GCSE
STATISTICS
8382/1H

Higher Tier Paper 1

Mark scheme

June 2020

Version: 1.0 Final

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

Level of response marking instructions

Level of response mark schemes are broken down into levels, each of which has a descriptor. The descriptor for the level shows the average performance for the level. There are marks in each level.

Before you apply the mark scheme to a student's answer read through the answer and annotate it (as instructed) to show the qualities that are being looked for. You can then apply the mark scheme.

Step 1 Determine a level

Start at the lowest level of the mark scheme and use it as a ladder to see whether the answer meets the descriptor for that level. The descriptor for the level indicates the different qualities that might be seen in the student's answer for that level. If it meets the lowest level then go to the next one and decide if it meets this level, and so on, until you have a match between the level descriptor and the answer. With practice and familiarity you will find that for better answers you will be able to quickly skip through the lower levels of the mark scheme.

When assigning a level you should look at the overall quality of the answer and not look to pick holes in small and specific parts of the answer where the student has not performed quite as well as the rest. If the answer covers different aspects of different levels of the mark scheme you should use a best fit approach for defining the level and then use the variability of the response to help decide the mark within the level, ie if the response is predominantly level 3 with a small amount of level 4 material it would be placed in level 3 but be awarded a mark near the top of the level because of the level 4 content.

Step 2 Determine a mark

Once you have assigned a level you need to decide on the mark. The descriptors on how to allocate marks can help with this. The exemplar materials used during standardisation will help. There will be an answer in the standardising materials which will correspond with each level of the mark scheme. This answer will have been awarded a mark by the Lead Examiner. You can compare the student's answer with the example to determine if it is the same standard, better or worse than the example. You can then use this to allocate a mark for the answer based on the Lead Examiner's mark on the example.

You may well need to read back through the answer as you apply the mark scheme to clarify points and assure yourself that the level and the mark are appropriate.

Indicative content in the mark scheme is provided as a guide for examiners. It is not intended to be exhaustive and you must credit other valid points. Students do not have to cover all of the points mentioned in the Indicative content to reach the highest level of the mark scheme.

An answer which contains nothing of relevance to the question must be awarded no marks.

1	2017	B1	

2	140	B1	

3	12	B1	

4	D	B1	

	A question regarding how far a person usually drives that (1) includes a time period (likely to be per week) and a unit for distance travelled (2) has at least 3 numerical response options that are exhaustive (3) has at least 3 numerical response options that are mutually exclusive	B3	B2 for a question involving how far a person usually drives that satisfies two of these three conditions B1 for a question that satisfies one of these three conditions

Additional Guidance

5(a)	**Condition 1**: The time period and unit for distance must be given in the question or be inferred from the response options. Condone other time periods (such as per day, per month etc). eg What distance, in miles, do you (usually) drive each week? **Condition 2**: The response options must cover all possible options. Allow distances to be rounded to the nearest unit / tenth etc. Allow 0 to be covered by 'I do not drive' oe eg 0–19.9 km 20–39.9 km 40–59.9 km 60 km or more eg Under 50 miles 50 miles $\leq x \leq$ 100 miles Over 100 miles Condone overlapping response options when considering whether all values are included. **Condition 3**: There should be no overlapping response options. Consider an option such as 'I do not drive' to overlap with an interval including 0. The response options do not need to be exhaustive for this condition to be met. Ignore any option boxes for 'Other' or 'Don't know' etc when considering these conditions. Example 1: How far do you drive per day? 1–15 16–30 Other (Condition 1 not satisfied as there is no unit for distance Condition 2 is not satisfied as the intervals are not exhaustive and because there are just 2 numerical options – do not allow an 'other' option to cater for missing values. Condition 3 is not satisfied as there are just 2 numerical options) Example 2: What distance do you usually drive each week? $\leq$ 50 km 50–100 km $>$ 100	B0 B2

