Mark Scheme (Results)

Summer 2025

Pearson Edexcel GCE
In Statistics (9ST0)
Paper 01: Data and Probability

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2025
Question Paper Log Number 76381
Publications Code 9ST0_01_2506_MS

General Marking Guidance

Total marks

The total number of marks for the paper is 80.

Mark types

The Edexcel Statistics mark schemes use the following types of marks:

- **M** **Method** marks,
  awarded for 'knowing a method and attempting to apply it',
  unless otherwise indicated.
- **A** **Accuracy** marks can only be awarded if the relevant method (M) marks have been earned.
- **B** **Unconditional accuracy** marks are independent of M marks
- **E** **Explanation** marks

NOTE: Marks should not be subdivided.

Abbreviations

These are some of the marking abbreviations that will appear in the mark schemes.

- ft follow through
- PI possibly implied
- cao correct answer only
- cso correct solution only
  (There must be no errors in this part of the question)
- awrt answers which round to
- awfw answers which fall within (a given range)
- SC special case
- nms no method shown
- oe or equivalent
- dep dependent (on a given mark or objective)
- dp decimal places
- sf significant figures
- ∗ The answer is printed on the paper

Further notes

- All candidates must receive the same treatment. Examiners must mark the first candidate in exactly the same way as they mark the last.
- Mark schemes should be applied **positively**. Candidates must be rewarded for what they have shown they can do rather than penalised for omissions.
- Examiners should mark according to the mark scheme not according to their perception of where the grade boundaries may lie.
- There is **no ceiling** on achievement. All marks on the mark scheme should be used appropriately.
- All the marks on the mark scheme are designed to be awarded. Examiners should always award full marks if deserved, i.e. if the answer matches the mark scheme. Examiners should also be prepared to award zero marks if the candidate's response is not worthy of credit according to the mark scheme.
- Where some judgement is required, mark schemes will provide the principles by which marks will be awarded and exemplification may be limited.
- When examiners are in doubt regarding the application of the mark scheme to a candidate's response, the team leader must be consulted.
- All A marks are 'correct answer only' (cao), unless shown, for example, as A1ft to indicate that previous wrong working is to be followed through.
- All M marks are 'possibly implied' (PI) unless specifically stated otherwise in the 'Notes' column.
- After a **misread**, the subsequent A marks affected are treated as A1ft, but manifestly absurd answers should never be awarded A marks.
- **Crossed out** work should be marked UNLESS the candidate has replaced it with an alternative response.
- If **two solutions** are given, each should be marked, and the resultant mark should be the mean of the two marks, rounded down to the nearest integer if needed.

Question	Scheme	Marks	AO	Notes
1 (a)	Median = 114	B1	1.1	
	IQR = 120 – 103	M1	1.1	
	= 17	A1	1.1	Must be seen in (a)
1 (b)	Median = 113	B1	1.1	
	IQR = 12	B1	1.1	Allow range for 2019-2020 = 35 and range for 2020-2021 = 44
	On average the number of points scored is similar as medians are similar	E1ft	2.1b	ft their median Allow 2020-2021 is less than 2019-2020
	2019-2020 is more varied/less consistent/more spread out than 2020-2021 as IQR is bigger in 2019-2020	E1ft	2.1b	ft their IQR or range Allow 2020-2021 is more varied/less consistent/more spread out than 2019-2020 as range is wider in 2020-2021
	Total	**7**		

Question	Scheme	Marks	AO	Notes
2	Mean for Pakistan = 356.8	B1	1.2	
	Standard deviation for Pakistan = 64.8	B1	1.2	
	Analysis draws on combination of both mean **and** standard deviation	E1		Standard deviation must have some reference to spread/consistency/varies
	Coefficient of variation is $\frac{64.8}{356.8}=0.182$	E1	2.1b	
	Or Coefficient of variation for Ethiopia is $\frac{36.1}{295.5}=0.122$		2.1b	
	Hence, relative standard deviation for Pakistan is higher, so that the conclusion is incorrect.	depE1	3.1a	Dependent on previous 2 E1 marks
	Total	**5**		

Alternative:

