



*Rewarding Learning*

**General Certificate of Secondary Education  
2023**

---

**Statistics**

Unit 2

Higher Tier

[GST22]

**MONDAY 19 JUNE, AFTERNOON**

---

**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 not usually illustrated in the published mark scheme.

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.

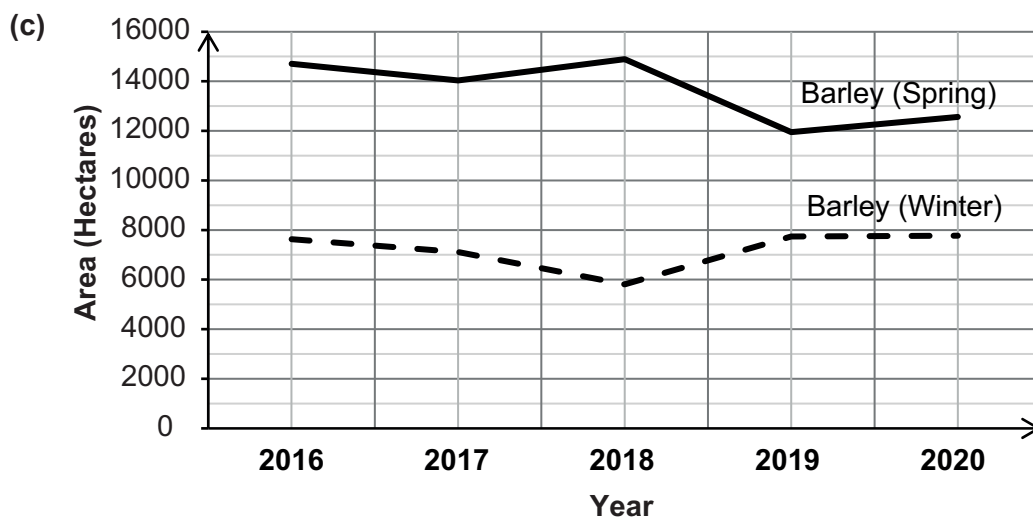
Some common examples of this occur in the following cases:

- (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).

1 (a) Increasing A1

(b) Rounding of numbers A1



(d) It is not clear from the diagram that the numbers are in thousands A1

2 (a) Members of the farming workforce in Northern Ireland A1

(b) It is unlikely that a suitable sampling frame exists A1

(c) A suitable hypothesis linking farm size and the number of hours worked, e.g. staff on large farms work more hours per week than those on small farms. A1

(d) Farm Size A1  
Average number of hours worked per week A1

(e) Since 'farm size' is a categorical variable and measured as small or large, the product-moment correlation coefficient is not suitable as both variables need to be numerical. A2

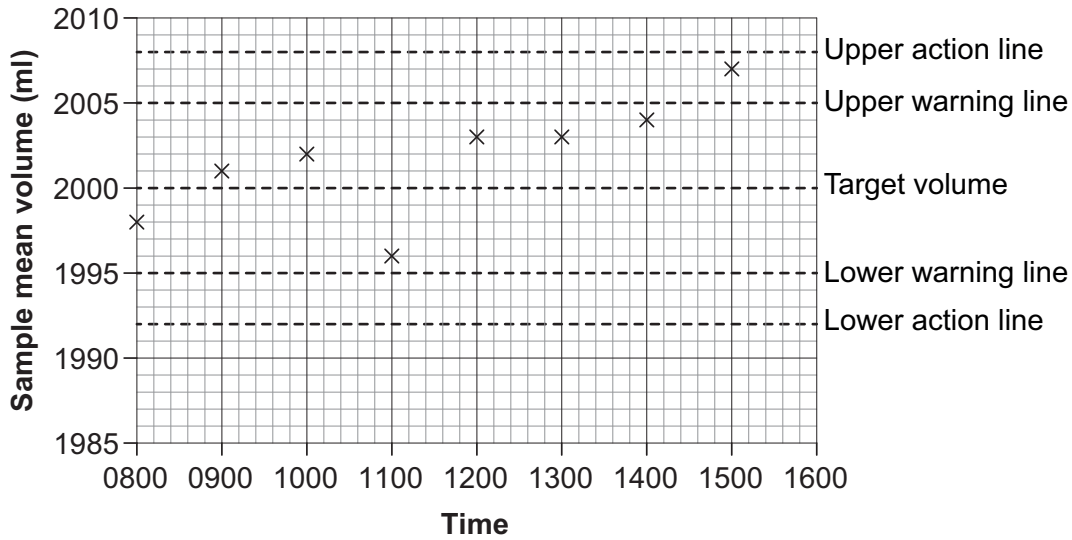
(f) Suitable question and response section. A2

