WORKSHEET - General 2 Mathematics

Data and Statistics

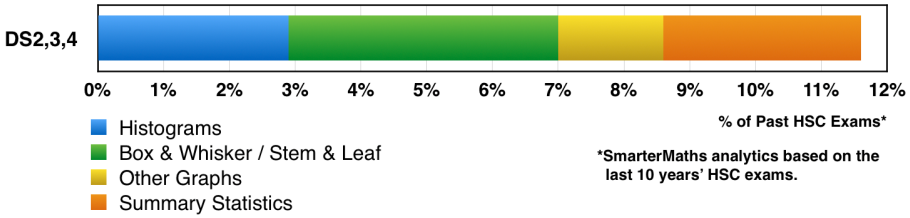
DS2/3/4 - Single Data Sets, Summary Stats and Interpretation

Stem & Leaf, Box & Whisker

Teacher: SmarterMaths

Exam Equivalent Time: 45 minutes *(based on HSC allocation of 1.5 minutes approx. per mark)*

The Big Picture: DS2,3,4 - Data & Statistics



HISTORICAL CONTRIBUTION

- ***DS2,3,4 Displaying and Interpreting Data Sets and Summary Statistics*** is one of the most important areas of the General 2 course, contributing **11.5%** on average, over the last decade.
- This analysis will focus on ***Box and Whisker*** and ***Stem and Leaf plots*** which have contributed an average of **4.1%** each year.

PAST HSC ANALYSIS - What to expect and common pitfalls

- ***Box and Whisker*** plots were examined in a solitary multiple choice question in both 2015 and 2016. In 2010, 2012, and 2014, this topic area was allocated between 4-6 marks in questions where students needed to compare two B&W plots. This question type is overdue to be tested in a significant way.
- ***Stem and Leaf*** plots were examined in a longer answer question in 2016 after 2 years of not appearing. Note that they were examined every year between 2007-2013 and remain a critical revision area.
- ***Double Stem and Leaf*** plots produced sub-50% mean marks the last 4 times they were examined (including 2016), so revision in this *specific area* is well advised.
- Although ***Box and Whisker*** and ***Stem and Leaf*** mark allocations often cannibalise each other in any one given year, they have been under-examined in recent times versus historical allocations and are an important revision focus.

Questions

1. Data, 2UG 2006 HSC 4 MC

A set of scores is displayed in a stem-and-leaf plot.

1	2 2 3
2	5 8
3	8 9
4	1 3 9

What is the median of this set of scores?

(A) 28

(B) 30

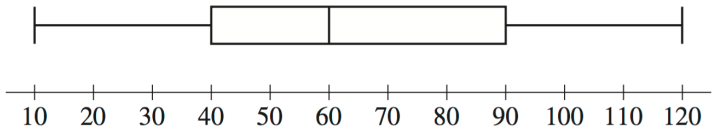
(C) 33

(D) 47

---

2. 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%

---

3. Data, 2UG 2010 HSC 16 MC

This back-to-back stem-and-leaf plot displays the test results for a class of 26 students.

Boys					Girls				
			1	2	1	2	4		
			3	3	0	2	3	5	
	9	7	4	4	4	4	5	9	9
6	4	2	2	5	3				
		3	0	6	1	9			

What is the median test result for the class?

(A) 44

(B) 46

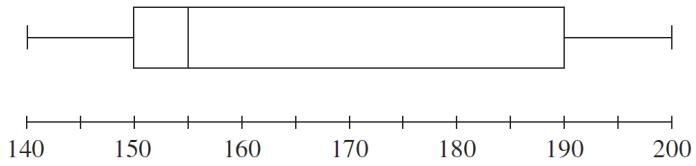
(C) 48

(D) 49

---

4. Data, 2UG 2011 HSC 7 MC

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

---

5. Data, 2UG 2016 HSC 19 MC

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

$2, 3, 3, 3, 5, 5, 8, 9, \ldots$

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

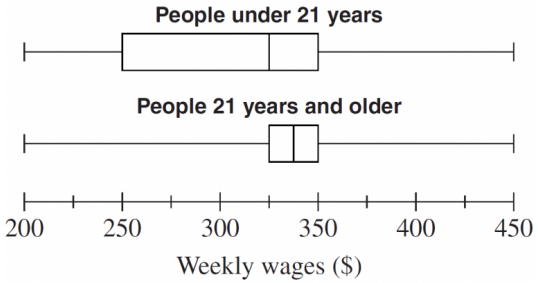
(C) $2, 3, 5, 8, 12$

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

---

6. 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.