Question	Scheme	Marks	AO	Notes
2	Mean for Pakistan = 356.8	B1	1.2	
	Standard deviation for Pakistan = 64.8	B1	1.2	
	Analysis draws on combination of both mean **and** standard deviation	E1		Standard deviation must have some reference to spread/consistency/varies
	Thus, the two means are within a factor of 1.2 of each other	E1	2.1b	
	Or			
	The standard deviations are within a factor of 1.8 of one another		2.1b	Accept nearly double/half
	Hence, relatively, the standard deviation for Pakistan is higher, so that the conclusion is incorrect.	depE1	3.1a	Dependent on previous 2 E1 marks
	Total	**5**		

Question	Scheme	Marks	AO	Notes
3 (a)	$P(2\text{ walk}) = 0.65\times0.45\times0.8$ $+0.65\times0.55\times0.2$ $+0.35\times0.45\times0.2$	M1	1.2	**Alt:** P(1 walk)=0.4505
	$P(3\text{ walk}) = 0.65\times0.45\times0.2$ (=0.0585)	M1	1.2	P(0 walk) = 0.154
	$P(\text{at least 2 walk})$ $= P(2\text{ walk}) + P(3\text{walk})$	ddM1	1.2	P(at least 2 walk) = 1- P(1 walk) - P(0 walk)
	$= 0.337 + 0.0585 = \frac{791}{2000} = 0.3955$	A1	1.2	awrt 0.396 Correct answer without working 4/4
3 (b) (i)	$P(\text{WW}) = 0.65\times0.7$ $= \frac{91}{200} = 0.455$ or $P(CC) = 0.1\times0.9$ $= \frac{9}{100} = 0.09$	M1	1.2	
	$P(\text{WW or CC}) = \frac{91}{200} + \frac{9}{100}$	M1	1.2	
	$= \frac{109}{200} = 0.545$	A1	1.2	Correct answer without working 2/2
3 (b) (ii)	$P(\text{Different}) = 1 - 0.545 = 0.455$	B1ft	1.2	**Alt:** 0.25+0.1x0.1+0.65x0.3 = 0.455
	Total	**8**		

Question	Scheme	Marks	AO	Notes
4	**Random variable *A***			
	There are a fixed number of seeds (30), 2 outcomes (germinate or not germinate) and the probability a seed germinates is constant (0.8)	E1	2.1a	E1 for any **two** of: • There are a fixed number of seeds • 2 outcomes (germinate or not germinate) • the probability a seed germinates is constant
	Assume that seeds germinate independently	E1	2.1a	
	So, B(30, 0.8) is a suitable model	dB1	1.2	Dependent on first E mark, 30 and 0.8 must be seen within their answer
	Random variable *B*			
	Birth weight is a continuous variable and both Poisson and binomial are discrete distributions	E1	1.2	
	So, neither distribution is suitable	dB1	1.2	Dependent on previous E mark
	Random variable *C*			
	Cars arrive at the toll booth singularly **and** at a constant rate (5 per minute)	E1	2.1a	Must have **both**
	Assume that cars arrive at the toll booth independently	E1	2.1a	
	So, Po(5) is a suitable model	dB1	1.2	Dependent on first E mark 5 must be seen within their answer
	Total	**8**		

Question	Scheme	Marks	AO	Notes
5 (a)	Scatter diagram To see any correlation/relationship /association/trend	E1 E1	1.1 1.1	 oe
5 (b)	0.947 (3 s.f.) There is a strong positive correlation between P [median house price] and E [median gross annual workplace-based earnings]	B1 E1	1.2 2.1a	awrt 0.947 **Actual** 0.94675… Must see **strong** and **positive**
5 (c)	$P = -676000 + 29.3E$ (3 s.f.) **SC**: Coefficients wrt 24000 & 0.0306 scores B1B0 This model would estimate that for every £1 increase in median gross annual workplace-based earnings the median house price would increase by '£29.30'	B1 B1 E1ft	1.2 1.2 2.1a	awrt –676000 & 29.3 **Actual:** a **or** b correct a = –675702.2206 b = 29.3163758 Correct variables **and** form $P = a + bE$ Condone 'we would expect' Condone b instead of '£29.30'
5 (d)	The data is based on regions of England and so may not be appropriate for Wales	E1	3.1a	oe