AVAILABLE MARKS

5

9

3 (a)



A2

- (b) Target line drawn at 2000  
 Warning lines drawn and labelled at 2005 and 1995  
 Action lines drawn and labelled at 2008 and 1992

A3

- (c) The shift manager should take another sample straight away as the sample mean is above the warning limit. If the new sample mean taken is also high, he should reset the machine.

A3

- (d) GM Advantage – If they do the process more often, they could pick up machine errors more quickly.

GM Disadvantage – This would take more time to carry out.

A2

- (e) PM Advantage – They would ensure that the lemonade bottles are filled more accurately as they would get more frequent warnings and actions.

PM Disadvantage – They may have to stop production and reset the machines more often if they reduce the limits.

A2

- (f) The range would show if there is consistency within the sample. The mean could be on target but some of the bottles could be too high or too low.

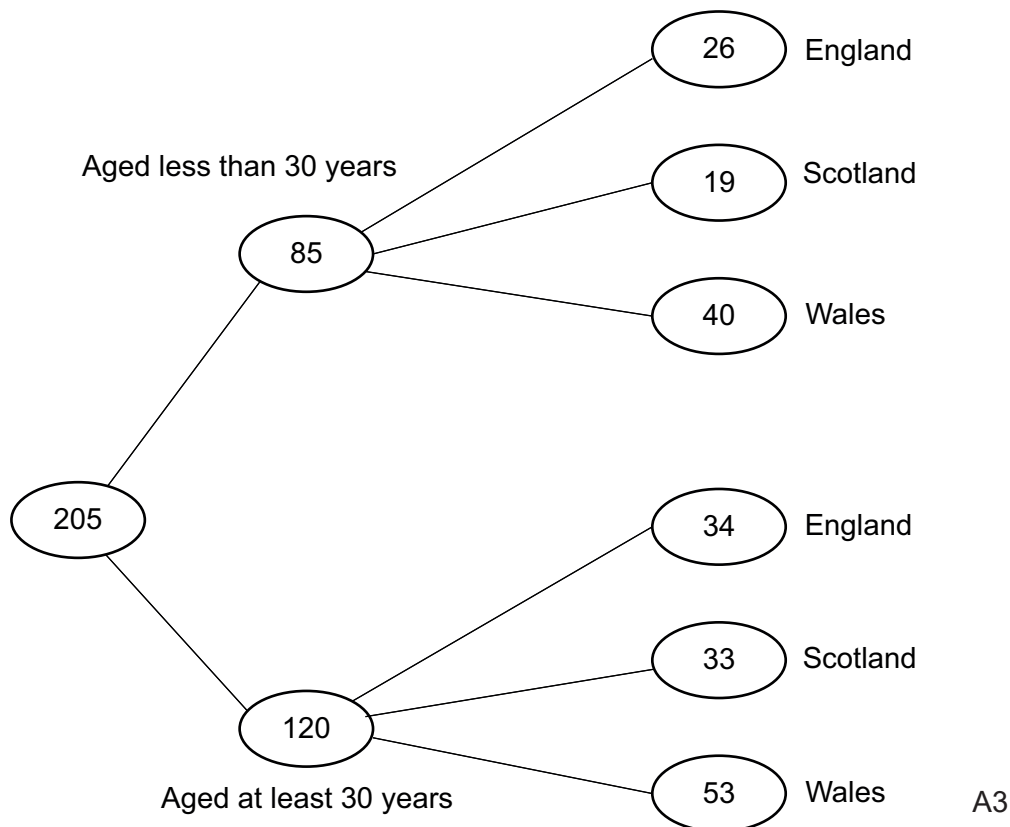
A2

AVAILABLE MARKS

14

		AVAILABLE MARKS
4	<p><b>(a)</b> Read 1 000 000 and 25 000 from graph</p> $\frac{1\,000\,000}{25\,000}$ <p>40 hectares</p>	MA1 MA1 A1
	<p><b>(b)</b> The number of hectares farmed stays very consistent, but the number of farms decreases over the time period.</p>	A2
