GCSE
STATISTICS
8382/2H

Higher Tier Paper 2

Mark scheme

June 2022

Version: 0.1 Final

2 2 6 G 8 3 8 2 / 2 H / M S

Mark schemes are prepared by the Lead Assessment Writer and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation events which all associates participate in and is the scheme which was used by them in this examination. The standardisation process ensures that the mark scheme covers the students' responses to questions and that every associate understands and applies it in the same correct way. As preparation for standardisation each associate analyses a number of students' scripts. Alternative answers not already covered by the mark scheme are discussed and legislated for. If, after the standardisation process, associates encounter unusual answers which have not been raised they are required to refer these to the Lead Examiner.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of students' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Further copies of this mark scheme are available from aqa.org.uk

Copyright information

AQA retains the copyright on all its publications. However, registered schools/colleges for AQA are permitted to copy material from this booklet for their own internal use, with the following important exception: AQA cannot give permission to schools/colleges to photocopy any material that is acknowledged to a third party even for internal use within the centre.

Copyright © 2022 AQA and its licensors. All rights reserved.

Glossary for Mark Schemes

GCSE examinations are marked in such a way as to award positive achievement wherever possible. Thus, for GCSE Statistics papers, marks are awarded under various categories.

If a student uses a method which is not explicitly covered by the mark scheme the same principles of marking should be applied. Credit should be given to any valid methods. Examiners should seek advice from their senior examiner if in any doubt.

M	Method marks are awarded for a correct method which could lead to a correct answer.
A	Accuracy marks are awarded when following on from a correct method. It is not necessary to always see the method. This can be implied.
B	Marks awarded independent of method.
ft	Follow through marks. Marks awarded for correct working following a mistake in an earlier step.
SC	Special case. Marks awarded for a common misinterpretation which has some mathematical worth.
M dep	A method mark dependent on a previous method mark being awarded.
B dep	A mark that can only be awarded if a previous independent mark has been awarded.
oe	Or equivalent. Accept answers that are equivalent. eg accept 0.5 as well as $\frac{1}{2}$
[a, b]	Accept values between a and b inclusive.
[a, b)	Accept values $a \leq$ value $< b$
3.14 …	Accept answers which begin 3.14 eg 3.14, 3.142, 3.1416
Use of brackets	It is not necessary to see the bracketed work to award the marks.

Examiners should consistently apply the following principles

Diagrams

Diagrams that have working on them should be treated like normal responses. If a diagram has been written on but the correct response is within the answer space, the work within the answer space should be marked. Working on diagrams that contradicts work within the answer space is not to be considered as choice but as working, and is not, therefore, penalised.

Responses which appear to come from incorrect methods

Whenever there is doubt as to whether a student has used an incorrect method to obtain an answer, as a general principle, the benefit of doubt must be given to the student. In cases where there is no doubt that the answer has come from incorrect working then the student should be penalised.

Questions which ask students to show working

Instructions on marking will be given but usually marks are not awarded to students who show no working.

Questions which do not ask students to show working

As a general principle, a correct response is awarded full marks.

Misread or miscopy

Students often copy values from a question incorrectly. If the examiner thinks that the student has made a genuine misread, then only the accuracy marks (A or B marks), up to a maximum of 2 marks are penalised. The method marks can still be awarded.

Further work

Once the correct answer has been seen, further working may be ignored unless it goes on to contradict the correct answer.

Choice

When a choice of answers and/or methods is given, mark each attempt. If both methods are valid then M marks can be awarded but any incorrect answer or method would result in marks being lost.

Work not replaced

Erased or crossed out work that is still legible should be marked.

Work replaced

Erased or crossed out work that has been replaced is not awarded marks.

Premature approximation

Rounding off too early can lead to inaccuracy in the final answer. This should be penalised by 1 mark unless instructed otherwise.

Continental notation

Accept a comma used instead of a decimal point (for example, in measurements or currency), provided that it is clear to the examiner that the student intended it to be a decimal point.

Q	Answer	Marks	Comments
1	$\frac{1}{16}$	B1	

Q	Answer	Marks	Comments
2	Skewness	B1	

Q	Answer	Marks	Comments
3	61	B1	

Q	Answer	Marks	Comments
4	Reliability	B1	

Q	Answer	Marks	Comments
5(a)	Insufficient sample size or Gender (possibly) irrelevant or Unequal numbers of males and females or Data for one female is repeated (it seems)	B1	
	Additional Guidance		
	Only from some classmates		B1
	Hasn't included everyone in the class		B1

Q	Answer	Marks	Comments
5(b)	The value is (always) above the healthy range/healthy BMI	B1	oe must be in context at this grade
	The data is increasing (slightly) or The data is staying similar or There is no sign of an improvement or The y-axis does not start at zero, so this does not give the true reflection of the increase	B1	oe reference to trend or lack of reduction over time referencing that the y-axis does not start at zero requires clarification
	Additional Guidance		
	Context can be evidenced in either comment even if the comment is incorrect Comments should be about the whole graph or range of years (10 years minimum) and not just consecutive years		
	(1) Overall BMI is not in the healthy range (2) BMI is higher in 2010 than in 1995	B1 B1	
	(1) BMI is unhealthy between 1995 and 2010 (2) BMI is higher in 2015 than in 2010	B1 B0	
	(1) Average/overall BMI is unhealthy	B1	
	Do not allow just values being stated but condone correct comparison of values between 25.8 and 28		
	The BMI is always above 24.9	B1	
	The line is above 24.9	B0	

Q	Answer	Marks	Comments
6(a)	Most/more people are against HS2 (than in favour of it)	B1	oe hypothesis (not question)
	Additional Guidance		
	Most people will have negative opinions about HS2		B1
	People are against HS2		B1
	More older people are against HS2 than younger people		B1
	HS2 will be disliked (by locals)		B1
	The reason people oppose HS2 is because it affects the countryside		B1
	Many people are unhappy with HS2's plans		B1
	HS2 will affect the countryside		B0
	HS2 will affect housing		B0
	HS2 doesn't affect the environment		B0
	HS2 will affect house prices		B0
	I/Tom believe(s) most people are against HS2		B0
	HS2 will ruin the countryside. Most people will use HS2		B0
	The sacrifice of the countryside is worth less than HS2		B0

Q	Answer	Marks	Comments
6(b)	Any two from • Comment about omission of over 70s • Comment about 21 – 50 group's width / uneven group widths • Reference to no unit (years) given • Reference to 'prefer not to say' type responses	B2	oe both responses may be seen in one comment B1 any one correct response
	Additional Guidance		
	Condone irrelevant/incorrect responses with correct response(s) as long as not contradictory		
	21 – 50 is too large and it doesn't say years (all in one comment)	B2	
	It's too personal/sensitive	B1	
	Some may not want to give their age	B1	
	Doesn't state all possible ages	B1	
	21 – 50 is a big age group/gap	B1	
	There are big age gaps	B0	
	Not enough option boxes	B0	
	It's not relevant	B0	
	People could lie	B0	

Q	Answer	Marks	Comments
6(c)	No time frame is given or No option of 'prefer not to say' type response	B1	oe do not accept same reason in parts (b) and (c) do not accept reference to no option boxes
	Additional Guidance		
	Condone irrelevant/incorrect responses with correct response(s) as long as not contradictory		
	People may not earn anything		B1
	Some may not want to share		B1
	People may be paid in Euros		B0
	Some will lie		B0
	Doesn't say before or after tax		B0

Q	Answer	Marks	Comments
6(d)	**Alternative method 1**		
	Number the stations (0)1 to 29	B1	oe
	Obtain five (two-digit) random numbers from the internet or other source to obtain the stations, disregarding repeats	B1	oe eg obtain five different/unique numbers using random number generator
	Alternative method 2		
	Put all 29 station names in a hat	B1	oe
	Draw out five at random without replacement	B1	oe eg draw out five different/unique names
	Additional Guidance		
	Accept random name generator if just using names eg Type all 29 names into random name generator and obtain five names without repeats	B2	
	Number the stations	B0	
	Pick five using random number generator	B0	
	Put 29 stations/names in a hat	B1	
	Put all the names in a hat	B1	
	Put names in a hat	B0	

Q	Answer	Marks	Comments
6(e)(i)	Convenience	B1	accept Opportunity or Judgement
	Additional Guidance		
	Accept poor spellings		
	Opportunity and systematic on answer line	B0	

Q	Answer	Marks	Comments
6(e)(ii)	Will be asking rail travellers or quick/convenient/easy/cheap/efficient	B1	oe do not accept "convenient" here if "convenience" given in e(i)

Q	Answer	Marks	Comments
6(e)(iii)	Will not be asking (m)any non-rail travellers	B1	oe comment that suggest widening the sample frame
	Additional Guidance		
	May not be / is not representative	B1	
	More likely to support HS2	B1	
	Only on Saturday afternoon	B1	
	Need to go on different days / at different time	B1	
	Might all be from same train/group	B1	
	Biased as the arrival time could be a variable	B1	
	Biased	B0	
	Might all be male/female	B0	
	Not asked the whole population	B0	

Q	Answer	Marks	Comments
6(e)(iv)	Many people are affected in other places (without stations)	B1	oe
	Additional Guidance		
	So it's (more) representative of the population		B1
	To get more opinions		B1
	To broaden the data		B1
	To compare data (of those who have a station with those that don't)		B1
	Those that won't have a station will/may have a different opinion		B1
	(Those) people will/may have different opinions		B1
	They'd have an unbiased opinion		B0
	Need everyone's opinion		B0
	To avoid bias		B0

Q	Answer	Marks	Comments
6(f)(i)	Dual bar chart	B1	accept multiple bar chart

Q	Answer	Marks	Comments
6(f)(ii)	[124, 129] and [64, 69] or [124, 129] – 60 or [64, 69] + 60	M1	accept in hours (without units stated) please check the graph for workings
	Yes ticked, and correct subtraction of their values in range or Yes ticked and [124, 129] – 60, with correct answer, compared to [64, 69] or Yes ticked and [64, 69] + 60, with correct answer, compared to [124, 129]	A1	
	Additional Guidance		
	127 and 65 seen. Yes, 62 minutes is about an hour (subtraction implied)		M1A1
	127 and 65 seen. Yes, it is about an hour (answer to subtraction not seen)		M1A0
	127 – 65 = 62 (no decision)		M1A0
	It is 59 minutes which is about an hour so Li Na is correct (no evidence)		M0A0

Q	Answer	Marks	Comments
6(g)	**Alternative Method 1**		
	44	B1	may be seen as part of a calculation
	$\frac{\text{their } 44}{128} \times 100$ or 34.375	M1	oe their 44 must be <128
	34.4	A1ft	ft their 44 to 1dp
	Alternative Method 2		
	44	B1	
	$\left(1-\frac{84}{128}\right) \times 100$ or 34.375 or $100-\frac{84}{128} \times 100$ or 65.625 and 34.375	M1	oe
	34.4	A1	
	Additional Guidance		
	Accept 34 or 34.3 or 34.37 or 34.38 in place of 34.375 Accept 65.6 or 65.62 or 65.63 in place of 65.625		
	44 seen then answer 34		B1M1A0
	34.375 only (nothing else seen)		B0M1A0
	$\frac{44}{208} \times 100$		B1M0A0

Q	Answer	Marks	Comments
7(a)	[1978, 1980]	B1	must be a year (natural number)
	[250, 260] or [2500, 3200]	B1	accept monthly or annual total
	Additional Guidance		
	Do not allow any follow through from the year to the estimate		

Q	Answer	Marks	Comments
7(b)	Reference to cyclic nature of data, eg roughly every [10, 12] years there is a peak (trough) in the number of sunspots	B1	oe
	Additional Guidance		
	The number of sunspots has decreased over the years		B0
	There are peaks and troughs		B0
	The data follows a pattern of up and down variation		B0

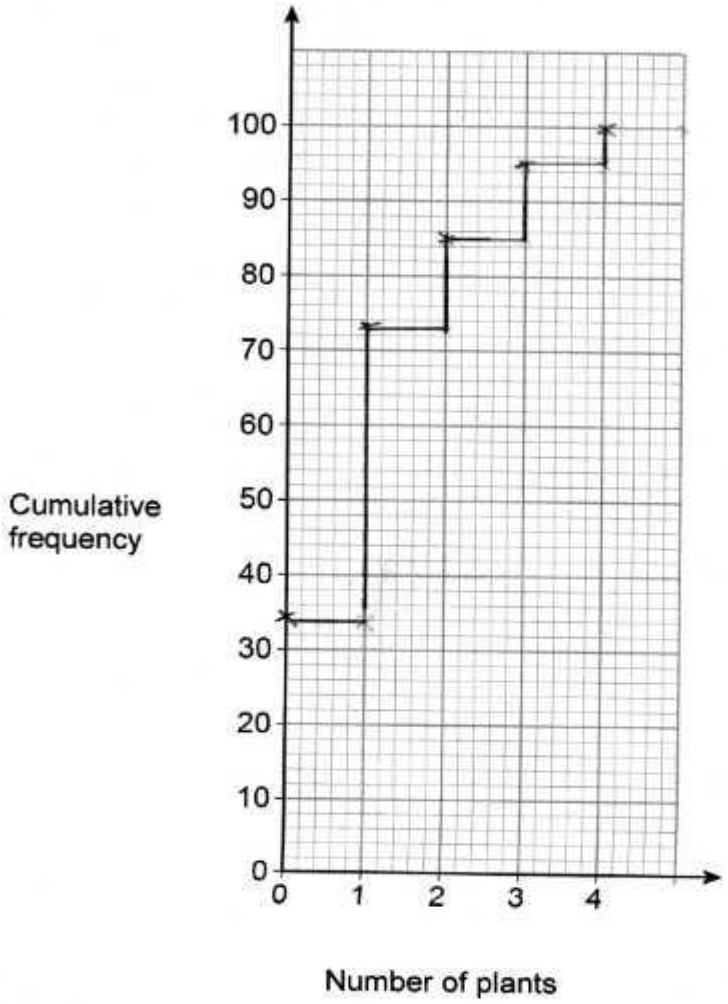
Q	Answer	Marks	Comments
7(c)	Calculate (or plot) 12-point moving averages	B2	B1 reference to moving averages but not 12-point accept 'rolling average' for 'moving average'

Q	Answer	Marks	Comments
8	54 in the D only area	B1	
	1 in the area outside the three circles	B1	
	0 in the central intersection of all three circles	B1	
	21 in BnDnL' and 21 in LnDnB'	B1	
	A total of 3 for the three numbers in the top three areas	B1ft	ft 100 – the sum of their five values
	Additional Guidance		
	Their 5 values must be integers (not negative) for the follow through mark		

Q	Answer	Marks	Comments
9(a)	**Alternative method 1**		
	$\frac{1}{5}$	M1	
	$\frac{1}{5} + \frac{4}{5} \times \frac{1}{4}$	M1dep	$\frac{1}{5} + \frac{1}{5}$ or $\frac{1}{5} \times 2$ unsupported is M1M0
	$\frac{2}{5}$	A1	oe accept 4/10 = 2/5 seen with little or no working
	Alternative method 2		
	Lists all the possible pairs for the last two players (either 10 unique pairs or 20 with either order)	M2	M1 lists at least 5 unique pairs or 10 with either order
	$\frac{2}{5}$	A1	oe

Q	Answer	Marks	Comments
9(b)(i)	This uses all the available data or This is a census	B1	oe
	Additional Guidance		
	The more data you use the more accurate it is		B0
	Gives most data		B0

Q	Answer	Marks	Comments
9(b)(ii)	Uses more recent data (as it will be more relevant) (for A, B or C) or Uses a reasonable sample size (for B or C)	B1	oe
	Additional Guidance		
	Do not allow reference to small sample size for A as is this is not an appropriate sample size		
	B or C – uses less data		B1
	A – uses less data		B0
	Condone any reference to 5, 20, 100 or all of the games instead of A, B, C or D as the option choice. If an option is not chosen, check the workings space.		

Q	Answer	Marks	Comments
10(a)(i)	Cumulative frequency values (34), 73, 85, 95, 100	B1	may be in table or evidenced in graph
	Correct plots at (0, 34), (1, 73), (2, 85), (3, 95) and (4, 100)	B1ft	ft their values if increasing $\pm\frac{1}{2}$ square tolerance
	Their points joined with straight lines to form the steps	B1ft	ft their values if increasing mark intention of a straight line
	Additional Guidance		
	Do not accept bars instead of, or as well as, the step polygon for the final B mark Ignore a step after the final plot at (4, 100)		
	If a bar chart is seen with no distinct plots then we can only award B1 if the CFs are correct		
			

Q	Answer	Marks	Comments
10(a)(ii)	The data are discrete	B1	oe

Q	Answer	Marks	Comments
10(b)	(Median Forest A) = 1	B1ft	ft cumulative graph
	(IDR Forest A) = 3	B1ft	ft cumulative graph
	The median is higher so there are more plants (on average) in forest B or There are more plants on average in forest B (as the median is higher)	B1ft	oe
	There is a larger/wider spread of the number of plants in Forest A (as the IDR is higher)	B1ft	oe
	Additional Guidance		
	If the median or the IDRs are the same, allow comments that the medians or IDRs are similar		
	If the median and/or IDR is correct in the answer space, ignore any contradiction on the graph		
	Comparison comments cannot be awarded if there are no median or IDR scores calculated		
	Answers should include an interpretation of the median / IDR in context. Plants must be seen in either response.		
	They cannot compare the range/IQR instead of the IDR or the mean instead of the median.		
	Accept reference to units		
	Ignore irrelevant comments as long as not contradictory. eg There are more plants in Forest B as the median and IDR are higher	B1	
	The number of plants in Forest A is less consistent/less varied	B1	
	There are more plants in Forest B	B0	
	There is a larger median of plants in Forest B	B0	
	There is a larger range of plants in Forest A	B0	
	The average is higher in Forest B (no context)	B0	
	The spread is larger in Forest A (no context)	B0	

Q	Answer	Marks	Comments
10(c)	(3) : Too long a gap between release and re-capture	B1	oe
	(4) : Too few voles in the re-captured group	B1	oe SC1 two correct responses but for the wrong steps
	Additional Guidance		
	Accept reference to tags rather than dye		
	Accept a valid suggestion for Step 3:		
	It is 6 months later so there may be births/may have died/hibernated or moved on	B1	
	The voles may have moulted	B1	
	The dye may have washed off	B1	
	6 months is a long time	B0	
	Accept a valid suggestion for Step 4:		
	She should collect more voles for a better estimate	B1	
	May not re-capture any voles with dye due to the very small sample	B1	
	They should have returned to the same part of the river/forest	B1	
	She should have collected 30 (any number greater than 5 implies more)	B1	
	Only caught 5	B1	
	She should have collected the same amount (not implied more)	B0	

Q	Answer	Marks	Comments
11(a)	4	B1	

Q	Answer	Marks	Comments
11(b)	(The birth date minus due date data) has to be a whole number of days	B1	oe eg all plots should be on vertical lines, but this plot is between the vertical lines
	Additional Guidance		
	Cannot have the 4.5^{th} of a month		B1
	It can't be 15.5 days, it's either 15 or 16 days		B1
	It can't be 15.5 days		B1
	A date can't be a decimal		B1
	It is plotted between 2 days		B1
	It isn't on a line on the x-axis		B1
	It isn't on a line		B0

Q	Answer	Marks	Comments
11(c)(i)	(4.01) is the expected mass/weight in kg of a baby born on its due date	B1	oe 4010 and reference to g
	Additional Guidance		
	Units of mass must be seen		
	Condone reference to a new-born to mean born on its due date		
	It is the initial/starting/beginning mass/weight in kg of the baby		B1
	It is the initial/starting/beginning mass/weight of the baby		B0

Q	Answer	Marks	Comments
11(c)(ii)	(0.04) is the increase in the baby's mass in kg for every additional day	B1	oe 40 g
	Additional Guidance		
	Units of mass must be seen		
	It is how much the baby's mass/weight increases in kg per day		B1
	How much the mass/weight changes per day		B0

Q	Answer	Marks	Comments
11(c)(iii)	Line of best fit that is • from x values of –49 to 10 • straight • passes through (–40, [2.3,2.5]) and (0, 4.01) and (10, [4.3,4.5])	B2	B1 a straight line with a positive gradient that passes through (0, 4.01) strict $\pm\frac{1}{2}$ square tolerance for plotting (0, 4.01)
	Additional Guidance		
	For the line of best fit, mark intention to be straight		

Q	Answer	Marks	Comments
11(d)	Sam can be estimated as it is interpolation	B1	oe accept yes/valid with justification e.g. It is within the range of the data
	their value from their line of best fit or 3.4(1) kg	B1ft	oe must be straight line if using their line of best fit
	Nim should not be estimated as her data are outside the range (of the scatter graph) / it would be extrapolation	B1	oe accept no/invalid with justification condone saying cannot be estimated condone the mention that it can be estimated but it would be unreliable due to extrapolation oe do not accept 'it goes off the graph'
	Additional Guidance		
	Check the graph for workings and an estimate		
	For the interpolation comment, allow mention of reliability for yes		
	Do not penalise if estimates given as long as unreliability/extrapolation referenced		
	Yes, interpolation	B1	
	Yes, it is within the data	B1	
	Interpolation	B0	
	Extrapolation comment, allow mention of unreliability for no		
	No, the trend/pattern may not continue	B1	
	The trend/pattern may not continue	B0	
	There is no data at that point	B0	
	The line (of best fit) does not go that far	B0	

Q	Answer	Marks	Comments
12(a)	**Alternative method 1**		
	800 × 50 or 40 000	M1	
	0.96 × 800 or 768	M1	oe
	their 768 × 300 or 230 400	M1dep	oe dep on 2nd method mark
	their 230 400 – their 40 000	M1dep	oe dep on 3rd method mark
	190 400	A1	
	Alternative method 2		
	0.96 × 800 or 768 or 0.04 × 800 or 32	M1	oe
	their 768 × 250 or 192 000	M1dep	oe dep on 1st method mark
	their (800 – their 768) × 50 or 1600	M1dep	oe dep on 1st method mark
	their 192 000 – their 1600	M1dep	oe dep on all previous method marks
	190 400	A1	
	Alternative method 3		
	0.96 x 250 or 240	M1	oe
	0.04 x 50 or 2	M1	oe
	their (240 – 2) or 238	M1dep	oe dep on M1M1
	their (240 – 2) or 238 x 800	M1dep	oe dep on 3rd method mark
	190 400	A1	
	Alternative method 4		
	0.04 x 300 or 12	M1	oe
	their (250 – 12) or 238	M2dep	oe eg 300 - 62
	their 238 x 800	M1dep	oe dep on M3
	190 400		
	Additional Guidance		
	There may be an attempt at more than one alternative method. Award the highest mark(s)		

Q	Answer	Marks	Comments
12(b)(i)	It is poor practice to take 5 in a row or The sample needs to be spread out more	B1	oe
	Additional Guidance		
	This is more about 5 in a row and not a poor sample size		
	Ignore any reference to other sampling methods		
	Not effective as 5 were chosen one after another		B1
	Poor sample size		B0
	Small sample so not representative/reliable		B0

Q	Answer	Marks	Comments
12(b)(ii)	$0.96^4 \times 0.04$ or $0.0339(7...)$	M1	oe
	$5 \times 0.96^4 \times 0.04$ or $5 \times 0.0339(7...)$	M1dep	
	0.17 or better	A1	oe 0.169(869…)
	Additional Guidance		
	If 0.17 or 0.169(869…) seen with no incorrect working scores full marks		

Q	Answer	Marks	Comments
13(a)	Positive correlation between the marks scored on the two papers	B1	oe eg positive agreement/ relationship/association between the rankings of the marks. eg students who did well on one paper, tended to do well on the other
	Additional Guidance		
	Ignore any adjectives describing the strength of the relationship		
	Some context should be included, eg reference to marks or papers		
	They do well on both tests		B0
	They do well		B0
	Positive correlation		B0
	The papers are closely related		B0

Q	Answer	Marks	Comments
13(b)(i)	Will be nearer to 1 or increases	B1	oe

Q	Answer	Marks	Comments
13(b)(ii)	1 - $\frac{6(\sum)d^2}{5(25-1)}$ = 0.8	M1	oe for forming a correct equation accept any variable for *d* or $(\sum)d^2$
	$(\sum)d^2$ = 4	A1	accept any variable for *d* or $(\sum)d^2$
	(1 -) $\frac{6 \times \text{their}\,4}{6(36-1)}$	M1dep	condone $\frac{6\times(\text{their}\,d)^2}{6(36-1)}$ must be clear what their $\sum d^2$ or *d* is
	[0.88, 0.9] or $\frac{31}{35}$	A1ft	oe ft for their $\sum d^2$ (must be ≥ 0) provided 0 < SRCC < 1
	Additional Guidance		
	For full marks, condone $1-\frac{6\times 2^2}{6(36-1)} =$ [0.88, 0.9] or $\frac{31}{35}$		
	If (1 -) is seen then it must be in the correct place in the formula for both M marks eg $\frac{1-6\sum d^2}{5(25-1)}(=0.8)$ scores M0M0 unless recovered		

Q	Answer	Marks	Comments
14(a)	Ticks cannot tell and explains that the data has been ordered so it is not clear whether the values of 0 are for consecutive days	B1	
	Additional Guidance		
	If the box is not ticked check the working space		
	Cannot tell ticked and it snowed on 3 consecutive days, but the data has been ordered by size, not chronologically so we can't tell	B1	
	Cannot tell ticked and the data is organised by size and not time	B1	
	(If they don't give the dates) we don't know if the first zero is linked to the second zero	B1	
	Cannot tell ticked and the data values of zero could have been split up over winter so not consecutive	B1	

Q	Answer	Marks	Comments
14(b)	(Median =) 2	B1	
	(Mean =) 5.1	B1	
	$1^2 + 1^2 + 3^2 + ...$ or $\sum x^2 = 927$	M1	workings may not be seen if calculator used
	(standard deviation =) 8.166…	A1	8.2 or better
	Substitution of their values into the skew formula	M1dep	dep on M1
	1.138…(= 1.14) or 1.139 (= 1.14)	A1	
	Additional Guidance		
	Do not allow any values substituted into the skew formula unless correct values/workings seen		
	Check the list of data values for evidence of workings to find the mean/median. Allow any indication of the median eg crossings out/circling		
	If candidates work backwards from +1.14, the maximum score possible is for correctly calculating the median and/or mean		
	For the 2nd A1, they must show at least 3 decimal places		

Q	Answer	Marks	Comments
14(c)	C	B1	