5 (e)	**Any 2 from** • House prices could influence salaries as much as salaries influence house prices • Supply of housing/particular type of housing may be restricted in a region • There may be high demand for housing due to high proportion of young people/increasing population/high number of immigrants • Existence of highly paid industry in the region (eg. IT, financial services, pharmaceuticals) • High levels of structural unemployment in the region leading to low average income • Level of rental accommodation available in the region	E1 E1	3.1b 3.1b	Accept other plausible reasons
	Total	**10**		

Question	Scheme	Marks	AO	Notes
6 (a)	• Colours of sweets are independent • Fixed number of sweets per packet • Only two possibilities green or not green • The probability of green sweets does not change over time	E1,E1	2.1a	E1 for each correct (total of 2)
6 (b) (i)	$P(G \leqslant 16) = 0.68387...$	B1	1.2	Awrt 0.684
6 (b) (ii)	$P(G = 6) = 0.0017708...$	B1	1.2	Awrt 0.00177
6 (b) (iii)	$P(9 < G < 20) = P(G \leqslant 19) - P(G \leqslant 9)$ $= 0.91519\ldots - 0.040231\ldots$	M1	1.2	
	$= 0.8749...$	A1	1.2	Awrt 0.875 gets 2/2
6 (c)	$\mu = 50 \times 0.3 = 15$	B1	1.2	
	$\sigma^2 = 15 \times 0.7 = 10.5$	B1	1.2	
	Mean is similar to expected but variance is considerably larger than expected	M1	2.1a	Comparisons of both mean **and** variance
	Jaymini's conclusion could be valid **and** should be investigated further.	A1	3.1a	
	Total	**10**		

Question	Scheme	Marks	AO	Notes
7 (a) (i)	Purpose is to ensure no bias exists in the assignment of an infant to a treatment Avoiding situations such as similar infants /sicker infants being allocated the olive oil/parents specifying what treatment they want will be avoided	E1 E1	3.1a 3.1a	Idea of no bias **Alt:** Every infant has an equally likely chance of getting either treatment
7 (a) (ii)	Both parents and researchers don't know which group they have been allocated to Ensures equal treatment and avoids prior expectations invalidating results	E1 E1	3.1a 3.1a	
7 (a) (iii)	Can compare the results of the olive oil against the results of the placebo To see if the differences can be ascribed to the olive oil specifically as the type of oil is the only difference in treatment	E1 E1	3.1a 3.1a	Comparison between olive oil and placebo Effect the olive oil has rather than placebo effect
7 (b) (i)	To give a baseline to work from which allows you to assess the effects of the treatment	E1	2.1a	Starting point for comparison
7 (b) (ii)	To give an indication of any adverse effects of the treatment so that subjects could be withdrawn from the trial	E1	2.1a	**Accept:** A measure of comparison at the **halfway** mark
7 (b) (iii)	This allows for an evaluation of the effectiveness of the treatment at the end of the trial so that conclusions can be drawn	E1	2.1a	Must draw conclusions

7 (c)	Abdominal massage with either olive oil or with paraffin oil is effective in reducing the symptoms of infant colic • Olive oil produced a greater decrease in mean crying duration • Olive oil produced a greater decrease in mean number of episodes of crying than use of paraffin oil. • Crying intensity showed no significant difference between the 2 groups Overall, massage with olive oil is more effective than with paraffin oil.	E1 E1 dE1	 3.1b 3.1b 3.1b	Massage with both oils is effective E1 for any correct **two** Dependent on second E mark
	Total	**12**		

