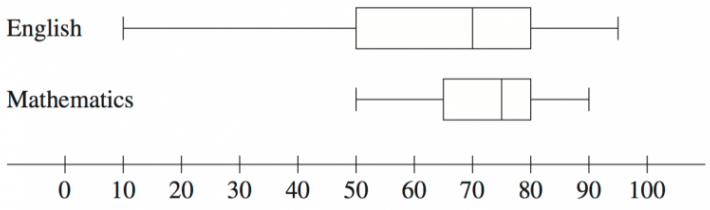
(i) Calculate the mean of this set of data. *(1 mark)*

(ii) What is the standard deviation of this set of data, correct to one decimal place? *(1 mark)*

---

31. Data, 2UG 2012 HSC 28d

The test results in English and Mathematics for a class were recorded and displayed in the box-and-whisker plots.



(i) What is the interquartile range for English? *(1 mark)*

(ii) Compare and contrast the two data sets by referring to the skewness of the distributions and the measures of location and spread. *(3 marks)*

---

32. Data, 2UG 2006 HSC 23c

Vicki wants to investigate the number of hours spent on homework by students at her high school.

(i) Briefly describe a valid method of randomly selecting 200 students for a sample. *(1 mark)*

(ii) Vicki chooses her sample and asks each student how many hours (to the nearest hour) they usually spend on homework during one week. The responses are shown in the frequency table.

Number of hours spent on homework in a week	*Frequency*
0 to 4	69
5 to 9	72
10 to 14	38
15 to 19	21

What is the mean amount of time spent on homework? *(2 marks)*

---

33. Data, 2UG 2015 HSC 27d

In a small business, the seven employees earn the following wages per week:

$$\$300,\ \$490,\ \$520,\ \$590,\ \$660,\ \$680,\ \$970.$$

i. Is the wage of $970 an outlier for this set of data? Justify your answer with calculations. *(3 marks)*

ii. Each employee receives a $20 pay increase.
What effect will this have on the standard deviation? *(1 mark)*

---

34. Probability, 2UG 2008 HSC 26b

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

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

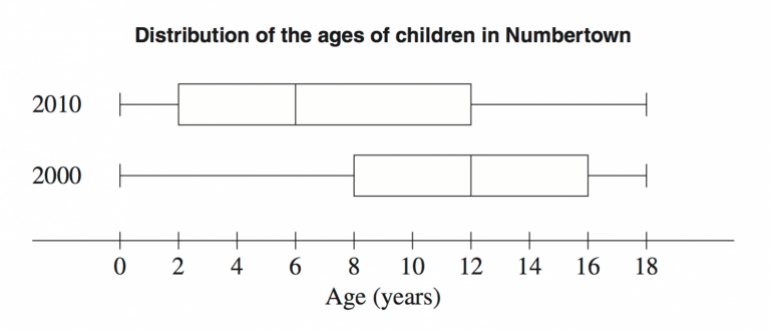
(i) What is the relative frequency of the 51–55 year retirement age? *(1 mark)*

(ii) Describe the distribution. *(1 mark)*

---

35. Data, 2UG 2010 HSC 27b

The graphs show the distribution of the ages of children in Numbertown in 2000 and 2010.



(i) In 2000 there were 1750 children aged 0–18 years.
How many children were aged 12–18 years in 2000? *(1 mark)*

(ii) The number of children aged 12–18 years is the same in both 2000 and 2010.
How many children aged 0–18 years are there in 2010? *(1 mark)*

(iii) Identify TWO changes in the distribution of ages between 2000 and 2010. In your answer, refer to measures of location or spread or the shape of the distributions. *(2 marks)*

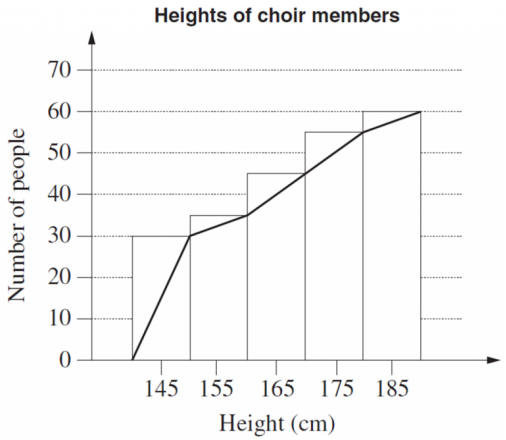
(iv) What would be ONE possible implication for government planning, as a consequence of this change in the distribution of ages? *(1 mark)*

---

36. Data, 2UG 2006 HSC 24c

The heights of the 60 members of a choir were recorded. These results were grouped and then displayed as a cumulative frequency histogram and polygon.

The shortest person in the choir is 140 cm and the tallest is 190 cm.



Draw an accurate box-and-whisker plot to represent the data. *(3 marks)*

---

37. Data, 2UG 2008 HSC 23f

Christina has completed three Mathematics tests. Her mean mark is 72%.

What mark (out of 100) does she have to get in her next test to increase her mean mark to 73%? *(2 marks)*

---

38. Data, 2UG 2017 HSC 30a

A set of data has a lower quartile (Q_L) of 10 and an upper quartile (Q_U) of 16.

What is the maximum possible range for this set of data if there are no outliers? *(2 marks)*

---

39. Data, 2UG 2018 HSC 26e

A cumulative frequency table for a data set is shown.

Score	*Cumulative frequency*
1	5
2	9
3	16
4	20
5	34
6	42

What is the interquartile range of this data set? *(2 marks)*

---

40. Data, 2UG 2005 HSC 27d

Nine students were selected at random from a school, and their ages were recorded.

Ages		
12	11	16
14	16	15
14	15	14

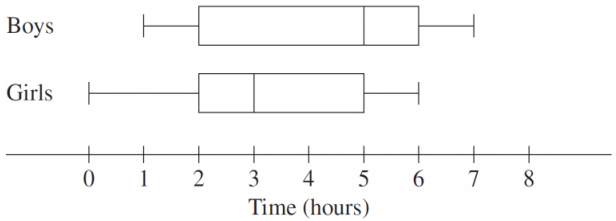
(i) What is the sample standard deviation, correct to two decimal places? *(2 marks)*

(ii) Briefly explain what is meant by the term *standard deviation*. *(1 mark)*

---

41. Data, 2UG 2009 HSC 26a

In a school, boys and girls were surveyed about the time they usually spend on the internet over a weekend. These results were displayed in box-and-whisker plots, as shown below.



(i) Find the interquartile range for boys. *(1 mark)*

(ii) What percentage of girls usually spend 5 or less hours on the internet over a weekend? *(1 mark)*

(iii) Jenny said that the graph shows that the same number of boys as girls usually spend between 5 and 6 hours on the internet over a weekend.
Under what circumstances would this statement be true? *(1 mark)*

Worked Solutions

1. Data, 2UG 2005 HSC 1 MC

$$\begin{aligned}\text{Mean} &= \frac{(3+4+5+6+6+8+8+8+15)}{9} \\ &= \frac{63}{9} \\ &= 7 \\ \Rightarrow B\end{aligned}$$

2. Data, 2UG 2017 HSC 1 MC

$\text{Median} = 30$

$\Rightarrow$ C

3. Data, 2UG 2009 HSC 3 MC

Eye colour is categorical data

$\therefore$ Only the mode can be found

$\Rightarrow C$

4. Data, 2UG 2004 HSC 12 MC

$$\begin{aligned}Q_1 &= 8 \\ Q_3 &= 11 \\ IQR &= Q_3 - Q_1 \\ &= 11 - 8 \\ &= 3 \\ \Rightarrow C\end{aligned}$$

5. Data, 2UG 2018 HSC 1 MC

$$\begin{aligned}\text{Range} &= \text{ upper extreme} - \text{lower extreme} \\ &= 9 - 2 \\ &= 7 \\ \Rightarrow &\ \text{D}\end{aligned}$$

