



*Rewarding Learning*

**General Certificate of Secondary Education**

---

**Statistics**

Unit 1

Higher Tier

[GST12]

---

**Assessment**

**MARK  
SCHEME**

## General Marking Instructions

### Introduction

The mark scheme normally provides the most popular solution to each question. Other solutions given by candidates are evaluated and credit given as appropriate; these alternative methods are illustrated in the Supplementary Marking Guidance (see page 3).

The marks awarded for each question are shown in the right hand column and they are prefixed by the letters **M**, **A** and **MA** as appropriate. The key to the mark scheme is given below:

**M** indicates marks for correct method.

**A** indicates marks for accurate working, whether in calculation, readings from tables, graphs or answers.

**MA** indicates marks for combined method and accurate working.

The solution to a question gains marks for correct method and marks for an accurate working based on this method. Where the method is not correct no marks can be given.

A later part of a question may require a candidate to use an answer obtained from an earlier part of the same question. A candidate who gets the wrong answer to the earlier part and goes on to the later part is naturally unaware that the wrong data is being used and is actually undertaking the solution of a parallel problem from the point at which the error occurred. If such a candidate continues to apply correct method, then the candidate's individual working must be **followed through** from the error. If no further errors are made, then the candidate is penalised only for the initial error. Solutions containing two or more working or transcription errors are treated in the same way. This process is usually referred to as "follow-through marking" and allows a candidate to gain credit for that part of a solution which follows a working or transcription error.

It should be noted that where an error trivialises a question, or changes the nature of the skills being tested, then as a general rule, it would be the case that not more than half the marks for that question or part of that question would be awarded; in some cases the error may be such that no marks would be awarded.

### Positive marking

It is our intention to reward candidates for any demonstration of relevant knowledge, skills or understanding. For this reason we adopt a policy of **following through** their answers, that is, having penalised a candidate for an error, we mark the succeeding parts of the question using the candidate's value or answers and award marks accordingly.

Some common examples of this occur in the following cases:

- (a) a numerical error in one entry in a table of values might lead to several answers being incorrect, but these might not be essentially separate errors;
- (b) readings taken from candidates' inaccurate graphs may not agree with the answers expected but might be consistent with the graphs drawn.

When the candidate misreads a question in such a way as to make the question easier only a proportion of the marks will be available (based on the professional judgement of the examiner)

## GCSE Statistics Supplementary Marking Guidance GST12 Unit 1 (Higher Tier)

### Introduction

This guidance supplements the Mark Scheme (see page 7) and gives additional detail on the awarding of all marks and, where appropriate, acceptable alternative solutions. When the professional judgement of the marker is required to award a mark, the rationale for the mark is given. Please note that this guidance is designed to indicate how marks are awarded rather than to supply model solutions.

### Detailed Marking Instructions

Question	Marks	Instructions
1	A2	One mark for each reason.
2 (a)	A1	Correct definition.
(b)	A1 A1	Any correct advantage, e.g. data collected is suitable. Any correct disadvantage, e.g. can be time consuming to collect.
(c)	A1	Any suitable reason, e.g. imbalance between agree/disagree options. NB do not accept criticisms of the question.
(d)	A2	One mark for each independent reason, e.g. reference to the sample size being too small, sample includes only Year 8 pupils and no others, etc.
3 (a)	A1 A1	One mark for stating that the mean takes account of outliers. One mark for acknowledging that the median does not.
(b)	A1	Any suitable, practical consequence, e.g. financial impact.
(c)	A1	Any suitable, practical consequence, e.g. customer dissatisfaction.
(d)	A1	510 g (correct answer only).
(e)	A1 A1	Lower action line correctly drawn and labelled. Upper action correctly drawn and labelled. Special case: if a candidate has correctly drawn both lines but not labelled either one, they can get A1 A0. If a candidate labels two incorrect lines, A0 A0.
(f)	MA1 A1	511 g (correct answer only). Do not accept correct working followed by incorrect answer. Mean correctly plotted on the control chart. If MA0 has been awarded, follow through candidate's answer if possible. Special case: if a candidate plots (11, 511) on the control chart with no supporting working, allow MA1 A1. If they plot (11, $x$ ) where $x \neq 511$ with no supporting working, then award MA0 A0.
(g)	A1 A1	Observation, i.e. the tenth sample mean is within the warning lines . . . so no action is required by Evelyn.