Question	Scheme	Marks	AO	Notes
8 (a)	Let A represent the person has the disease Let B represent the test gives a positive result			Converting information into probabilities
	$A = 0.001 \quad A' = 0.999$ $\mathrm{P}(B \mid A) = 0.95 \quad \mathrm{P}(B \mid A') = 0.02$	B1 B1	2.1a 2.1a	B1 any two correct B2 all four correct Can be seen on tree diagram
	$\mathrm{P}\left(\dfrac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B \mid A)\mathrm{P}(A) + \mathrm{P}(B \mid A')\mathrm{P}(A')}\right)$	M1	1.2	Must see Bayes formula, or can be assumed by a correct substitution
	$= \mathrm{P}\left(\dfrac{0.95 \times 0.001}{0.95 \times 0.001 + 0.02 \times 0.999}\right)$	M1	1.2	
	$= 0.0454$	A1	1.2	
8 (b)	Only 4.54% of positive responses come from cases in which the disease is present	E1ft	2.1b	Accept answers which follow from their answer to (a)
	The test is unsatisfactory	E1ft	2.1b	Follows first E mark
	Total	**7**		

Question	Scheme	Marks	AO	Notes
9 (a)	Bell shaped curve (accept symmetrical about the mean) Asymptotic $\frac{2}{3}$ of observations lie within $\mu \pm \sigma$ 95% of observations lie within $\mu \pm 2\sigma$ 99.7% of observations lie within $\mu \pm 3\sigma$	 E1 E1	 2.1a 2.1a	Statements including standard deviation must include reference to the mean Any feature up to 2 marks
9 (b)	$\mathrm{P}(X < 28) = 0.98609...$	B1	1.2	Awrt 0.986
9 (c) (i) **See below for alternatives**	$\mu = 22.5 - 22.5 = 0$	B1	1.2	Can be seen in new distribution $\sim \mathrm{N}(0,\ \sigma^2)$
	$\sigma^2 = 2 \times 2.5^2 = \frac{25}{2}$	B1	1.2	Allow $\sigma = \frac{5}{\sqrt{2}}$
	$(T_2 - T_3) \sim \mathrm{N}(0,\ 12.5)$	M1	1.2	
	$\mathrm{P}(T_2 - T_3 > 5) = 0.078649...$ (Tables gives 0.0793)	A1	1.2	Awfw 0.0786 – 0.0793
9 (c) (ii)	$\mathrm{P}\left(T_1 > \frac{T_1 + T_2 + T_3 + T_4}{4} + 2\right)$ $\Rightarrow \mathrm{P}(3T_1 - T_2 - T_3 - T_4 > 8)$	M1	2.1a	
	$\mu = 3 \times 22.5 - 22.5 - 22.5 - 22.5 = 0$	B1	1.2	Can be seen in new distribution $\sim \mathrm{N}(0,\ \sigma^2)$
	$\sigma^2 = 9 \times 2.5^2 + 2.5^2 + 2.5^2 + 2.5^2 = 75$	M1 A1	1.2 1.2	
	$3T_1 - T_2 - T_3 - T_4 \sim N(0,75)$	M1	1.2	
	$\mathrm{P}(3T_1 - T_2 - T_3 - T_4 > 8) = 0.1778$ (Tables gives 0.1788)	A1	1.2	Awfw 0.1778 – 0.1790
	Total	**13**		

Question	Scheme	Marks	AO	Notes
Alt 1 **9 (c) (i)**	$\mu = 5$	B1	1.2	Can be seen in new distribution $\sim \mathrm{N}(5,\ \sigma^2)$
	$\sigma^2 = 2\times 2.5^2 = \frac{25}{2}$	B1	1.2	Allow $\sigma = \frac{5}{\sqrt{2}}$
	$Y \sim \mathrm{N}(5,\ 12.5)$	M1	1.2	
	$\mathrm{P}(Y \leq 0) = 0.078649\ldots$ (Tables gives 0.0793)	A1	1.2	Awfw 0.0786 – 0.0793
Alt 2 **9 (c) (i)**	$\mu = -5$	B1	1.2	Can be seen in new distribution $\sim \mathrm{N}(-5,\ \sigma^2)$
	$\sigma^2 = 2\times 2.5^2 = \frac{25}{2}$	B1	1.2	Allow $\sigma = \frac{5}{\sqrt{2}}$
	$Y \sim \mathrm{N}(-5,\ 12.5)$	M1	1.2	
	$\mathrm{P}(Y \geq 0) = 0.078649\ldots$ (Tables gives 0.0793)	A1	1.2	Awfw 0.0786 – 0.0793
	Total	**13**		

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom