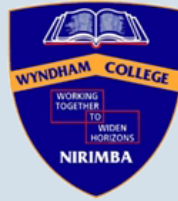


## STD 1: Statistical Analysis (Std 1), S1 Data Analysis (Y11)

### Other Charts (Std 1)

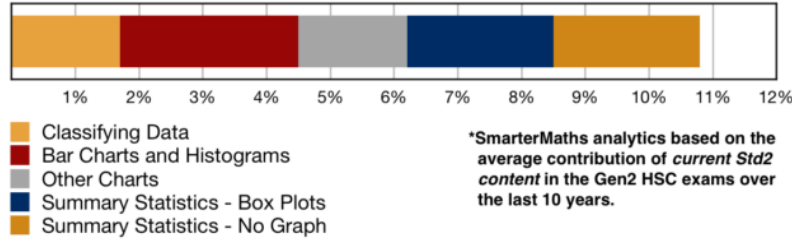
Teacher: Kirtana Hariharan

Exam Equivalent Time: 69 minutes (based on HSC allocation of 1.5 minutes approx. per mark)



### General 2 Exam Contribution History S1 Data Analysis

#### MS-S1 Data Analysis



### 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.
- This analysis looks at the largest Other Charts (1.7%).

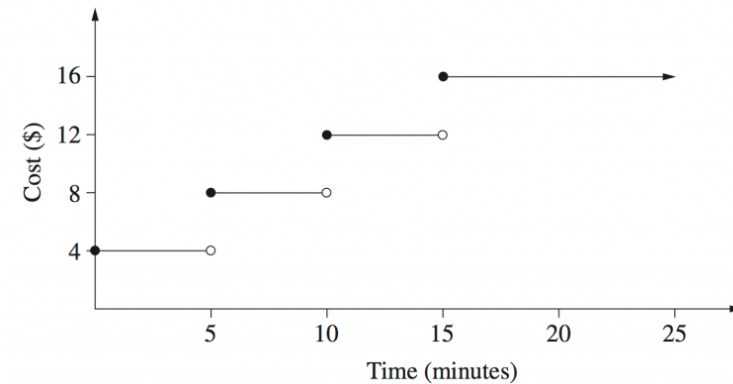
### ANALYSIS - What to Expect and Common pitfalls

- The new Std2 syllabus states generally that students "[need to be able to] .. summarise and interpret grouped and ungrouped data through appropriate graphs."
- An understanding of Stem & Leaf plots is recommended in response to this syllabus requirement, albeit not with the focus required in the old Gen2 course (where S&L plots' importance was similar to box-and-whisker charts).
- 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.
- Area charts have caused notable problems in the past and deserve particular attention.
- Other charts tested that students have answered well include pie charts, segregated bar charts, step graphs and dot plots.
- Familiarity with Pareto charts is highly recommended. This chart style gets a specific mention in the new syllabus and should be revised.

## Questions

### 1. Data, 2UG 2014 HSC 1 MC

The step graph shows the cost of telephone calls to China.




Bella made a telephone call at 9.50 pm to a friend in China and spoke to him until 10.07 pm.

How much did this telephone call cost?

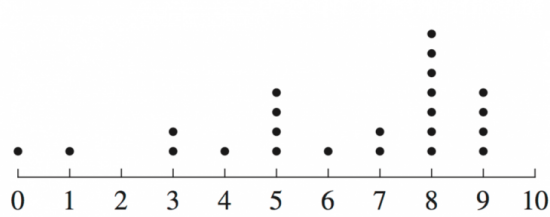
- (A) \$4
- (B) \$8
- (C) \$12
- (D) \$16

## 2. Data, 2UG 2010 HSC 1 MC

 RAP Data - Bottom 6%: School result (76%) was equal to state average (76%)


The results of a survey are displayed in the dot plot.

What is the range of this data?