---

7. Data, 2UG 2011 HSC 25d

Data was collected from 30 students on the number of text messages they had sent in the previous 24 hours. The set of data collected is displayed.

Male		*Female*
9 9 8 7 6 5 5 4 2 1	0	8 9
1 1 0 0	1	1 1 2 5 6 8 8 8
0	2	0 1 7
	3	4
	4	
	5	
	6	
1	7	

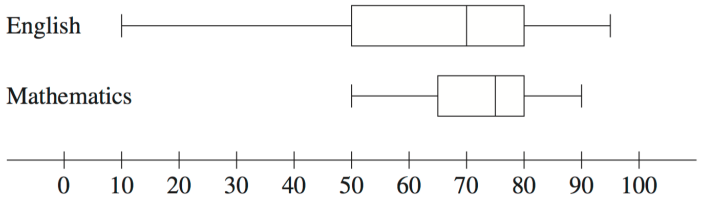
(i) What is the outlier for this set of data? *(1 mark)*

(ii) What is the interquartile range of the data collected from the female students? *(1 mark)*

---

8. 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)*

---

9. Data, 2UG 2013 HSC 26f

Jason travels to work by car on all five days of his working week, leaving home at 7 am each day. He compares his travel times using roads without tolls and roads with tolls over a period of 12 working weeks.

He records his travel times (in minutes) in a back-to-back stem-and-leaf plot.

Travel time (minutes)

Without tolls		*With tolls*
9	3	5 8 9 9
9 9 8 7 7 6 5 5 4 4 3 2 0	4	0 1 2 6 7 7 8 8 8 9
9 8 7 5 4 3 3 3 2 2 2 2 1 1 0	5	2 4 4 5 6 8 9
1	6	1 3 5 7
	7	0 2 8
	8	2
	9	0

(i) What is the modal travel time when he uses roads without tolls? *(1 mark)*

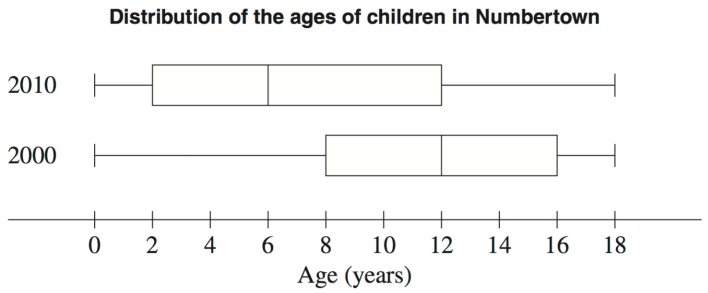
(ii) What is the median travel time when he uses roads without tolls? *(1 mark)*

(iii) Describe how the two data sets differ in terms of the spread and skewness of their distributions. *(2 marks)*

---

10. 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)*

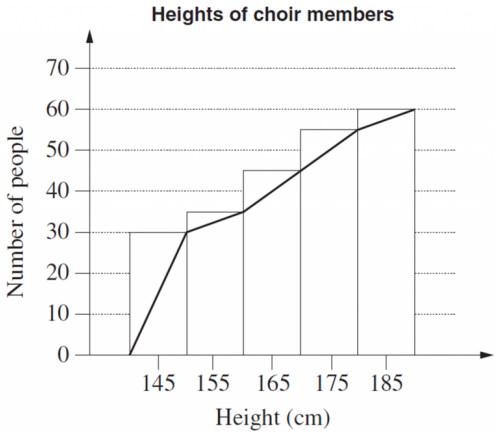
(iv) What would be ONE possible implication for government planning, as a consequence of this change in the distribution of ages? *(1 mark)*

---

11. 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)*

---

12. Data, 2UG 2016 HSC 29c

The ages of members of a dance class are shown in the back-to-back stem-and-leaf plot.

