



*Rewarding Learning*

**General Certificate of Secondary Education**

---

**Statistics**

Unit 2

Higher Tier

[GST22]

---

**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 GST22 Unit 2 (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	(a) (i)	A1 –2000 (allow a suitable approximation).	
		(ii) A1 Statement in context, e.g. 2000 more people left than arrived.	
	(b)	A1 Gradual increase between 2001 and 2012	
		A1 Decrease after 2012 Note: 2012 can be approximated but a cut-off point must be mentioned. A response such as “increases then decreases” would earn A1 A0.	
	(c)	(i)	A1 2001
			A1 2010
			A1 2014
		(ii)	A1 Zero migration.
			A1 Overall change was due to natural change only.
	2	(a)	MA1 100
(b)		M1 Trying to find $\Sigma fx$ and divide it by 100 (may be implied by working).	
		MA1 Values used in $\Sigma fx$ calculation all correct.	
		A1 0.93 Note: if the pupil divides by 5 (or 4), maximum is M0 MA1 A0	
(c)		MA1 4	
(d)		A1 Any suitable diagram.	
		A1 Reference to discrete data.	
(e)		(i) A1 Opportunity sampling (accept convenience sampling).	
		(ii) A1 Possible reason, e.g. some of those surveyed may belong to the same family.	
3		(a)	A1 Suitable question, e.g. How has the age and gender distribution of the population of Northern Ireland changed between June 2008 and June 2018? Note: if expressed as a hypothesis, award A0. Accept mid-2008 for June 2008, etc here and throughout.
	(b)	A1 Age.	
		A1 Gender.	
		A1 For both mid-2008 and mid-2018	
	(c)	A1 Any suitable secondary source, e.g. NISRA website. Note: the source must be named. Do not accept descriptions of how primary data could be obtained or any reference to sampling.	

- (d) A1 Population pyramid (accept histogram).  
A1 . . . for each year.  
Note: for the second A1, there must be an understanding that the diagrams will be used for comparison, so a superimposed line graph over a population pyramid, as in the pre-release materials, is acceptable.
- (e) A1 A suitable calculation, e.g. median age.  
A1 Reference to comparing values for June 2008 and June 2018  
A1 Suitable interpretation, e.g. larger median indicates older population.
- 4 (a) A1 1.714 million (must be in millions – do not accept 1 714 000).
- (b) A1 Concluding statement, i.e. yes.  
A1 Justification, e.g. reference to the gradual increase in proportion of the population aged 85+ between mid-2001 and mid-2017
- (c) A1 Any reference to double counting.  
A1 Specific reference to 23,000  
A1 Mention of affected number 224 100
- (d) A1 Suitable graph, e.g. line graph or time series graph.
- (e) A1 100
- (f) A1 Mid-2017  
A1 10.8%  
A1 Mid-2001
- 5 (a) A1  $20 < r \leq 30$  (accept 20 – 30).
- (b) MA2 All cumulative frequencies correct. Penalise one mark per error to a maximum of two marks.
- (c) MA1  $20 < r \leq 30$  (accept 20 – 30).
- (d) MA1 Cumulative frequencies plotted correctly using upper class boundaries.  
MA1 Points joined (smooth curve or straight lines acceptable).  
A1 Curve/line starts at (0, 0).
- (e) (i) MA1 26 mm (follow pupil's diagram).  
(ii) MA1 33 – 19 (follow pupil's diagram).  
A1 14 mm (pupil's value).  
(iii) MA1 106 seen on diagram or in working (pupil's value at 38 mm).  
MA1 Percentage calculation presented correctly.  
A1 11.7% (follow pupil's working).
- (f) A1 Reference to grouped data.
- 6 (a) A1 No sampling frame (accept a description of this).  
Answer must refer to why it is not possible rather than why it might be difficult, so do not accept reasons like 'it would take too long'.
- (b) A1 Any suitable advantage, e.g. efficiency.  
A1 Any suitable disadvantage, e.g. low response rate.