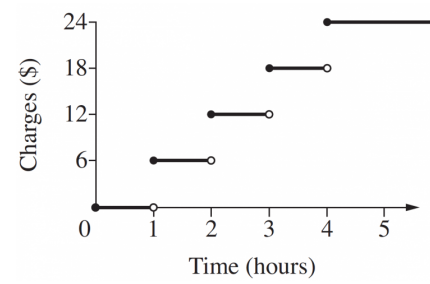


- (A) 7
  - (B) 8
  - (C) 9
  - (D) 10
- 

## 3. Data, 2UG 2009 HSC 2 MC

 RAP Data - Bottom 22%: School result (89%) was 6% above state average (83%)

The step graph shows the charges for a carpark.



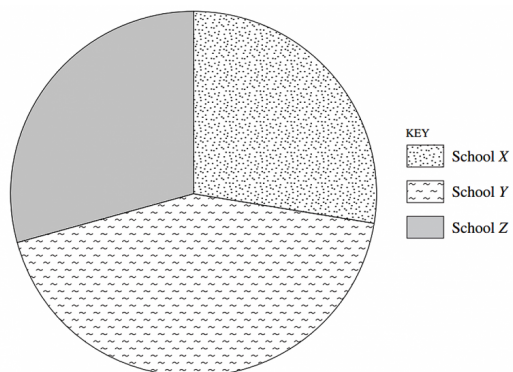
Maria enters the carpark at 10:10 am and exits at 1:30 pm.

How much will she pay in charges?

- (A) \$6
  - (B) \$12
  - (C) \$18
  - (D) \$24
-

4. Data, 2UG 2004 HSC 8 MC

This sector graph shows the distribution of 116 prizes won by three schools: X, Y and Z.

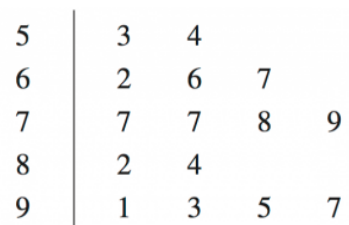


How many prizes were won by School X?

- (A) 26
  - (B) 32
  - (C) 81
  - (D) 99
- 

5. Data, 2UG 2012 HSC 1 MC

A set of 15 scores is displayed in a stem-and-leaf plot.



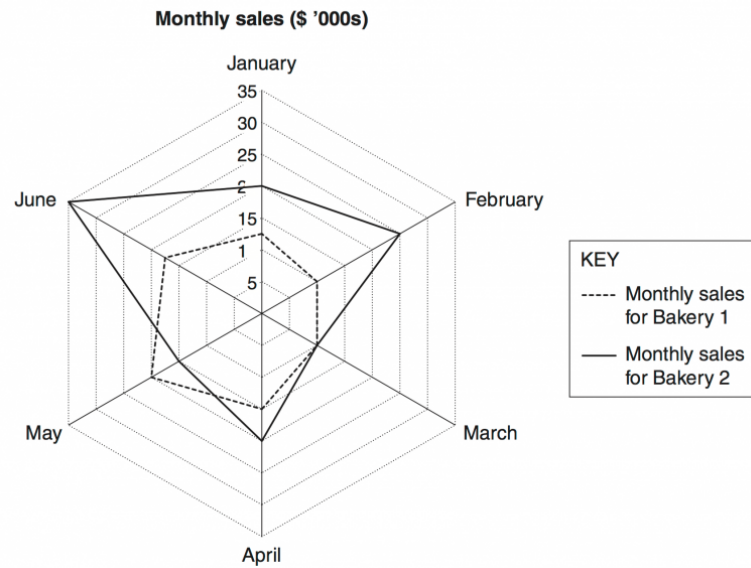
What is the median of these scores?

- (A) 7
  - (B) 8
  - (C) 77
  - (D) 78
-

6. Data, 2UG 2012 HSC 7 MC

⚡ RAP Data - Bottom 15%: School result (64%) was 3% above state average (61%)

The Pi Company has two bakeries. The radar chart displays the monthly sales for the bakeries.



What was the difference in sales in June between the two bakeries?