	<p><b>(c)</b> There are fewer farms as smaller farms may have been bought over by bigger farms or combined with other small farms.</p>	A1
	<p><b>(d)</b> 0.933</p>	A2
	<p><b>(e) (i)</b> Positive correlation</p>	A1
	<p><b>(ii)</b> As the number of large farms increases the total number of pigs increases <b>or</b> Large farms are more likely to keep more pigs.</p>	A1
		10

5 (a)

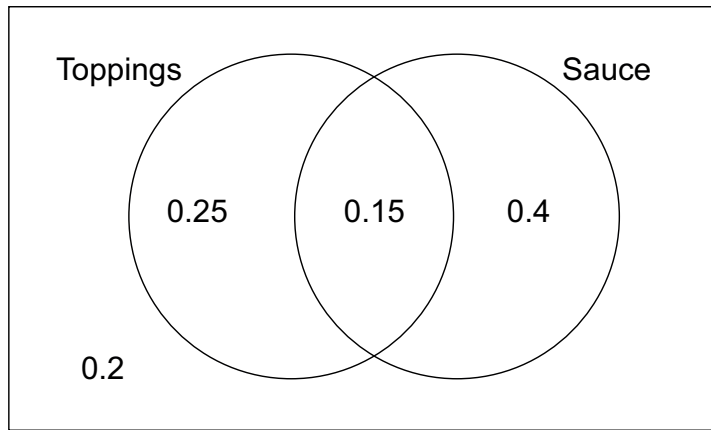


- (b) (i)  $\frac{19}{205}$  MA2
- (ii)  $\frac{26 + 34}{205} = \frac{60}{205} = \frac{12}{41}$  MA2
- (c)  $\frac{19 + 26}{85} = \frac{45}{85} = 9/17$  MA2
- (d) Janet is incorrect because 47.1% ( $\frac{40}{85}$ ) of those aged under 30 go on holiday to Wales which is greater than the 44.2% ( $\frac{53}{120}$ ) of those aged 30 or over who go on holiday to Wales. MA2 A1
- (e) The frequency tree shows the totals for the sub-categories. A1

AVAILABLE  
MARKS

13

6 (a)



MA2

(b)  $0.25 + 0.4 = 0.65$

MA2

(c)  $P(\bar{S} | \bar{T}) = \frac{0.2}{0.2 + 0.4} = \frac{1}{3}$

MA2

(d) (i)  $0.8 \times 80 = 64$

MA1

(ii) Knowing expected sales figures is useful for ensuring that there is sufficient stock available to meet demand

A1

8

AVAILABLE  
MARKS

7 (a)

Journey time, $t$ (minutes)	Frequency	Cumulative frequency
$0 < t \leq 60$	2	2
$60 < t \leq 120$	3	5
$120 < t \leq 180$	7	12
$180 < t \leq 210$	8	20
$210 < t \leq 240$	6	26
$240 < t \leq 300$	4	30