---

6. Data, STD2 S1 SM-Bank 2 MC

$$\begin{aligned}\text{Range} &= \text{ Max Value} - \text{Min Value} \\ 8 &= 15 - \text{Min Value} \\ \therefore\ \text{Min Value} &= 15 - 8 \\ &= 7 \\ \Rightarrow &\ D\end{aligned}$$

---

7. Data, 2UG 2004 HSC 6-7 MC

Question 6

$$\begin{aligned}\text{Range} &= \text{High} - \text{Low} \\ &= 12 - 1 \\ &= 11 \\ \Rightarrow &\ C\end{aligned}$$

Question 7

9 scores

$$\begin{aligned}\therefore\ \text{Median} &= \frac{9+1}{2} \\ &= 5\text{th score} \\ &= 4\end{aligned}$$

$$\text{Mode} = 3$$

$$\Rightarrow D$$

---

8. Data, 2UG 2011 HSC 14 MC

Since an even amount of scores are added below and above the existing median, it will not change.

$$\Rightarrow B$$

---

9. Data, 2UG 2011 HSC 17 MC

$$\begin{aligned}\text{Old Mean} &= (1.8 + 1.83 + 1.84 + 1.86 + 1.92) \div 5 \\ &= \frac{9.25}{5} \\ &= 1.85\ \text{ m}\end{aligned}$$

Since the new mean $= 1.86\text{m}$ (given)

$$\text{New Mean} = \text{Height of all 6 players} \div 6$$

$$\therefore 1.86 = \frac{9.25 + h}{6} \quad (h = \text{height of new player})$$

$$\begin{aligned}h &= (6 \times 1.86) - 9.25 \\ &= 1.91\ \text{ m} \\ \Rightarrow &\ C\end{aligned}$$

---

10. Data, 2UG 2013 HSC 14 MC

Mean increases because new house is sold above the existing average.

$$\text{Initial median} = \frac{607\,000 + 607\,000}{2} = 607\,000$$

$$\text{New median} = 607\,000 \quad (4\text{th value in a list of 7})$$

$$\Rightarrow C$$

---

11. Data, 2UG 2008 HSC 10 MC

$$\text{IQR} = \text{Upper Quartile} - \text{Lower Quartile}$$

$$\Rightarrow D$$

---

12. Data, 2UG 2008 HSC 13 MC

Since the mean doesn't change

$\Rightarrow$ 2 absent students must have a mean height of 160 cm.

Considering each option given,

$(149 + 171) \div 2 = 160$

$\Rightarrow C$

---

13. Data, 2UG 2008 HSC 8 MC

$$\begin{aligned}\text{Median} &= \frac{n+1}{2} \\ &= \frac{33+1}{2} \\ &= \text{17th score}\end{aligned}$$

$\therefore$ Median is 14

$\Rightarrow C$

---

14. Data, 2UG 2006 HSC 12 MC

Mean of 5 scores $= 62$

$\therefore$ Total of 5 scores $= 62 \times 5 = 310$

Add a score of 14

Total of 6 scores $= 310 + 14 = 324$

$$\begin{aligned}\therefore \text{ New mean} &= \frac{324}{6} \\ &= 54\end{aligned}$$

$\Rightarrow B$

---

15. Data, 2UG 2007 HSC 16 MC

By checking row and column total, the number of females part-time work is incorrect

$\Rightarrow D$

---

16. Data, 2UG 2015 HSC 6 MC

$Q_1 = 40$

Median $= 60$

$\therefore$ % Students between 40 and 60

$= 50\% - 25\%$

$= 25\%$

$\Rightarrow C$

---

17. Data, 2UG 2016 HSC 22 MC

In History

$Q_3 = 30$

$\therefore$ Students scoring over 30

$= 25\% \times 112$

$= 28$

In Geography:

Median $= 30$

$\therefore$ Students that completed Geography

$= 2 \times 28 = 56$

$\Rightarrow C$

---

18. Data, 2UG 2018 HSC 11 MC

Mode $= 7$ (highest frequency of 9)

Median = average of 15th and 16th data points.