- (A) \$7.50
- (B) \$17.50
- (C) \$7500
- (D) \$17 500

7. 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

8. Data, 2UG 2008 HSC 3 MC

⚠ RAP Data - Bottom 9%: School result (54%) was 1% above state average (53%)

The stem-and-leaf plot represents the daily sales of soft drink from a vending machine.

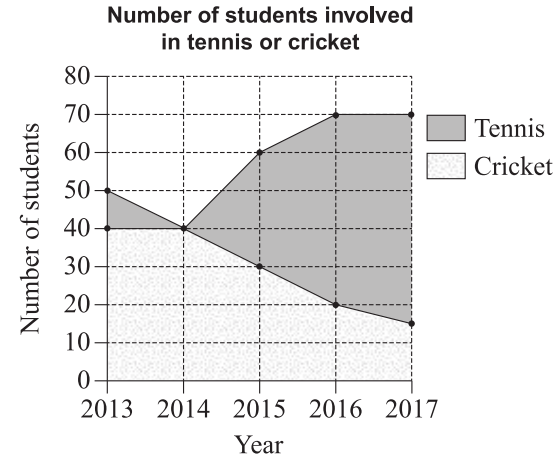
If the range of sales is 43, what is the value of  $N$  ?

2	$N$	5	5	
3	4	7	7	9
4	0	5	8	
5	2			
6	0	7		

- (A) 4
  - (B) 5
  - (C) 24
  - (D) 25
- 

9. Data, 2UG 2018 HSC 17 MC

The area chart shows the number of students involved in tennis or cricket at a school over a number of years.



In which year was the number of students involved in tennis equal to the number of students involved in cricket?

- A. 2013
  - B. 2014
  - C. 2015
  - D. 2016
-

10. Data, 2UG 2010 HSC 16 MC

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

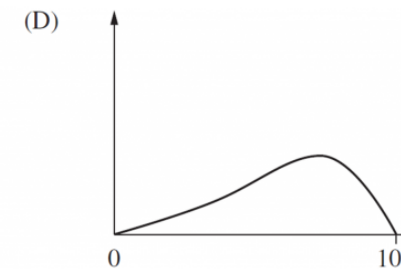
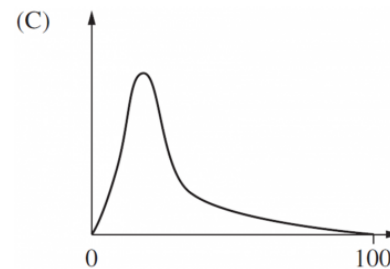
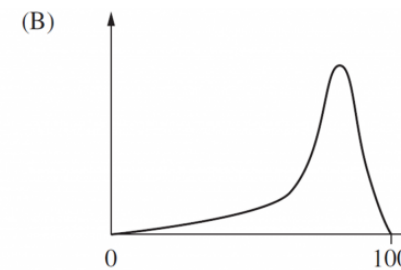
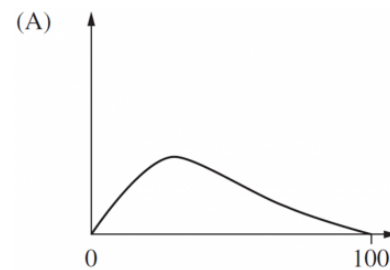
Boys					Girls			
		1	2	1	2	4		
		3	3	0	2	3	5	
	9 7	4	4	4	4	5	9	9
6	4 2	2	5	3				
	3 0	6	6	1	9			

What is the median test result for the class?

- (A) 44
- (B) 46
- (C) 48
- (D) 49

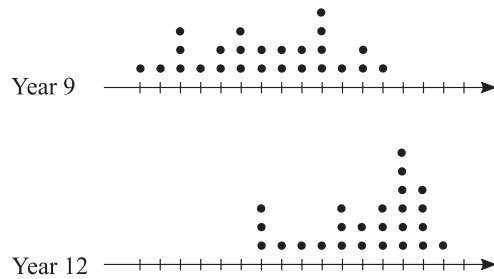
11. Data, 2UG 2006 HSC 8 MC

Which of these graphs best represents positively skewed data with the smaller standard deviation?



12. Data, STD2 SM-Bank 2 MC

The dot plots show the height of students in Year 9 and Year 12 in a school. They are drawn on the same scale.



Which statement about the change in heights when comparing Y9 to Y12 is correct?

- A. The mean increased and the standard deviation decreased.
- B. The mean decreased and the standard deviation decreased.
- C. The mean increased and the standard deviation increased.
- D. The mean decreased and the standard deviation increased.

13. FS Health, 2UG 2015 HSC 19 MC

The table shows the life expectancy (expected remaining years of life) for females at selected ages in the given periods of time.

**Life expectancy for females**

<i>Period of time</i>	<i>Remaining years of life</i>			
	at age 0	at age 25	at age 45	at age 65
1965–1967	74.2	51.2	32.3	15.7
1975–1977	76.6	53.1	34.0	17.1
1985–1987	79.2	55.4	36.1	18.6
1995–1997	81.3	57.1	37.7	19.8

In 1975, a 45-year-old female used the information in the table to calculate the age to which she was expected to live. Twenty years later she recalculated the age to which she was expected to live.

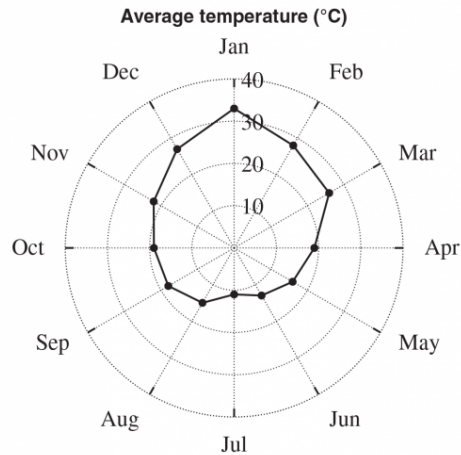
What is the difference between the two ages she calculated?

- (A) 2.7 years
- (B) 3.1 years
- (C) 3.7 years
- (D) 5.8 years

#### 14. Data, 2UG 2008 HSC 23a

You are organising an outside sporting event at Mathsville and have to decide which month has the best weather for your event. The average temperature must be between  $20^{\circ}\text{C}$  and  $30^{\circ}\text{C}$ , and average rainfall must be less than 80 mm.

The radar chart for Mathsville shows the average temperature for each month, and the table gives the average rainfall for each month.



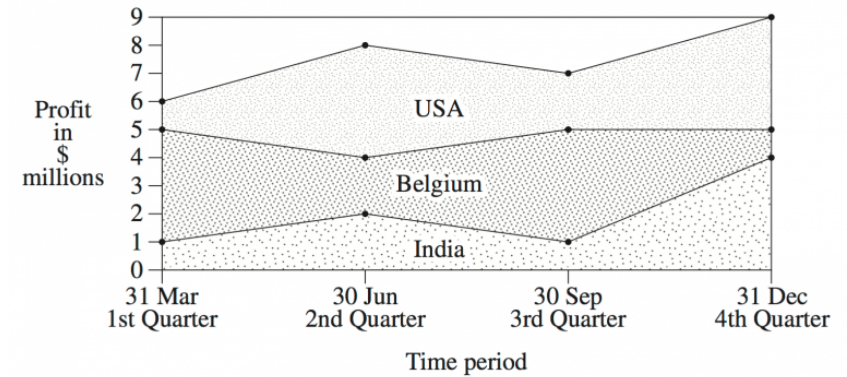
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
(mm)	150	162	86	95	104	140	59	81	60	72	70	90

- (i) If you consider only the temperature data, there are a number of possible months for holding the event. Name ONE of these months. (1 mark)
- (ii) If both rainfall and temperature data are considered, which month is the best month for the sporting event? (1 mark)

#### 15. Data, 2UG 2009 HSC 24b

- ⚡ Part i: RAP Data - Bottom 20%: School result (61%) was 6% above state average (55%)
- ⚠ Part ii: RAP Data - Bottom 9%: School result (76%) was 1% above state average (75%)

Tayvan is an international company that reports its profits in the USA, Belgium and India at the end of each quarter. The profits for 2008 are shown in the area chart.



- (i) What was the total profit for Tayvan on June 30? (1 mark)
- (ii) What was Tayvan's profit in Belgium on March 31? (1 mark)

#### 16. Data, 2UG 2009 HSC 24a

The diagram below shows a stem-and-leaf plot for 22 scores.

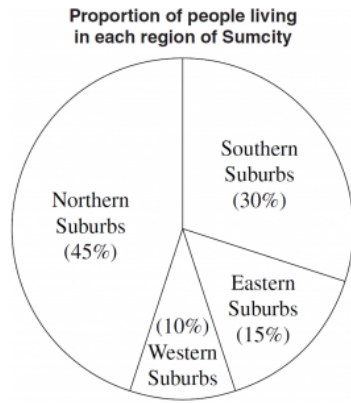
2	3	5	9		
3	1	4	7	9	
4	2	4	4	5	7
5	1	2	4		
6	2	3	7		
7	5	8	8	8	

- (i) What is the mode for this data? (1 mark)
- (ii) What is the median for this data? (1 mark)



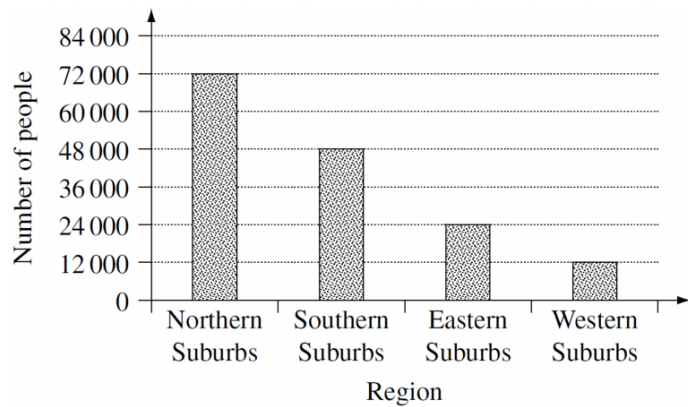
17. Data, 2UG 2005 HSC 24d

The sector graph shows the proportion of people, as a percentage, living in each region of Sumcity. There are 24 000 people living in the Eastern Suburbs.



(i) Show that the total number of people living in Sumcity is 160 000. (1 mark)

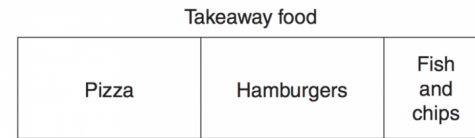
Jake used the information above to draw a column graph.



(ii) The column graph height is incorrect for one region. Identify this region and justify your answer. (2 marks)

18. Data, 2UG 2008 HSC 23e

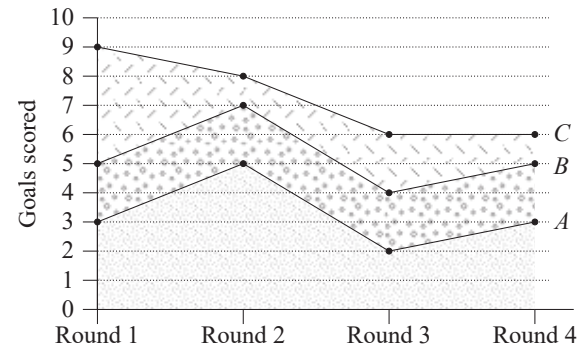
In a survey, 450 people were asked about their favourite takeaway food. The results are displayed in the bar graph.



How many people chose pizza as their favourite takeaway food? (2 marks)

19. Data, 2UG 2017 HSC 26f

The area chart shows the number of goals scored by three hockey teams, A, B and C, in the first 4 rounds.



(i) How many goals were scored by team C in round 1? (1 mark)  
 (ii) In which round did all three teams score the same number of goals? (1 mark)

20. Data, 2UG 2007 HSC 24d

Barry constructed a back-to-back stem-and-leaf plot to compare the ages of his students.

**Ages of students attending Barry's Ballroom Dancing Studio**

<i>Females</i>		<i>Males</i>
9	1	1 2 3
7	2	0 2 2 2 4 5
5	3	0 0 1 7
5 2	4	6 7
3 2 0	5	2
4 4 2 1	6	4 4

- (i) Write a brief statement that compares the distribution of the ages of males and females from this set of data. (1 mark)
  
- (ii) What is the mode of this set of data? (1 mark)
  
- (iii) Liam decided to use a grouped frequency distribution table to calculate the mean age of the students at Barry's Ballroom Dancing Studio.

For the age group 30 – 39 years, what is the value of the product of the class centre and the frequency? (2 marks)

- (iv) Liam correctly calculated the mean from the grouped frequency distribution table to be 39.5.

Caitlyn correctly used the original data in the back-to-back stem-and-leaf plot and calculated the mean to be 38.2.

What is the reason for the difference in the two answers? (1 mark)

---

21. 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.

<i>Male</i>		<i>Female</i>
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)
- 

22. Data, 2UG 2005 HSC 24a

- (i) Draw a stem-and-leaf plot for the following set of scores.

21 45 29 27 19 35 23 58 34 27 (2 marks)

- (ii) What is the median of the set of scores? (1 mark)
  - (iii) Comment on the skewness of the set of scores. (1 mark)
-

### 23. Data, 2UG 2013 HSC 26f

⚡ Part i: RAP Data - Bottom 17%: School result (77%) was 5% above state average (72%)

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 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

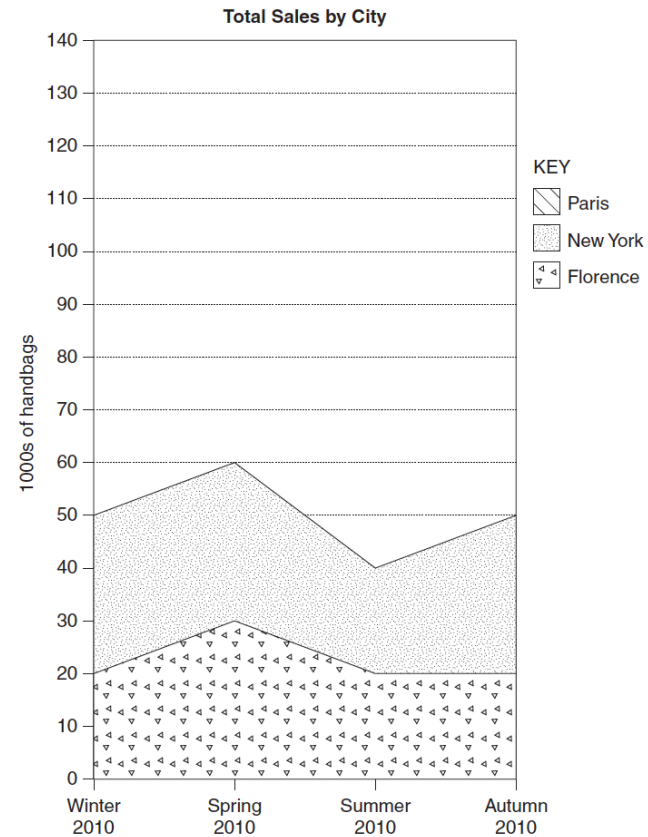
- What is the modal travel time when he uses roads without tolls? (1 mark)
- What is the median travel time when he uses roads without tolls? (1 mark)
- Describe how the two data sets differ in terms of the spread and skewness of their distributions. (2 marks)

### 24. Data, 2UG 2011 HSC 27a

A company sells handbags in Paris, New York and Florence.

Use the data in the table to complete the area chart below. (2 marks)

	Winter 2010	Spring 2010	Summer 2010	Autumn 2010
Number of handbags sold in Paris	60 000	70 000	50 000	70 000



## 25. Data, 2UG 2016 HSC 29c

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

<i>Women</i>	3	<i>Men</i>
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)

---

Copyright © 2004-18 The State of New South Wales (Board of Studies, Teaching and Educational Standards NSW)

Worked Solutions

## 1. Data, 2UG 2014 HSC 1 MC

$\Rightarrow D$

---

## 2. Data, 2UG 2010 HSC 1 MC

Range = High – Low

$$= 9 - 0$$

$$= 9$$

$\Rightarrow C$

---

## 3. Data, 2UG 2009 HSC 2 MC

Time in carpark = 3h 20m

From graph, charge will be \$18

$\Rightarrow C$

---

## 4. Data, 2UG 2004 HSC 8 MC

Centre angle of School X sector

$$= 100^\circ \text{ (by measurement)}$$

$\therefore$  Prizes won by school X

$$= \frac{100}{360} \times 116$$

$$= 32.22 \dots$$

$\Rightarrow B$

---

**5. Data, 2UG 2012 HSC 1 MC**

15 scores  $\Rightarrow$  Median is 8th

$\therefore$  Median is 78

$\Rightarrow D$

---

**6. Data, 2UG 2012 HSC 7 MC**

Bakery 1 Sales in June = \$17 500

Bakery 2 Sales in June = \$35 000

$\therefore$  Difference in Sales = 35 000 – 17 500  
= 17 500

$\Rightarrow D$

---

**7. Data, 2UG 2006 HSC 4 MC**

10 scores

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

$\Rightarrow C$

---

**8. Data, 2UG 2008 HSC 3 MC**

Range = High – Low = 43

$\therefore 67 - \text{Low} = 43$

Low = 24

$\therefore N = 4$

$\Rightarrow A$

---

**9. Data, 2UG 2018 HSC 17 MC**

Area Charts are cumulative,

Consider 2015,

Cricket players = 30

Tennis players = 60 – 30 = 30

$\Rightarrow C$

---

**10. Data, 2UG 2010 HSC 16 MC**

26 results given in the data

$\Rightarrow$  Median is average of 13<sup>th</sup> and 14<sup>th</sup>

♦♦ Mean mark 35%

$$\begin{aligned}\therefore \text{Median} &= \frac{45 + 47}{2} \\ &= 46\end{aligned}$$

$\Rightarrow B$

---

**11. Data, 2UG 2006 HSC 8 MC**

By elimination

Positive skew when the tail on the right side is longer.

$\therefore$  NOT  $B$  or  $D$

A smaller standard deviation occurs

when data is clustered more closely.

$\therefore$  NOT  $A$  where data is more widely spread.

$\Rightarrow C$

---

## 12. Data, STD2 SM-Bank 2 MC

Mean has increased (Y9 to Y12)

The Year 12 data is more tightly distributed around the mean.

∴ Standard deviation has decreased (Y9 to Y12)

⇒ A

## 13. FS Health, 2UG 2015 HSC 19 MC

In 1975, her life expectancy

$$= \text{age} + \text{remaining years}$$

$$= 45 + 34$$

$$= 79$$

◆ Mean mark 39%.

In 1995, her life expectancy

$$= 65 + 19.8$$

$$= 84.8$$

∴ Difference =  $84.8 - 79$

$$= 5.8 \text{ years}$$

⇒ D

## 14. Data, 2UG 2008 HSC 23a

(i) One of Feb, Mar, Nov, Dec

(ii) November

## 15. Data, 2UG 2009 HSC 24b

(i) \$8 000 000 (from graph)

(ii) Belgium profit (30 March)

$$= \$5\,000\,000 - \$1\,000\,000$$

$$= \$4\,000\,000$$

**MARKER'S COMMENT:** Area charts provide *cumulative totals*. Many students did not know this and incorrectly answered 14 million.

