

Y1S2 XMQs and MS

(Total: 26 marks)

1. P31(AS)_2019 Q4 . 8 marks - Y1S1 Data collection
2. P31_2020 Q3 . 10 marks - Y1S2 Measures of location and spread
3. P31_2021 Q3 . 8 marks - Y1S2 Measures of location and spread

4. Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

- (a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	Frequency	Midpoint (x mm)
$0 \leq r < 0.5$	121	0.25
$0.5 \leq r < 1.0$	10	0.75
$1.0 \leq r < 5.0$	24	3.0
$5.0 \leq r < 10.0$	12	7.5
$10.0 \leq r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

- (b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

- (c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

- (d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

- (ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

- (iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

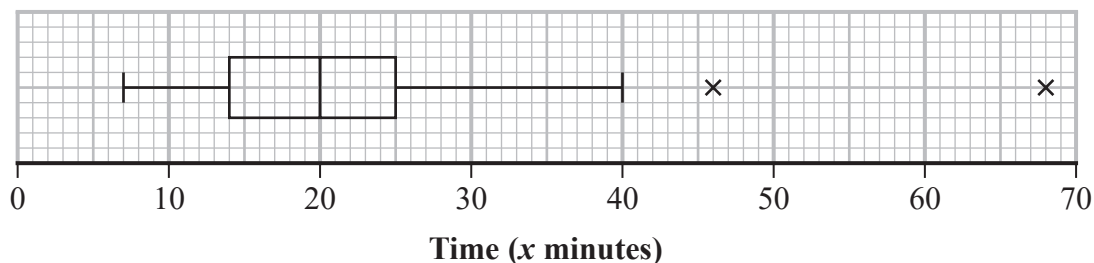


Question	Scheme	Marks	AOs
4 (a)	Tr(ace) (data needs to be converted to numbers before the calculation can be carried out)	B1	2.4
		(1)	
(b)	$[1+] \frac{138-131}{24} \times 4$	M1	2.1
	= 2.1666.... awrt 2.17	A1	1.1b
		(2)	
(c)	$\sigma = \sqrt{\frac{7704.1875}{184} - \left(\frac{539.75}{184}\right)^2} = 5.7676... \quad \sigma = \text{awrt } \underline{5.77}$	M1 A1	1.1b 1.1b
		(2)	
(d)(i)	Using class midpoints to estimate the mean assumes that the values are uniformly distributed within the class(es) .	B1	2.4
(ii)& (iii)	This is not the case here as the majority of the data (in the first class) are 0.	B1	2.3
	The actual mean is likely to be <u>smaller</u> than the estimate (since the first group has more values at 0 and close to 0)	dB1	2.2b
		(3)	
(8 marks)			
Notes			
(a)	B1: Identifying tr(ace) data Ignore comments about n/a, missing data, anomalies, etc.		
(b)	M1: Correct fraction $\frac{7}{24} \times 4$ allow working down $[5] - \frac{155-138}{24} \times 4$ allow a correct equation leading to a correct fraction e.g. $\frac{x-1}{5-1} = \frac{138-131}{155-131}$ for M1 Use of $(n + 1)$ with 138.75 allow $\frac{7.75}{24} \times 4$ A1: awrt 2.17 (condone $\frac{13}{6}$) awrt 2.29 from $(n + 1)$ (condone $\frac{55}{24}$)		
(c)	M1: Correct expression for standard deviation (allow mean = awrt 2.93) A1: awrt 5.77 correct answer only scores M1A1 (allow $s = 5.78$) SC: 5.76 with no working scores M1A0		
(d)(i)	B1: Explaining that data assumed to be spread evenly across each class (o.e.) e.g. The midpoint of each class is the <u>mean</u> of each class or all the values in the class are located at the midpoint condone normally distributed within each class		
Mark together (ii)&(iii)	B1: Demonstrating an understanding of the LDS that the majority of data values (in the first class) are at 0 or close to 0 (trace). dB1: (dependent upon 2 nd B1) Correct inference based on knowledge of the LDS SC: If B1 is scored in (i) for ‘The data are spread evenly across each class,’ then in (ii) ‘The data are not evenly distributed in the classes’ scores B1 but in (iii) ‘the actual mean is smaller’ with no further justification scores B0		

3. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



(a) Find the range of the times. (1)

(b) Find the interquartile range of the times. (1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

(c) calculate the mean time taken to complete the puzzle, (1)

(d) calculate the standard deviation of the times taken to complete the puzzle. (2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer. (1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.