	(Condition 1 is satisfied – condone missing unit on 3rd option as a unit is seen elsewhere Condition 2 is satisfied – condone overlap when considering whether intervals are exhaustive Condition 3 is not satisfied as 50 is included in two intervals.) If intervals are poorly expressed, award a maximum of 2 marks, eg How many miles do you usually drive every day? < 50 ☐ ≥ 50 < 100 ☐ ≥ 100 < 200 ☐ ≥ 200 ☐ Other ☐	B2

5(b)(i)	It is a sensitive topic / may not want to say **or** To ensure he gets truthful responses **or** To improve the response rate to the question	B1	oe
	Additional Guidance		
	So that his results are unbiased/accurate unless further information is provided explaining why his results will be better Results are anonymous Some may lie		B0 B0 B0
5(b)(ii)	Some of those answering ‘Yes’ did not answer his question (about driving faster than the speed limit) About 50 /some said yes because they threw an odd number Some were told to say yes (implying dice roll)	B1	oe
	Additional Guidance		
	About 50 of these people would have said ‘Yes’ because they rolled an odd number Only (about) 10 people answered ‘Yes’ to his question The correct proportion is 20%		B1 B1 B1

6(a)	2 + 1 + 5 + 2 + 10 + 8 + 25 + 13 or 66	M1	oe
	66, this is about two-thirds	A1	Any indication
	Additional Guidance		
	Condone one error or omission in the addition for M1		
	If calculated two thirds must equal 0.67 or 0.66 or better or 67% or 66% or better, use of two thirds = 0.6 cannot score the A mark		
	$\frac{66}{100}$ is about $\frac{2}{3}$		M1A1
	66% = two thirds		M1A0
	Working with 101: 66 is two thirds of 101 66 and two thirds of 101 is 67 or 67.3(…) 66 is 65.3% of 101 so they are about the same		 M1A0 M1A0 M1A0
	Working with 2000: 66% of 2000 is 1320, two thirds of 2000 is 1333 or better or 1334, so they are about the same		M1A1
	Any reference to 66 being 66 adults is A0 eg 66 adults chose to work earlier		M1A0

6(b)	Ticks 'Cannot Tell' and Due to rounding (there could be a few who chose 11.30 but out of 2000 people this is almost zero %) or Some of the people put 'Don't know' (some of them may want to start at 11.30)	B1	oe
	Additional Guidance		
	A few needs to be less than 10		
	Reference to rounding, eg: It could be due to rounding It could be 0.49% It could be due to rounding, it could be 0.9%		 B1 B1 B0
	Some may have answered, but not enough for it to become 1 percent		B0
	Ticks 'Cannot Tell', it may have been a really small percentage		B0

6(c)	Not all British working adults work an 8-hour day / have fixed hours	B1	oe
	Additional Guidance		
	Any mention of shift work / working nights		B1
	Some people work flexible hours		B1
	People have different work commitments People have different commitments		B1 B0
	Some people may be part-time		B0
	Some people are self employed		B0
	Reference to sample size, asking more people etc		B0
	Reference to representation, eg other workers may work differently		B0

7(a)(i)	$\frac{332}{600}$ or $\frac{83}{150}$ or 0.55 or better or 55% or better	B2	oe B1 sight of 332 or $\frac{n}{600}$; $n < 600$
	Additional Guidance		
	Ignore any attempt to convert or simplify once the correct answer is seen		
	For B2, ignore probability words unless contradictory and on the answer line		

7(a)(ii)	$\frac{529}{600}$ or 0.88 or better or 88% or better	B2	oe B1 $\frac{71}{600}$ or 0.12 or 0.118 or better or 12% or 11.8% or better or sight of 529
	Additional Guidance		
	Ignore any attempt to convert or simplify once the correct answer is seen		
	For B2, ignore probability words unless on the answer line and contradictory		

7(b)	$\frac{11}{71}$ or 0.15 or better or 15% or better	B2	oe B1 sight of 71 or 11 as numerator in a probability
	Additional Guidance		
	Ignore any attempt to convert or simplify once the correct answer is seen		
	For B2, ignore probability words unless on the answer line and contradictory		