## 16. Data, 2UG 2009 HSC 24a

(i) Mode = 78

(ii) 22 scores

⇒ Median is the average of 11th and 12th scores

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

$$= 46$$

## 17. Data, 2UG 2005 HSC 24d

(i) Let the population of Sumcity =  $P$

$$15\% \times P = 24\,000$$

$$\therefore P = \frac{24\,000}{0.15}$$

$$= 160\,000 \dots \text{as required}$$

(ii) Western Suburbs population

$$= 10\% \times 160\,000$$

$$= 16\,000$$

The column graph has this population as

12 000 people which is incorrect.

### 18. Data, 2UG 2008 HSC 23e

Number of people who chose pizza

$$\begin{aligned} &= \frac{\text{Length of pizza section}}{\text{Total length of bar}} \times 450 \\ &\approx \frac{7}{18} \times 450 \\ &\approx 175 \end{aligned}$$

**COMMENT:** This question required measurement of the actual image on the exam. The same methodology works here.

$\therefore$  175 people chose pizza.

### 19. Data, 2UG 2017 HSC 26f

(i) Goals by team C =  $9 - 5$   
 $= 4$

(ii) In Round 3 (all teams scored 2 goals).

### 20. Data, 2UG 2007 HSC 24d

(i) More males attend than females and a higher proportion of those are younger males, with the distribution being positively skewed. Female attendees are generally older and have a negatively skewed distribution.

(ii) Mode = 64 (4 times)

(iii) Class centre =  $\frac{30 + 39}{2}$   
 $= 34.5$

Frequency = 5

$\therefore$  Class centre  $\times$  frequency  
 $= 34.5 \times 5$   
 $= 172.5$

(iv) The difference in the answers is due to the class centres used in group frequency tables distorting the mean value from the exact data.

### 21. 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) =  $20 - 9 = 11$

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

## 22. Data, 2UG 2005 HSC 24a

(i)

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

(ii) 10 scores

$$\begin{aligned}\therefore \text{Median} &= \frac{(5\text{th} + 6\text{th})}{2} \\ &= \frac{27 + 29}{2} \\ &= 28\end{aligned}$$

(iii) The data has a tail that stretches to the right

$\therefore$  Data is positively Skewed.

## 23. Data, 2UG 2013 HSC 26f

(i) Modal time = 52 minutes

(ii) 30 times with no tolls

$$\begin{aligned}\text{Median} &= \text{Average of 15th and 16th} \\ &= \frac{50 + 51}{2} \\ &= 50.5 \text{ minutes}\end{aligned}$$

(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.

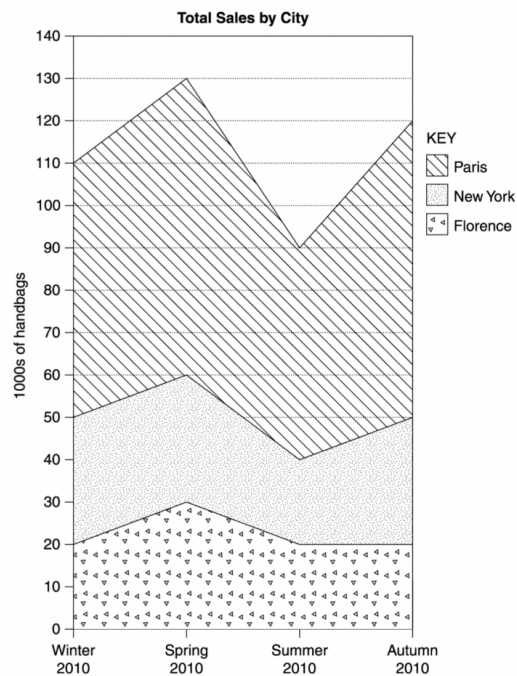
◆ Mean mark 36%

**MARKER'S COMMENT:** Finding a median proved challenging for many students. Take note!

◆ Mean mark 39%



## 24. Data, 2UG 2011 HSC 27a



♦♦ Mean mark 27%

COMMENT: Most common error was to not realise Area charts show *cumulative data*.

## 25. Data, 2UG 2016 HSC 29c

Women:

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

♦ Mean mark 44%.

Men:

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

∴ Pat is correct.