M1 A1

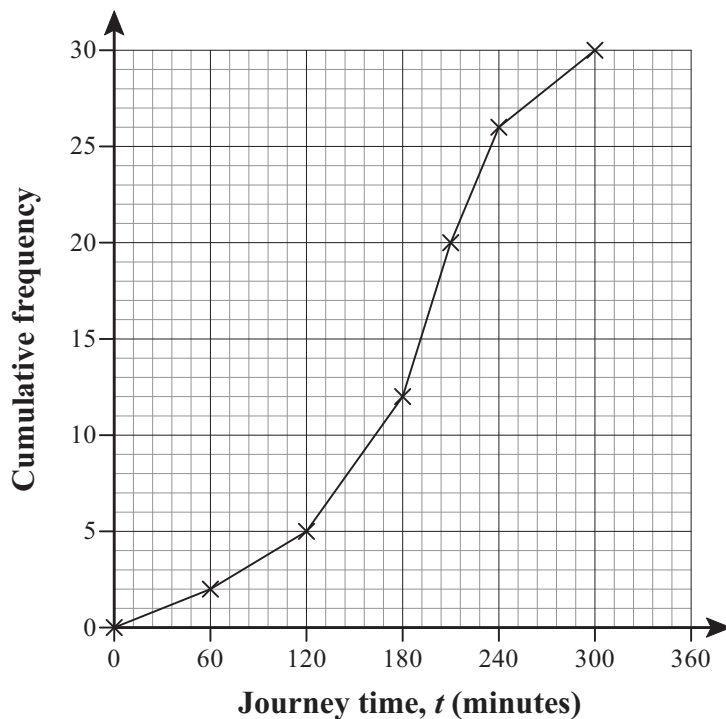
(b) 5

A1

(c)  $180 < t \leq 210$

A1

(d)



MA3

(e)

	Median (minutes)	Interquartile range (minutes)
<b>Group A</b>	192	81
<b>Group B</b>	50	40

MA4

(f) The journey times for the passengers in group B were, on average, much shorter than those in group A since the median journey time was smaller.

A2

The journey times for the passengers in group B showed less variation than those in group A since the interquartile range of the journey times was smaller.

A2

15

			AVAILABLE MARKS		
8	(a)	0.63	MA1	8	
	(b)	$n = 5$ $p = 0.37$	A1 A1		
	(c)	(i) Let $D$ = the number of days Kate takes sandwiches to school. $P(D = 1) = 5pq^4$ $= 5(0.37)(0.63)^4$ $= 0.2914$	MA1 A1		
		(ii) $P(D \geq 1) = 1 - P(D = 0)$ $= 1 - (0.63)^5$ $= 0.9008$	M1 MA1 A1		
9	(a)	$A_{2015} = \frac{8616}{108} \times 100$ $= 7978$ hectares	M1 MA1 A1		8
	(b)	$C_{2020} = \frac{7132}{8104} \times 100$ $= 88.0$	MA1 A1		
	(c)	The crop area for wheat in Northern Ireland decreased by 12% between 2019 and 2020	A3		
10	(a)	Mean = $\frac{42 + 23 + 34 + 18 + 22 + 55 + 20 + 36 + 27 + 43}{10} = \frac{320}{10} = 32$	MA1		10
	(b)	$\Sigma x^2 = 11556$ $\sigma = \sqrt{\frac{11556}{10} - 32^2}$ $= 11.47$ years	M1 MA1 A1		
	(c)	$z_A = \frac{28 - 32}{11.47} = -0.3487\dots$ $z_S = \frac{28 - 29.5}{8.3} = -0.1807\dots$ The person is more likely to belong to the soprano section of the choir as the standardised score is closer to zero	M1 MA1 MA1 A2		
	(d)	The small number of members of the alto section of the choir may mean that using the normal distribution as a model may not be appropriate.	A1		
			<b>Total</b>	<b>100</b>	