- 4 (a) A1 Histogram (do not accept bar chart).
- (b) A1 Must mention “positive”. Award A0 for a description such as ‘skewed to the right’.
- (c) A1 Any reference to the unavailability of specific data points because of grouping.
- (d) M1 Clear attempt to find the total heights of the bars  
A1 57 (correct answer only)  
Special case: if 57 is seen and nothing else, M1 A1  
Candidates trying to use frequency density get M0 A0
- (e) MA1 Correct midpoints seen (may be within working).  
M1 Trying to find  $\Sigma fx$  and divide it by  $\Sigma f$ . This can be implied if totals are correct.  
MA1 Calculation for previous mark uses all correct values.  
A1 45.3 minutes.
- (f) A1 Conclusion, i.e. runners took longer to complete the previous race.  
A1 Justification, e.g. since the mean time taken was greater.
- 5 (a) (i) A1 117  
(ii) M1 Clear attempt to arrange the numbers in ascending or descending order  
A1 98  
(iii) MA1 108 - 92  
A1 16
- (b) A1 Reference to the mode being unrepresentative as it’s the largest value.
- (c) MA1 Minimum value, maximum value, median and quartiles correctly plotted (allow follow through for pupil’s values from part (a)).  
MA1 Correct box plot shape.
- (d) A1 Median time after the course is less than the median time before the course  
A1 . . . so the course was effective.  
A1 The inter-quartile range (IQR) after the course is less than the IQR before the course  
A1 . . . so the course resulted in more consistent times.
- 6 (a) A1 Data is discrete.
- (b) MA1 Clear attempt to read across from 50 on the vertical axis and down to the horizontal axis.  
A1 5  
Note: allow MA1 A1 if only 5 is seen without supporting evidence on the diagram.
- (c) (i) MA1 Clear attempt to read across from 90 on the vertical axis and down to the horizontal axis.  
A1 8  
Note: allow MA1 A1 if only 8 is seen without supporting evidence on the diagram.  
(ii) A1 Reference to 90%  
A1 Interpretation, e.g. names had 8 letters or fewer.  
Note: accept equivalent interpretation based on 10%.

- (d) A1 Comparison of median values.  
A1 Interpretation, e.g. girls have longer names on average.
- (e) (i) M1  $7 - 3$   
A1 4  
(ii) A1 90%  
A1 3 and 7
- 7 (a) A1 Reference to representativeness of subgroups.
- (b) M1 Knowing to work out one fraction, i.e. either  $\frac{x}{90}$  or  $\frac{112}{720}$   
MA1 Equation in x, e.g.  $\frac{x}{90} = \frac{112}{720}$  or  $x = \frac{112}{720} \times 90$   
A1  $x = 14$
- (c) A1 Any suitable category such as gender, age, etc.
- (d) MA1 Attempt to find the total number of full-time staff:  $(104 + 280 + 128) = 512$   
A1 One mark for writing the correct probability:  $\frac{32}{45}$  (or equivalent).
- Note: probability can be expressed as an unsimplified vulgar fraction, decimal or percentage.
- 8 (a) M1 Any sampling fraction seen, e.g.  $\frac{45}{x}$  or  $\frac{6}{85}$   
MA1 Correct equation in x, e.g.  $\frac{45}{x} = \frac{6}{85}$   
A1  $= 637.5$   
A1 Rounding to either 637 or 638
- (b) A2 One mark each for any 2 valid assumptions, e.g.  
- Population has not changed (no births/deaths)  
- No tags were lost  
- Tagged eels mixed with general population  
- Sample size was sufficient  
- All eels equally likely to being caught (samples are random)
- (c) A1 Continuous circled or otherwise identified.
- (d) MA1 Cumulative frequencies: 4, 9, 18, 34, 45 all correct.  
M1 Any clear method to find median, e.g. use of a correct formula or interpolation method.  
MA1 Above method using correct values.  
A1 81.5625 (allow correctly rounded value).
- (e) A1 Reference to grouping or unavailability of individual data items.
- 9 (a) (i) A1 95%  
(ii) A1 Reference to two standard deviations on each side of the mean.
- (b) A1 Positive z-score.  
A1 Weight of the rod is greater than the mean.