7(c)	$\frac{67}{200}$ or 0.335 or 33.5%	M1	oe
	$\frac{67}{200} \times \frac{66}{199}$ or $\frac{4422}{39800}$ or 0.11 or 0.1111… or 11.11(%) or 11.1105…(%) or 11.1106(%)	M1dep	oe
	0.111 or 11.111%	A1	SC1 for $\frac{4489}{40000}$ or 0.112 or 11.223%
	Additional Guidance		
	Ignore any attempt to convert or simplify once the correct answer is seen		
	For A1, ignore probability words unless on the answer line and contradictory		

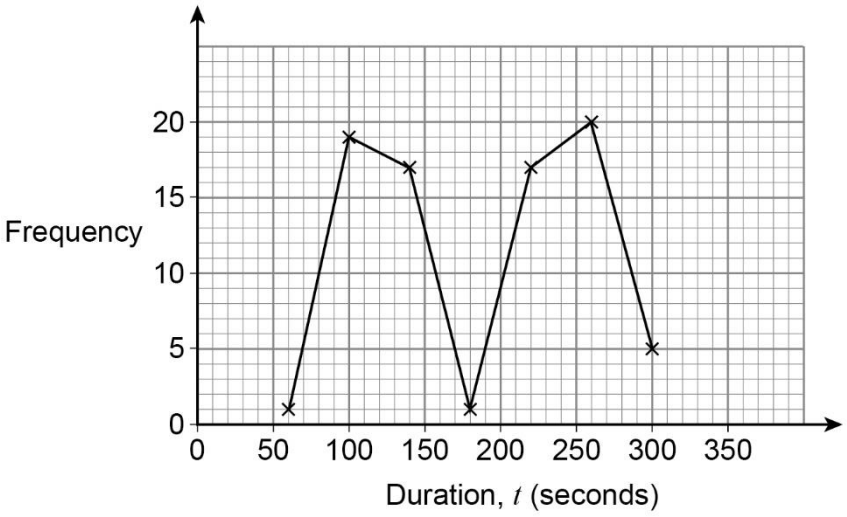
7(d)	Statement 1: Ticks Yes and comments that over 300 (332) went to social media first that day	B1	oe eg 'over half'
	Statement 2: Ticks Cannot tell and comments that these results are just for one day (and might not be true for every day) or Ticks No and comments that fewer than 100 (88) went on social media first that day (if it's not true on the first day it cannot be true every day)	B1	oe
	Additional Guidance		
	For the first statement, do not award B1 if 332 or its calculation is wrong		

8(a)	A comment relating to sample size / accuracy: eg 6 people is not enough eg The results will not be reliable enough with just 6 people	B1	oe
	A comment relating to ethics: eg The researcher cannot infect randomly chosen people with a deadly disease eg The people taking part in the experiment may die	B1	oe
	Additional Guidance		
	No placebo (people may have recovered without the drug) No control group		First B1 First B1
	It has to be voluntary (all 6 could have volunteered)		First B0
	Use people who already have the disease The disease could be infectious (and so people might spread it)		Second B1 Second B1
	The people could be unhealthy The people could be really old and die anyway		Second B0 Second B0
	Any reference to problems with the drug is second B0, eg: The drug might not be suitable They might be allergic to the drug The drug might be dangerous The drug might have long term effects		 Second B0 Second B0 Second B0 Second B0

8(b)	Patients should be anonymous	B1	oe eg She shouldn't include the name of the patients
	Additional Guidance		
	It's confidential		B1
	The names (are given)		B1
	It's too personal		B1
	It might be hurtful as their names have been published It might be hurtful for those people to read it		B1 B0
	Consent is needed / Some people might not want to be included (missed the point, publishing names should be avoided)		B0
	It's rude/offensive		B0

9	$\left(\frac{1}{2}\right)^3$ **or** 120 ÷ 8 **or** a list of the 8 possible outcomes **or** a tree diagram with H and T on each branch	M1	oe
	15	A1	

10(a)



Points plotted at correct heights	M1	($\pm \frac{1}{2}$ small square) Condone one error Condone bars
Points plotted at midpoints of intervals	M1	($\pm \frac{1}{2}$ small square) Condone one error
Fully correct frequency polygon	A1	Ignore any attempt to draw lines down to horizontal axis.