Women		*Men*
2	3	4 6
4 2	4	2 2 5 6 8
8 8 5 4 0 0	5	3
9 4 3 3	6	3

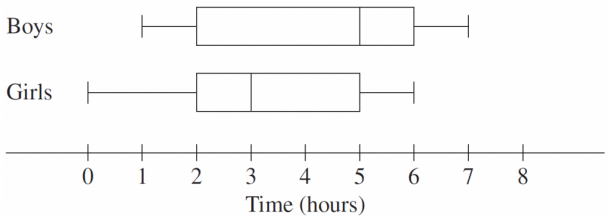
Pat claims that the women who attend the dance class are generally older than the men.

Is Pat correct? Justify your answer by referring to the median and skewness of the two sets of data. *(3 marks)*

---

13. 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 2006 HSC 4 MC

10 scores

$$\text{Median} = \frac{5\text{th} + 6\text{th}}{2}$$
$$= \frac{28 + 38}{2}$$
$$= 33$$

$\Rightarrow C$

---

2. Data, 2UG 2015 HSC 6 MC

$Q_1 = 40$

Median $= 60$

$\therefore$ % Students between 40 and 60

$= 50\% - 25\%$

$= 25\%$

$\Rightarrow C$

---

3. Data, 2UG 2010 HSC 16 MC

♦♦ *Mean mark 35%*

26 results given in the data

$\Rightarrow$ Median is average of 13^{th} and 14^{th}

$$\therefore \text{Median} = \frac{45 + 47}{2}$$
$$= 46$$

$\Rightarrow B$

Worked Solutions

4. Data, 2UG 2011 HSC 7 MC

♦ *Mean mark 47%*

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

$\Rightarrow B$

---

5. Data, 2UG 2016 HSC 19 MC

♦ *Mean mark 46%.*

Since range is 10,

Last data point = 12

$Q_1 = 3$

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

Median = 5

$\Rightarrow A$

---

6. 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.

$\Rightarrow B$

---

7. Data, 2UG 2011 HSC 25d

(i) Outlier is 71

(ii) Lower quartile = 9 (4th female data point)

Upper quartile = 20 (11th female data point)

$\therefore$ Interquartile range (female) $= 20 - 11 = 9$

♦♦ *Mean mark 34%*
COMMENT: Ensure you can quickly and accurately find quartile values using stem and leaf graphs!

8. Data, 2UG 2012 HSC 28d

(i) IQR (E) $= 80 - 50$

$= 30$

(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.*

9. Data, 2UG 2013 HSC 26f

(i) Modal time = 52 minutes

(ii) 30 times with no tolls

Median = Average of 15th and 16th

$$= \frac{50 + 51}{2}$$

$= 50.5$ minutes

♦ *Mean mark 36%*
MARKER'S COMMENT: Finding a median proved challenging for many students. Take note!

♦ *Mean mark 39%*

(iii) Spread

Times without tolls have a much tighter spread (range = 22) than times with tolls (range = 55).

Skewness

Times without tolls shows virtually no skewness while times with tolls are positively skewed.

10. Data, 2UG 2010 HSC 27b

(i) Since the median = 12 years

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

$\therefore$ # Children 12-18 $=$ 50% $\times$ 1750

$= 875$

♦ *Mean mark 45%*

(ii) Upper quartile (2010) = 12 years

Children in upper quartile = 875 (from part (i))

$\therefore$ # Children aged 0-18 $= 4 \times 875$

$= 3500$

♦♦ *Mean mark 25%*

(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.

11. Data, 2UG 2006 HSC 24c

Low $= 140$

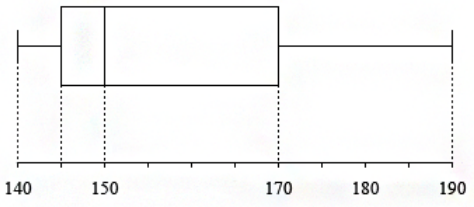
High $= 190$

Median $= 150$ (# People = 30)

$Q_1 = 145$ (# People = 15)

$Q_3 = 170$ (# People = 45)

Box and Whisker



12. Data, 2UG 2016 HSC 29c

Women:

The median is 55 in a data set that is negatively skewed.

Men:

The median is 45 in a data set that is positively skewed.

$\therefore$ Pat is correct.

♦ *Mean mark 44%.*

13. Data, 2UG 2009 HSC 26a

(i) Interquartile range $= 6 - 2$

$= 4$

(ii) 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.