$\therefore$ Median $= 6$

$\Rightarrow$ B

---

19. Data, 2UG 2011 HSC 7 MC

Since the median (155) is closer to the lower quartile (150) and lower extreme (140) than the upper equivalents, it is positively skewed.

$\Rightarrow B$

♦ *Mean mark 47%*

---

20. Data, 2UG 2007 HSC 21 MC

$5, 6, 11, x, 13, 18, 25$

$$\begin{aligned} \text{Range} &= 2x - 6 \\ 25 - 5 &= 2x - 6 \\ 2x &= 26 \\ x &= 13 \end{aligned}$$

$\therefore$ Median $= 13$

Q1 $= 6$ Q3 $= 18$

$\therefore$ IQR $= 12$

$\Rightarrow D$

---

21. Data, 2UG 2014 HSC 14 MC

Using the class centres

$$\begin{aligned} \text{Total hours} &= (1 \times 5) + (4 \times 10) + (7 \times 3) + (10 \times 2) \\ &= 5 + 40 + 21 + 20 \\ &= 86 \end{aligned}$$

$$\text{Mean hours} = \frac{86}{20} = 4.3$$

$\Rightarrow B$

♦ *Mean mark 45%*
COMMENT: The mean is calculated using "class centres" in grouped data.

---

22. Data, 2UG 2016 HSC 19 MC

Since range is 10,

Last data point $= 12$

$Q_1 = 3$

$$Q_3 = \frac{8 + 9}{2} = 8.5$$

Median $= 5$

$\Rightarrow A$

♦ *Mean mark 46%.*

---

23. Data, 2UG 2016 HSC 21 MC

Using the centre of each class interval:

$$\begin{aligned} \text{Mean} &= \frac{3 \times 3 + 8 \times 6 + 13 \times 8 + 18 \times 9}{3 + 6 + 8 + 9} \\ &= 12.42\ldots \end{aligned}$$

$\Rightarrow D$

♦ *Mean mark 43%.*

24. Data, 2UG 2009 HSC 21 MC

♦♦♦ *Mean mark 28%.*

If $\bar{x}$ of 10 scores $= 14$

$\Rightarrow$ Sum of 10 scores $= 10 \times 14 = 140$

With 2 additional scores, $\bar{x} = 16$

$\Rightarrow$ Sum of 12 scores $= 12 \times 16 = 192$

$\therefore$ Value of 2 extra scores $= 192 - 140$

$= 52$

$\therefore$ Mean of 2 extra scores $= \dfrac{52}{2} = 26$

$\Rightarrow D$

---

25. Data, 2UG 2005 HSC 22 MC

Considering A

50% of Under 21 group earned over \$325

and 75% of Over 21 group did. NOT TRUE.

Considering B

75% of Under 21 group earned below \$350

is TRUE.

C and D can both be proven to be untrue

using their median and quartile values.

$\Rightarrow B$

---

26. Data, 2UG 2007 HSC 24a

(i) Total of scores

$= 3 + 5 + 5 + 6 + 8 + 8 + 9 + 10 + 10 + 50$

$= 114$

$\therefore$ Mean $= \dfrac{114}{10} = 11.4$

(ii) Mean

If the outlier (50) is removed, the mean

would become lower.

Median

The current median (10 data points)

$= \dfrac{5\text{th} + 6\text{th}}{2} = \dfrac{8+8}{2} = 8$

The new median (9 data points)

$=$ 5th value

$= 8$

$\therefore$ Median will NOT change.

---

27. Data, 2UG 2013 HSC 26b

12, 12, 12, 16, 18, 24

NB. There are many correct solutions.

---

28. Data, STD2 S1 SM-Bank 1

Minimum value: 3

First quartile: 7

Median: $\dfrac{11+8}{2} = 9.5$

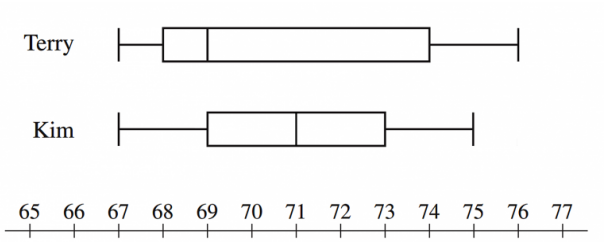
Third quartile: 13

Maximum value: 18

---

29. Data, 2UG 2014 HSC 29c

(i)



(ii) 50%

♦ *Mean mark 39%*

(iii) Terry's results are more positively skewed than Kim's and also have a higher limit high. However, Kim's results are more consistent, showing a tighter IQR. They also have a significantly higher median than Terry's and are evenly skewed.

$\therefore$ Kim's results were better.

♦♦ *Mean mark 29%*
COMMENT: Examiners look favourably on using *language of location* in answers, particularly the areas they have specifically pointed students towards (skewness in this example).

---

30. Data, 2UG 2017 HSC 27a

(i) Mean $= (220 + 105 + 101 + 450 + 37 + 338 + 151 + 205) \div 8$

$= 200.875$

(ii) Std Dev $= 127.357\ldots$ (by calc)

$= 127.4$ (1 d.p.)

♦ *Mean mark part (ii) 47%.*
IMPORTANT: The *population* standard deviation is required here.

---

31. Data, 2UG 2012 HSC 28d

(i) $IQR_{(\text{English})} = 80 - 50$

$= 30$

(ii) Skewness

- English has greater negative skew
- Maths is more normally distributed

Location and Spread

- English has a range of 85, Maths has 40.
- English has larger IQR than Maths (30 vs 15)
- Maths' Median (75) is higher than English (70)
- Same upper quartile marks (80)
- English has highest and lowest individual mark

♦ *Mean mark 35%*
MARKER'S COMMENT: Markers are looking for students to use the correct language of location and spread such as *mean, median, interquartile range, standard deviation and skewness.*

32. Data, 2UG 2006 HSC 23c

(i) A valid method would be using a stratified sample.
The number of students sampled in each year is proportional to the size of each year.

(ii)

Hours	Class Centre (x)	Frequency (f)	fx
0 - 4	2	69	138
5 - 9	7	72	504
10 - 14	12	38	456
15 - 19	17	21	357
		Total	**1455**

$$\begin{aligned}\text{Mean} &= \frac{\text{Sum of Scores}}{\text{Total scores}} \\ &= \frac{1455}{200} \\ &= 7.275 \text{ hours}\end{aligned}$$

MARKER'S COMMENT: This "routine" exercise of finding a mean from grouped data was incorrectly answered by most students! The best responses copied the table and inserted a class-centre column (see solution).

33. Data, 2UG 2015 HSC 27d

i. 300, 490, 520, 590, 660, 680, 970

$$\begin{aligned}\text{Median} &= 590 \\ Q_1 &= 490 \\ Q_3 &= 680 \\ \therefore IQR &= 680 - 490 = 190\end{aligned}$$

Outlier if value is greater than

$$\begin{aligned}&Q_3 + 1.5 \times IQR \\ &= 680 + 1.5 \times 190 \\ &= 965\end{aligned}$$

$\therefore$ The wage \$970 is an outlier.

♦ *Mean mark 39%.*

ii. All values increase by \$20, but so too does the mean. Therefore the spread about the new mean will not change and therefore the standard deviation will remain the same.

34. Probability, 2UG 2008 HSC 26b

(i) Relative frequency $(51 - 55)$

$$\begin{aligned}&= \frac{\#\text{ People (51-55)}}{\text{Total People}} \\ &= \frac{35\,000}{2\,000\,000} \\ &= \frac{7}{400}\end{aligned}$$

(ii) Distribution is negatively skewed because as age increases, so does the number of people in each age bracket.