Additional Guidance

If polygon and histogram are both drawn and the midpoints are marked correctly (maximum of 2 marks) **or** a histogram only drawn then the heights must be correct ($\pm \frac{1}{2}$ small square)	M1 M1 A0 M1 M0 A0

10(b)	187 (seconds)	B1	

10(c)

The mean falls in an interval with frequency of 1	B1	oe

Additional Guidance

Ignore any reference to the mode/median/grouped data	
The durations are either short or long (with not many in the middle)	B1
The distribution is bi-modal	B1
Not many values are close to 187 **or** a large drop around 187	B1

11(a)	Bar for females aged 10–19 drawn at 190	B2	$\frac{1}{2}$ square tolerance B1 for sight of 160 (thousand) **or** 190 (thousand)

11(b)(i)	400 (thousand) **or** 540 (thousand)	M1	Implied by 940 (thousand). Allow ±10 (thousand)
	$\frac{\text{their } 400 + \text{their } 540}{1400 + 1550}$ $(\times 100)$	M1dep	
	[31,32.6] (%) or 33 (%), if correct working is seen.	A1	
	Additional Guidance		
	The mark for M1 may be on the diagram		

11(b)(ii)	(It is the age when…) people (often) leave their parent's home **or** people move to start a new job / leave university **or** people start a family and need to move / moving in with partner **or** People buying/renting their first home	B1	oe
	Additional Guidance		
	They are getting married		B0
	They have enough money for a mortgage		B0
	They have a new job		B0

12(a)	$\frac{8\ 290\ 000}{15\ 590\ 000}$ **or** [0.53, 0.532] **or** $\frac{2\ 750\ 000}{50\ 450\ 000}$ **or** [0.054, 0.055]	M1	oe
	$\frac{8\ 290\ 000}{15\ 590\ 000}$ **or** [0.53, 0.532] **and** $\frac{2\ 750\ 000}{50\ 450\ 000}$ **or** [0.054, 0.055]	M1 dep	oe
	$\frac{8\ 290\ 000}{15\ 590\ 000} \div \frac{2\ 750\ 000}{50\ 450\ 000} = [9.7, 9.8]$ **or** $\frac{2\ 750\ 000}{50\ 450\ 000} \div \frac{8\ 290\ 000}{15\ 590\ 000} = 0.1(\ldots)$ **or** [0.53, 0.532] = 10 x [0.054, 0.055] **and** a correct conclusion	A1	Acceptable conclusions include Mike's statement is correct **or** The risk of hearing loss for those aged 60 and over is just less than 10 times greater oe
	Additional Guidance		
	$\frac{8\ 290\ 000}{15\ 590\ 000} \div \frac{2\ 750\ 000}{50\ 450\ 000} = [9.7, 9.8]$ He is correct		M1 M1 A1
	Mike is correct as 0.53 ÷ 10 is about 0.055		M1 M1 A1

12(b)	[1700 000, 1732 222] integer values only	B1	oe Accept 2 000 000 with working.

13(a)	16%	B1	oe

13(b)	**Sophie is correct**, with working $500 + 4 \times 3 = 512$ **and** $500 - 4 \times 3 = 488$ **or** $\frac{512-500}{4} = 3$ **and** $\frac{488-500}{4} = -3$ **or** $\frac{512-500}{3} = 4$ **and** $\frac{500-488}{3} = 4$ **or** $\frac{512-488}{6} = 4$	B2	B1 for one calculation **or** B1 for a link made between 3 sds and 'almost all' or 99(.7…)% or 99.8% or almost 100%
	Additional Guidance		
	Stating that reducing the standard deviation to 4 (ml) would work, with correct working		B2

14 (a)	5005 (million)	B1	
	Additional Guidance		
	Mark the table first but If the table is blank check the answer space		

14 (b)	Increasing trend (in the value of clothing imports) Moving averages increase	B1	oe The value of clothing imports is increasing
	Additional Guidance		
	Reference to correlation		B0
	Moving averages show a constant increase		B0