- (c) A1 To see the effect of the installing the new checkout  
A1 . . . without actually installing it.
- 7 (a) MA1 Ranks correct (may rank in either order). Averaging for tied ranks must be used  
MA1  $\Sigma d^2 = 6.5$   
M1 Trying  $n = 8$  and their  $\Sigma d^2$  in the correct formula. This mark can be awarded even if the pupil has made a calculation error in finding  $\Sigma d^2$ .  
A1 0.923  
Note: if pupils calculate the product-moment correlation coefficient (0.8965) award no marks.
- (b) A1 Description – (strong) positive correlation.  
A1 Interpretation in context – tastes are in general agreement.
- (c) A1 Bivariate.  
A1 Random on random.
- 8 (a) MA1 4420  
MA1 2800
- (b) M1 Identifying 4420 and  $(5200 + 2800)$   
MA1 0.5525 (accept answer as a vulgar fraction).
- (c) (i) M1 Trying to divide 780 by 5200  
A1 0.15  
(ii) MA1 0.8125
- (d) M1 Answer to (c)(ii) divided by answer to (c)(i).  
A1 5.42 (allow follow through from (c)(ii)).
- (e) A1 More than 5 times more likely to have a cracked screen  
A1 . . . if they had not applied a screen protector than if they had.
- 9 (a) A1 Explicit reference to measurement of age in completed/full years.  
A1  $30 - 15 = 15$
- (b) MA2 All frequency densities correct (penalise one mark per mistake to a maximum of two marks).  
MA2 All bars correctly plotted as per pupil's table (allow follow through).  
Note: if frequency density has not been used, MA0.
- (c) A1 Paula may not be correct.  
A1 Reason, e.g. sample may not include those aged 90 or over.
- (d) A1 No.  
A1 Justification, e.g. histogram shows a positive skew.
- 10 (a) M1 Correct structure for calculation, which need not be in one go:  $\frac{p_1}{I_1} \times 100$   
MA1 Using  $p_1 = 110750$  and  $I_1 = 108.7$   
A1 £101,886 (or £101,885.93)  
Note: If answer is given as £101,885.925 or anything other than rounding to the nearest £ or p, then award M1 MA1 A0.

- (b) MA1  $136179/129601 \times 100$   
 A1 105.1
- (c) A1 House prices were 5.1% more expensive  
 A1 . . . in 2018 than 2017
- (d) (i) A1 Geometric mean (must specify ‘geometric’ even if correctly calculated in (ii)).
- (ii) MA1  $\sqrt[5]{(108.7 \times 107.2 \times 105.3 \times 103.6 \times 105.1)}$   
 A1 106.0  
 Note: five index numbers must be used for MA1. Allow follow through using pupil’s answer to (b).
- (e) M1 Trying to find  $\Sigma I_i p_i$  divided by  $\Sigma p_i$   
 MA1 Values used correctly.  
 A1 106.6

			AVAILABLE MARKS	
1	(a) (i)	-2000	A1	9
	(ii)	Approximately 2000 more people left Northern Ireland than arrived.	A1	
	(b)	Gradual increase until 2012 followed by a decrease.	A2	
	(c) (i)	2001, 2010, 2014	A3	
	(ii)	These are the three years when migration was zero so overall change was due to natural change only. <ul style="list-style-type: none"> <li>• three years net migrating very small (near zero)</li> <li>• overall change due to natural change</li> </ul>	A2	
2	(a)	$27 + 58 + 11 + 3 + 1 = 100$	MA1	9
	(b)	Mean = $\frac{0 \times 27 + 1 \times 58 + 2 \times 11 + 3 \times 3 + 4 \times 1}{100}$ = 0.93	M1 MA1 A1	
	(c)	Range = $4 - 0 = 4$	MA1	
	(d)	Bar chart, since the number of pets is a discrete variable.	A2	
	(e) (i)	Opportunity sampling	A1	
	(ii)	Some of the people surveyed may be from the same family.	A1	
3	(a)	How has the age and gender distribution of the population of Northern Ireland changed between June 2008 and June 2018?	A1	10
	(b)	Age and gender of the population in mid-2008 and mid-2018	A3	
	(c)	NISRA website	A1	
	(d)	Comparative population pyramid	A2	
	(e)	She could calculate the median age of the population by gender for both years. The larger median age would indicate that the population was, on average, older at that time.	A3	
4	(a)	1.714 million	A1	11
	(b)	Yes because there has been a gradual increase in the proportion of the population aged 85+ between mid-2001 and mid-2017	A2	
	(c)	The final figure (23 500) is included in the previous one (224 100) so Teresa's total includes the 23 500 twice.	A3	
	(d)	Line graph.	A1	
	(e)	100	A1	
	(f)	The population in <u>mid-2017</u> is <u>10.8%</u> more than it was in <u>mid-2001</u>	A3	

5 (a)  $20 < r \leq 30$

A1

(b)

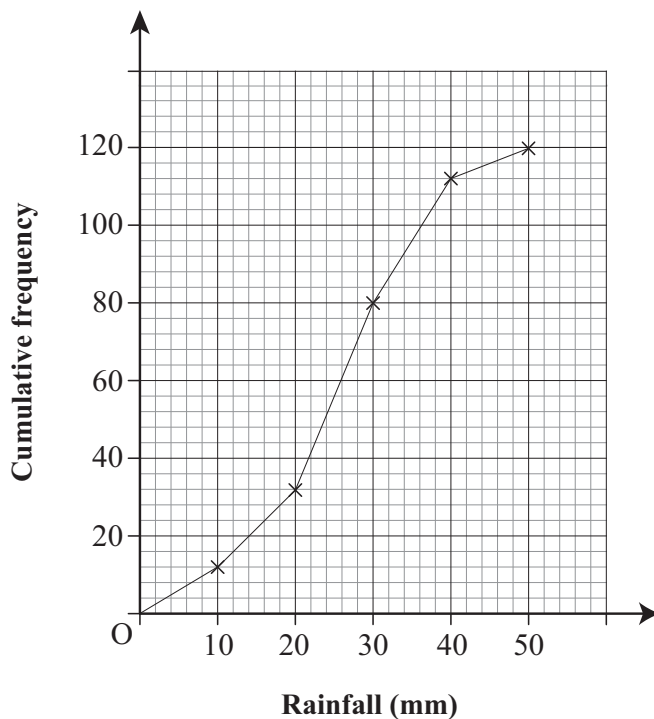
Rainfall, $r$ , (mm)	Number of days	Cumulative frequency
$0 < r \leq 10$	12	12
$10 < r \leq 20$	20	32
$20 < r \leq 30$	48	80
$30 < r \leq 40$	32	112
$40 < r \leq 50$	8	120

MA2

(c)  $20 < r \leq 30$

MA1

(d)



MA2 A1

(e) (i) 26 mm

MA1

(ii)  $\text{IQR} = 33 - 19$   
 $= 14 \text{ mm}$

MA1

A1

(iii)  $\frac{120 - 106}{120} \times 100$

MA2

$= 11.7\%$

A1

(f) A reference to grouped data given in the table

A1

AVAILABLE  
MARKS

14

- 6 (a) No sampling frame. A1
- (b) Advantage: no need for someone to carry out the survey. A1  
Disadvantage: possible low response rate. A1
- (c) It allows the manager to see whether installing a new checkout would have the desired effect without going to the expense of actually installing it. A2

7 (a)

$R_E$	4	1	7	5	3	2	8	6
$R_R$	4	2	6.5	5	1	3	8	6.5
$d$	0	-1	0.5	0	2	-1	0	-0.5
$d^2$	0	1	0.25	0	4	1	0	0.25

MA1

$\Sigma d^2 = 6.5$  MA1

$$r_s = 1 - \left( \frac{6 \Sigma d^2}{n(n^2 - 1)} \right)$$

$$= 1 - \left( \frac{6 \times 6.5}{8(8^2 - 1)} \right)$$

M1

$$= 0.923$$

A1

- (b) There is a strong positive correlation between the scores so Emma's and Ronan's tastes are in general agreement. A2
- (c) Bivariate data A1  
Random on random data A1