35. Data, 2UG 2010 HSC 27b

(i) Since the median $= 12$ years

$\Rightarrow$ 50% of children are aged 12-18 years

♦ *Mean mark 45%*

$$\therefore \ \#\text{ Children 12-18} = \ 50\% \ \times 1750$$
$$= 875$$

(ii) Upper quartile (2010) $= 12$ years

$\#$ Children in upper quartile $= 875$ (from part (i))

♦♦ *Mean mark 25%*

$$\therefore \ \#\text{ Children aged 0-18} = 4 \times 875$$
$$= 3500$$

(iii) Changes in distribution (only 2 req'd)

- the lower quartile age is lower in 2010
- the median is lower in 2010
- the upper quartile age is lower in 2010
- the interquartile range is greater in 2010
- 2010 is positively skewed while 2000 is negatively

♦ *Mean mark 35%*
MARKER'S COMMENT: A number of students incorrectly identified "positive" skew as "negative" skew here.

(iv) Implication for government planning

- since the children are getting younger in 2010, approve and build more childcare facilities
- Build more school and public playgrounds.

♦ *Mean mark 46%*
MARKER'S COMMENT: This is worth only 1 mark and long-winded answers are not required.

---

36. Data, 2UG 2006 HSC 24c

Low $= 140$

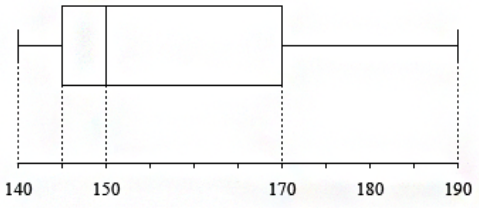
High $= 190$

Median $= 150 \quad (\#\text{ People} = 30)$

$Q_1 = 145 \quad (\#\text{ People} = 15)$

$Q_3 = 170 \quad (\#\text{ People} = 45)$

Box and Whisker



---

37. Data, 2UG 2008 HSC 23f

Total marks in 3 tests

$= 3 \times 72$

$= 216$

We need 4-test mean $= 73$

i.e. Total Marks (4 tests) $\div 4 = 73$

Total Marks (4 tests) $= 292$

$$\therefore \ 4\text{th test score} = 292 - 216$$
$$= 76$$

---

38. Data, 2UG 2017 HSC 30a

$IQR = 16 - 10 = 6$

If no outliers,

♦♦ *Mean mark 34%.*

$$\begin{aligned}\text{Upper limit} &= Q_U + 1.5 \times IQR \\ &= 16 + 1.5 \times 6 \\ &= 25\end{aligned}$$

$$\begin{aligned}\text{Lower limit} &= Q_L - 1.5 \times IQR \\ &= 10 - 1.5 \times 6 \\ &= 1\end{aligned}$$

$$\begin{aligned}\therefore \text{ Maximum range} &= 25 - 1 \\ &= 24\end{aligned}$$

---

39. Data, 2UG 2018 HSC 26e

$$42 \text{ data points} \Rightarrow \text{median} = \frac{21\text{st} + 22\text{nd}}{2}$$

♦♦ *Mean mark 27%.*

$\text{Q}_1 = 11\text{th data point} = 3$

$\text{Q}_3 = 32\text{nd data point} = 5$

$$\begin{aligned}\therefore \text{ IQR} &= 5 - 3 \\ &= 2\end{aligned}$$

---

40. Data, 2UG 2005 HSC 27d

(i) Sample standard deviation

$$\begin{aligned}&= 1.6914\ldots \text{ (by calculator)} \\ &= 1.69 \quad \text{(to 2 d.p.)}\end{aligned}$$

(ii) Standard deviation is a measure of how much members of a data group differ from the mean value of the group.

---

41. Data, 2UG 2009 HSC 26a

(i) $\text{Interquartile range} = 6 - 2$

$$= 4$$

(ii) $\text{Upper quartile} = 5$

♦♦ *Mean mark 31%*

$\therefore$ 75% of girls spend 5 or less hours

♦♦♦ *Mean mark 9%*

(iii) 5-6 hours for girls accounts for 25% of all girls.

5-6 hours for boys accounts for 25% of all boys,

(median to the upper quartile represents 25%.)

$\Rightarrow$ This will only be the same number if the number of all girls surveyed equals the number of boys surveyed.