14(c)	A suitable comment about the seasonal variation, eg (Clothing imports are) highest in Q3 / higher at the end of a year (Imports are) lowest in Q2 / lower at the start of a year	B1	oe
	Additional Guidance		
	Imports are above the trend line in Q3		B1
	Increase in Q3 / decrease in Q2		B1
	Most imports in the summer/autumn		B1
	In Q2 the imports are less expensive (this is not referring to the amount of imports)		B0

14(d)(i)	[5140, 5160]	B1	
	$\frac{-40 + (-150)}{2}$ **or** -95	M1	oe
	[5045, 5065]	A1	Unless their calculation is incorrect

14(d)(ii)	The trend continues in the same way **or** The seasonal pattern remains the same	B1	oe
	Additional Guidance		
	Trade is not affected by a recession / global events etc		B1
	Importing the same amount of clothes / price remains the same		B1

15(a)	$\frac{48.7-45.5}{2.4}$ **or** $1\frac{1}{3}$ **or** 1.3(3…)	M1	
	$\frac{x-41.7}{1.8}=\frac{48.7-45.5}{2.4}$ **or** (x =) 41.7 + 1.8 × their $1\frac{1}{3}$	M1 dep	oe
	44.1 (seconds)	A1	

<table>
<tr><td rowspan="8">15(b)</td><td>$\frac{43.7-45.5}{2.4}$ or $\frac{44.3-45.5}{2.4}$

or $\frac{40.5-41.7}{1.8}$ or $\frac{40.3-41.7}{1.8}$</td><td>M1</td><td>Correct calculation for, or value of, any standardised score</td></tr>
<tr><td>All standardised scores correct.

<table><tr><th></th><th>Kim</th><th>Pria</th></tr><tr><td>Race 1</td><td>–0.75 or –0.8 or $-\frac{3}{4}$</td><td>–0.5 or $-\frac{1}{2}$</td></tr><tr><td>Race 2</td><td>–0.6(6…) or –0.67 or – 0.7 or $-\frac{2}{3}$</td><td>–0.7(7…) or –0.78 or –0.8 or $-\frac{7}{9}$</td></tr></table></td><td>A2</td><td>oe

A1 for 2 or 3 correct standardised scores</td></tr>
<tr><td>Kim swam better in Race 1 with a reason,

eg as –0.75 < –0.67

or as her standardised score in Race 1 was lower than her score in Race 2</td><td>B1ft</td><td>ft from their standardised scores for Kim</td></tr>
<tr><td>Pria swam better in Race 2 with a reason,
eg as –0.78 < –0.5

or as her standardised score in Race 2 was lower than her score in Race 1</td><td>B1ft</td><td>ft from their standardised scores for Pria</td></tr>
<tr><td colspan="3">Additional Guidance</td></tr>
<tr><td colspan="2">If they misread the question and compare Kim with Pria in each race

Kim swam better than Pria in Race 1 with a reason,
eg as –0.75 < –0.67
or as her standardised score in Race 1 was lower

Pria swam better than Kim in Race 2 with a reason,
eg as –0.78 < –0.5
or as her standardised score in Race 2 was lower</td><td>B1ft B1ft from their standardised scores for each race</td></tr>
<tr><td colspan="2">If the table is blank check the answer space</td><td></td></tr>
<tr><td colspan="2">Furthest from the mean can only be awarded if both standardised scores are negative</td><td></td></tr>
<tr><td colspan="2">Kim swam quicker in race 1 but had a lower standardised score</td><td>B0</td></tr>
</table>

16(a)	A hypothesis should not be a question He has asked a question He has not predicted what will happen	B1	oe
	Additional Guidance		
	His hypothesis is a question, rather than a statement		B1
	His hypothesis should express his views about which round of golf will take less strokes		B1
	His hypothesis should be: On average players will take fewer stokes on Round 2		B1