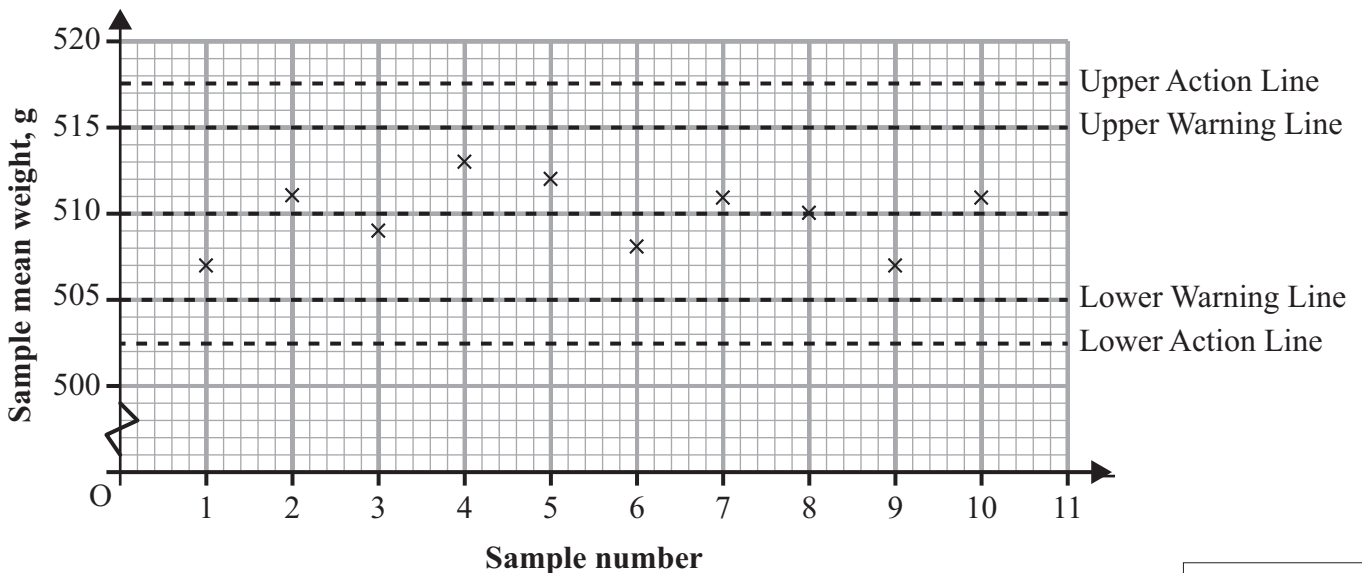
- (c) M1 Trying to use the formula  $z = \frac{x - \mu}{\sigma}$   
 MA1 Values used correctly in the above formula.  
 A1  $z = 28.08$   
 Note: if formula is stated but not used, award M0. If the formula is correctly used but has not been stated algebraically, award M1 MA1 and A1 if correct.
- (d) MA1 Correct method to find upper limit, i.e.  $24.8 + 3 \times 0.8$   
 A1 27.2  
 A1 Comparison and conclusion, i.e.  $27.4 > 27.2$  hence outlier.  
 Note: comparison with 27.4 must be clear for final A1.
- (e) A1 Machine B.  
 A1 Correct reason, i.e. z-score is closer to zero.
- 10 (a) A1 0.38  
 A2 All probabilities on branches correct. Penalise one mark per error to a maximum of two marks.
- (b) M1 Any valid method for finding the required probability, i.e.  