- 8 (a) 4420 (screen protector, no crack) MA1  
2800 (no screen protector) MA1

(b)  $\frac{4420}{5200 + 2800} = \frac{221}{400} = 0.5525$  M1 MA1

(c) (i)  $\frac{780}{5200} = 0.15$  M1 A1

(ii)  $\frac{2275}{2800} = 0.8125$  MA1

(d) Relative Risk =  $\frac{0.8125}{0.15}$  M1  
= 5.42 A1

- (e) Mobile phone owners in the town are more than 5 times more likely to have a cracked screen after one year if they have not applied a screen protector than if they had applied one. A2

AVAILABLE  
MARKS

5

8

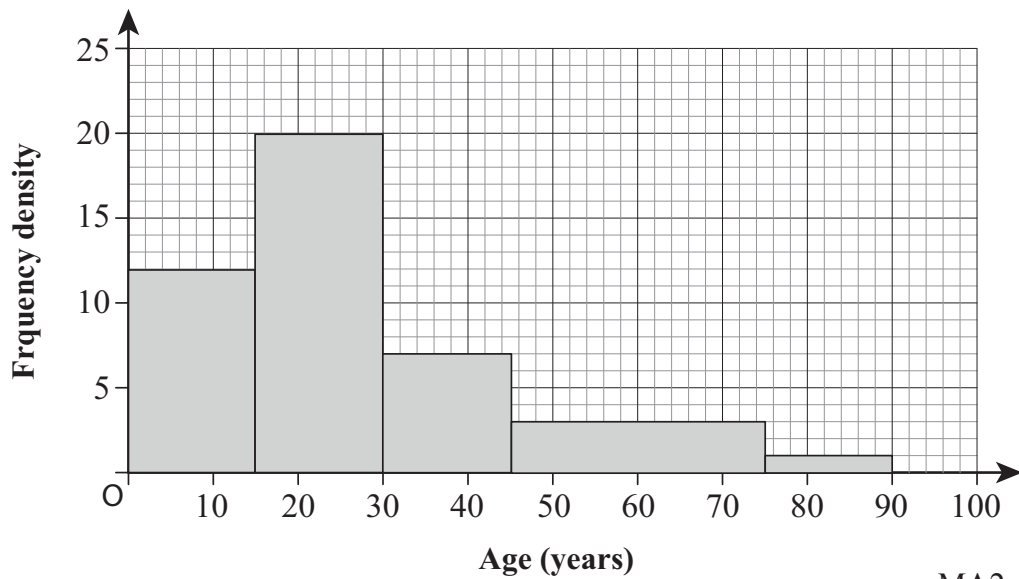
11

- 9 (a) The upper class boundary is 30 because age is measured in completed years, so the class width is  $30 - 15 = 15$  A2

(b)

Age, $A$ , (years)	Frequency	Class width	Frequency Density
$0 \leq A \leq 14$	180	15	12
$15 \leq A \leq 29$	300	15	20
$30 \leq A \leq 44$	105	15	7
$45 \leq A \leq 74$	90	30	3
$75 \leq A \leq 89$	15	15	1
$A > 90$	0		0

MA2



MA2

- (c) This is a sample of residents and no information is given on the sampling method so those aged 90 or more may have been left out, so Paula may not be correct. A2
- (d) The normal distribution would not be an appropriate model for this data as the histogram indicates a positive skew to the data. A2

10

10 (a) $\frac{110750}{108.7} \times 100$	M1 MA1	AVAILABLE MARKS
= £101 886	A1	
(b) $\frac{136179}{129601} \times 100$	MA1	
= 105.1	A1	
(c) The house prices were 5.1% more expensive in 2018 than 2017	A2	
(d) (i) Geometric mean	A1	
(ii) $\sqrt[5]{108.7 \times 107.2 \times 105.3 \times 103.6 \times 105.1}$	MA1	
= 106.0	A1	
(e) $\frac{(523 \times 103.6) + (671 \times 107.2) + (458 \times 109.1)}{523 + 671 + 458}$	M1 MA1	
= 106.6	A1	
<b>Total</b>	13	
	<b>100</b>	