When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b , explaining how your values satisfy the above conditions. (3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d). (1)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Qu 3	Scheme	Marks	AO
(a)	$[68 - 7 =]$ 61 (only)	B1 (1)	1.1b
(b)	$[25 - 14] =$ 11	B1 (1)	1.1b
(c)	$\left[\mu \text{ or } \bar{x} = \frac{607.5}{27} = \right] =$ 22.5	B1 (1)	1.1b
(d)	$\sigma = \sqrt{\frac{17\,623.25}{27} - "22.5"{}^2}$ <u>or</u> $\sqrt{146.4629\dots}$ $= 12.10218\dots$ awrt 12.1	M1 A1 (2)	1.1b 1.1b
(e)	$\mu + 3\sigma = "22.5" + 3 \times "12.1\dots" =$ awrt 59 so only one outlier	B1ft (1)	1.1b
(f)	Median increases implies that both values must be > 20 Mean is the same means that $a + b = 45$ So possible values are: e.g. $b = 21$ and $a = 24$ (o.e.)	M1 M1 A1 (3)	3.1b 1.1b 2.2b
(g)	Both values will be less than 1 standard deviation from the mean and so the standard deviation of all 29 values will be smaller	B1 (1)	2.4
		(10 marks)	
Notes			
(a)	B1 for correctly interpreting the box plot to find the range (more than 1 answer is B0)		
(b)	B1 for correct understanding of IQR and answer of 11		
(c)	B1 for 22.5 only (or exact equivalent such as $\frac{45}{2}$). Allow 22 mins and 30 secs.		
(d)	M1 for a correct expression including square root. Allow $\sqrt{146}$ or better. Ft their mean A1 for awrt 12.1 NB Allow use of $s = 12.3327\dots$ or awrt 12.3		
(e)	B1ft for a correct calculation or value based on their μ and σ and compatible conclusion		
(f)	1 st M1 Correct start to the problem and a correct statement about the values based on median Allow if their final two values are both > 20 2 nd M1 for a correct explanation leading to equation $a + b = 45$ (o.e. e.g. equidistant from mean) Allow if their final two values sum to 45 A1 for a correct pair of values (both > 20 with a sum of 45) and at least some attempt to explain how their values satisfy at least one of the conditions (both > 20 <u>or</u> $a + b = 45$). Ignore $a =$ or $b =$ labels NB The values for a and b do not need to be integers.		
(g)	B1 for a correct explanation. Must mention that both values are less than 1 sd (ft their answer to (d)) from the mean		

3. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x , using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \quad \sum y^2 = 5912$$

- (a) State the units of the variable x (1)
- (b) Find the mean Daily Mean Pressure for these 30 days. (2)
- (c) Find the standard deviation of Daily Mean Pressure for these 30 days. (3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

- (d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer. (2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA



Qu 3	Scheme	Marks	AO
(a)	Hectopascal <u>or</u> hPa	B1 (1)	1.2
(b)	$\bar{x} = \bar{y} + 1010$ <u>or</u> $\frac{214}{30} + 1010$ $= 1017.1333\dots$ awrt 1017	M1 A1 (2)	1.1b 1.1b
(c)	$\sigma_x = \sigma_y$ (or statement that standard deviation is not affected by this type of coding) $[\sigma_y =] \sqrt{\frac{5912}{30} - ("7.13[33\dots]")^2}$ <u>or</u> $\sqrt{146.1822\dots}$ $= 12.0905\dots$ awrt 12.1	M1 M1 A1 (3)	3.1b 1.1b 1.1b
(d)	High pressure (since approx. mean + sd) so clockwise Locations are (from North to South): Leuchars, Heathrow, Hurn Wind direction is direction wind blows <u>from</u> So: Heathrow (NE) Hurn (E) Leuchars (W)	B1 B1 (2)	2.4 2.2a
		(8 marks)	
Notes			
FYI	1 hPa = 100 Pa; 10hPa = 1 kPa; 1Pa = 1 Nm ⁻²		
(a)	B1 for “hectopascal” <u>or</u> hPa (condone pascals, allow millibars <u>or</u> mb) o.e. Do NOT allow kPa <u>or</u> kilopascals <u>or</u> Pa on its own		
(b)	M1 for a strategy to find \bar{x} Allow an attempt to find $\sum x$ that gets as far as $\sum x = \sum y - 30 \times 1010 [= 30\,514]$ A1 for awrt 1017 (accept 1020) [Ignore incorrect units]		
(c)	1 st M1 for an overall strategy using the fact $\sigma_x = \sigma_y$ (can be implied by correct <u>final</u> ans) <u>or</u> for $\sum x = 30\,514$ and $\sum x^2 = 31\,041\,192$ (both seen and correct) 2 nd M1 for a correct expression (with $\sqrt{\quad}$) (ft their \bar{y} to 3sf) allow awrt 146 for 146.1822.. <u>or</u> for correct expression in x can ft their $\sum x > 30\,000$ or their answer to (b) A1 (dep on 2 nd M1) for awrt 12.1 [Ignore incorrect units] Final ans of awrt 12.1 scores 3/3 but if they then adjust for x e.g. add 1010 (M0M1A1)		
Final answer	(d) 1 st B1 for at least one of these reasons (these 2 lines) clearly stated (may see diagram) Need “high pressure” and “clockwise” to score on 1 st line Contradictory statements B0 e.g. correct N~S list but say “anticlockwise” 2 nd B1 (indep of 1 st B1) for deducing the 3 correct directions either in the table or stated as above If the answers in table and text are different we take the table (as question says)		