$$P(S_1 \cap S_2) = P(S_1 \cap \bar{S}_2) + P(\bar{S}_1 \cap S_2)$$
 or  

$$P(S_1 \cap S_2) = 1 - [P(S_1 \cap \bar{S}_2) + P(\bar{S}_1 \cap S_2)]$$
  
 MA1 Correct values used for chosen method.  
 A1 0.4712
- (c) A1 Fixed number of calls.  
 A1 Constant probability of a sale.
- (d) M1 Correct method used to find the required probability, i.e.  
 $P(S > 2) = P(S = 4) + P(S = 3)$   
 MA1 Values correctly substituted into the relevant terms.  
 A1 0.51 (accept suitable rounding).
- (e) M1 Knowing to calculate  $P(S = 4)$ ,  $P(S = 3)$  and  $P(S = 2)$   
 MA1 Identifying 0.3622 is the maximum probability.  
 MA1 3

- 1 The 3-dimensional effect makes comparing sizes difficult. A1  
 Extracting the newspaper sector makes it appear bigger. A1
- 2 (a) Primary data is collected by a researcher from first-hand sources. A1  
 (b) Any advantage, e.g. the data is more reliable. A1  
 Any disadvantage, e.g. more costly/time consuming. A1  
 (c) Suitable reason, e.g. The responses are not balanced between agree/disagree, respondents could tick more than one box. A1  
 (d) 1. The sample size is too small A1  
 2. The sample does not include pupils from other year groups. A1
- 3 (a) The mean is affected by outlying values whereas the median is not. A2  
 (b) Any appropriate consequence, e.g. loss of profit due to using more flour than necessary. A1  
 (c) Any appropriate consequence, e.g. complaints from customers due to the bags having less flour than expected. A1  
 (d) 510 g A1  
 (e) Action lines plotted at 502.5 g and 517.5 g. A2

AVAILABLE MARKS
2
6



(f)  $\bar{x}_{10} = \frac{509.7 + 511.6 + 509.3 + 512.9 + 511.5}{5} = 511 \text{ g}$  MA1

Point correctly plotted at (10, 511) A1

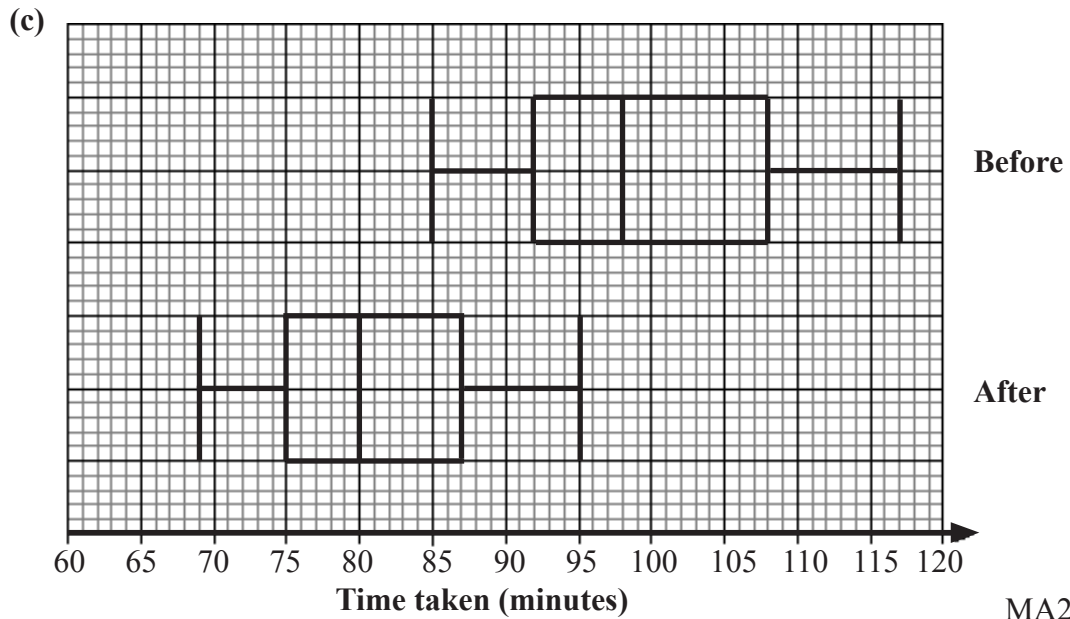
- (g) The tenth sample mean is within the warning lines so Evelyn needs to take no action. A2