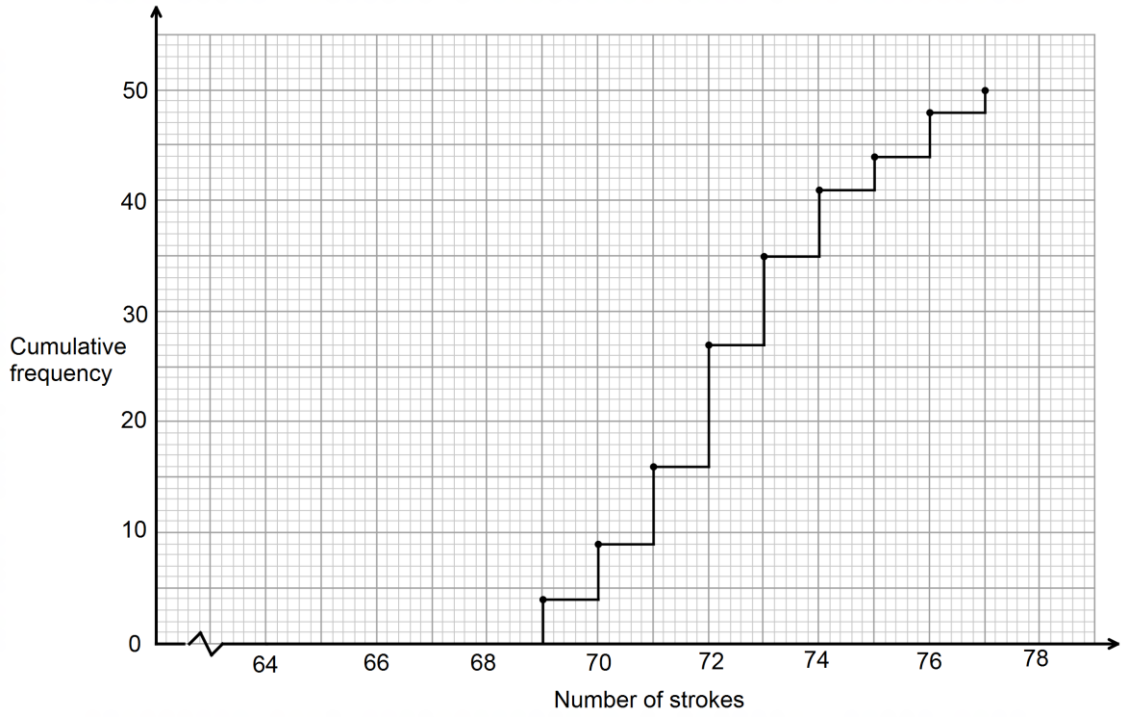
16(b)	The data are discrete	B1	oe
	Additional Guidance		
	The data can only take integer values		B1
	The data set is not continuous		B1
	You cannot have a decimal/fraction of a stroke		B1
	It shows exact data values		B0
	It is continuous		B0

16(c)	16 seen	M1	oe Allow ±0.5
	32(%)	A1	Allow ±1

16(d)	(UQ =) 76	B1	

16(e)	**Alternative 1** their 76 – 72 **or** 4	M1	
	72 – 1.5 × (their 76 – 72) **or** 72 – 1.5 × 4 **or** 66	M1	
	(Their lower fence) **or** 66 **and** A suitable conclusion, eg • 65 < lower fence **or** 65 < 66 • 65 is an outlier	A1ft	
	Alternative 2 their 76 – 72 **or** 4	M1	
	$\frac{65-72}{\text{their } 4}$ **or** $\frac{72-65}{\text{their } 4}$ **or** (–)1.75	M1	
	(–)their 1.5 × IQR **or** (–)1.75 A suitable conclusion involving 1.5, eg • 65 is more than 1.5 × IQR below the LQ • – their 1.5 × IQR < –1.5 **or** –1.75 < –1.5	A1	
	Additional Guidance		
	The conclusion could be implied by a preamble, eg An outlier is a value more than 1.5 × IQR below the LQ. $\frac{65-72}{4}$ = –1.75 so 65 is an outlier		