11
----

- 4 (a) Histogram A1
- (b) Positively skewed A1
- (c) It is not possible to identify the least and greatest times since the data appears in groups. A1
- (d)  $12 + 19 + 14 + 8 + 4 = 57$  M1  
A1
- (e) Midpoints: 30, 40, 50, 60, 70 MA1
- Mean =  $\frac{12 \times 30 + 19 \times 40 + 14 \times 50 + 8 \times 60 + 4 \times 70}{57}$  M1 MA1  
= 45.3 minutes A1
- (f) Runners took longer, on average, to complete the previous race since the mean time was greater. A2

11

- 5 (a) (i) 117 minutes A1
- (ii) 85 88 91 92 94 95 97 98 100 103 104 108 109 117 117 M1  
98 minutes A1
- (iii)  $108 - 92 = 16$  minutes MA1  
A1
- (b) Because the mode (117) is the maximum value in the dataset and not representative of 'central value'. A1



- (d) The median time after the course is less than the median time before so the training course was effective. A1  
A1
- The IQR of times after the course is less than the IQR before the course so the training course resulted in more consistent times. A1  
A1

12

			AVAILABLE MARKS
6	(a) Because the data is discrete.	A1	13
	(b) 5	MA1 A1	
	(c) (i) 8	MA1 A1	
	(ii) 90% of the boys surveyed had names which were 8 letters long or less.	A2	
	(d) Median for girls (6) is larger than median for boys (5) therefore girls have longer names on average.	A2	
	(e) (i) $7 - 3 = 4$	M1A1	
	(ii) Among the girls, <b>80%</b> had names containing between <b>3</b> and <b>7</b> letters.	A2	
7	(a) To ensure each subgroup of the population is adequately represented in the sample.	A1	7
	(b) $\frac{x}{90} = \frac{112}{720}$	M1 MA1	
	$x = \frac{112}{720} \times 90$		
	$x = 14$	A1	
	(c) Any suitable category: gender, age etc.	A1	
	(d) $104 + 280 + 128 = 512$	MA1	
	$\frac{32}{45}$	A1	
8	(a) $\frac{45}{x} = \frac{6}{85}$	M1 MA1	12
	$6x = 3825$		
	$x = 637.5$	A1	
	637 or 638	A1	
	(b) Any 2 valid assumptions: Population has not changed (no births/deaths) <b>OR</b> No tags were lost <b>OR</b> Tagged eels mixed with general population <b>OR</b> Sample size was sufficient <b>OR</b> All eels equally likely to be caught (samples are random) etc.	A1 A1	
	(c) Continuous	A1	
	(d) Cumulative Frequencies: 4, 9, 18, 34, 45	M1	
	$\frac{m - 80}{85 - 80} = \frac{23 - 18}{34 - 18}$ (or appropriate alternative method)	MA2	
	$m = 81.5625$	A1	
	(e) Because the data is grouped.	A1	

- 9 (a) (i) 95% A1
- (ii) 95% of the data lie within two standard deviations of the mean. A1
- (b) The z-score is positive so the weight of the rod is greater than the mean. A2
- (c)  $z = \frac{x - \mu}{\sigma}$  M1
- $-0.1 = \frac{x - 28.2}{1.2}$  MA1
- $x = 28.08 \text{ g}$  A1
- (d) Upper limit =  $24.8 + 3 \times 0.8$  MA1
- $= 27.2$  A1
- $27.4 > 27.2$  hence outlier A1
- (e) Machine B A1
- The z-score is closer to zero. A1

AVAILABLE  
MARKS

12