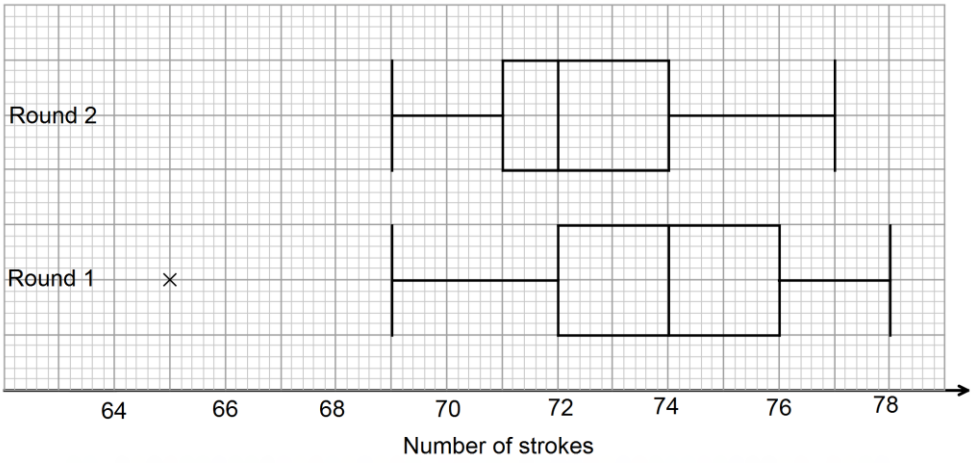
16(f)



Correct cumulative frequencies seen in table or implied by graph	B1	4 9 16 27 35 41 44 48 50
Their cumulative frequencies, which must increase, plotted at correct horizontal values (69, their 4), (70, their 9), (71, their 16), etc	M1	($\pm \frac{1}{2}$ small square) At least 7 points correct. Can be implied by graph passing through correct points.
Fully correct cumulative frequency step polygon including a vertical line from (69, 0) to (69, 4)	A1	

Additional Guidance

A cumulative frequency curve/polygon can score the first 2 marks.	

16(g)			
	For **Round 1**: Mark at 65 to indicate outlier **and** Lowest value marked at 69 **and** Maximum value marked at 78 (could be a dot)	M1	Outlier marked as cross or asterisk. The outlier should not be connected to the rest of the box plot.
	Round 1 Correct marks at the LQ (72), median (74) and their UQ	M1	ft UQ from part (d) provided that 74 < their UQ < 78
	Round 2 Correct marks at the LQ (their 71), median (their 72) and UQ (their 74)	M1	ft their cumulative frequency step polygon from part (f) Do not follow through from a cf frequency polygon/curve
	Both box plots fully correct with vertical lines marking the minimum and maximum values	A1ft	Follow through from sensible UQ from part (d) and from cumulative frequency step polygon from part (f).
	Additional Guidance		
	Tolerance of ($\pm \frac{1}{2}$ small square)		
	Both box plots fully correct but wrong way round		SC3

	(The median value for Round 2 is lower showing that) players generally needed fewer strokes on Round 2	B1ft	oe
	The scores on Round 2 are less spread out (because the range/IQR for Round 2 is less)	B1ft	oe
	Additional Guidance		
16(h)	Ignore references to skew Any figures given must be correct for their box plots **Comments about location** On average players used more strokes on Round 1 Players generally used fewer strokes on Round 2 because Round 2 has a smaller mean (BOD) The average number of strokes for Round 2 was lower (so players did better on Round 2 on average) The average number of strokes for Round 2 was 72 whereas in Round 1 it was 71 (insufficient comparison) The median number of strokes for Round 2 is smaller (median value not interpreted) The median value for Round 2 is lower showing that players generally did worse on Round 2 (incorrect interpretation)		B1 B1 B1 B0 B0 B0
	Comments about spread Scores in Round 2 are less varied/more consistent Round 2 has a smaller IQR (no interpretation of IQR)		B1 B0

16(i)	A suitable factor that could explain the lower median value for Round 2, eg Players become more familiar with the course Weather conditions Time of day the rounds were played	B1	oe
	Additional Guidance		
	Players improve (with practice)		B1
	More favourable pin positions in Round 2		B1
	Course conditions have changed		B1
	Difficulty of the course(s)		B1