Mark Scheme (Results)

Summer 2019

Pearson Edexcel In GCE Statistics

Paper 9ST0_01

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 2019
Publications Code 9ST0_01_1906_MS

General Marking Guidance

- 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.
- Crossed out work should be marked UNLESS the candidate has replaced it with an alternative response.

EDEXCEL GCE Statistics

General Instructions for Marking

1. The total number of marks for the paper is 75.

2. The Edexcel Mathematics mark schemes use the following types of marks:

- **M** marks: method marks are awarded for 'knowing a method and attempting to apply it', unless otherwise indicated.
- **A** marks: Accuracy marks can only be awarded if the relevant method (M) marks have been earned.
- **B** marks are unconditional accuracy marks (independent of M marks)
- Marks should not be subdivided.

3. Abbreviations

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

- bod – benefit of doubt
- ft – follow through
- the symbol $\sqrt{}$ will be used for correct ft
- cao – correct answer only
- cso - correct solution only. There must be no errors in this part of the question to obtain this mark
- isw – ignore subsequent working
- awrt – answers which round to
- SC: special case
- oe – or equivalent (and appropriate)
- dep – dependent
- indep – independent
- dp decimal places
- sf significant figures
- $\ast$ The answer is printed on the paper
- ⌈ The second mark is dependent on gaining the first mark

4. All A marks are 'correct answer only' (cao.), unless shown, for example, as A1 ft to indicate that previous wrong working is to be followed through. After a misread however, the subsequent A marks affected are treated as A ft, but manifestly absurd answers should never be awarded A marks.

Paper 1 mark scheme

Question	Scheme	Marks	AO	Notes
1(a)	**Example sampling method**			
	Eva should look at the membership database/list.	E1	1.1	Defining their population as all members, or view a list
	She should produce queries to produce two lists, members aged 60 and over, and members aged under 60, using their dates of birth. **or** She should divide the list into members aged 60 and over, and members aged under 60.	E1	1.1	Clear division on age
	She should select 25 people from each list…	E1	1.1	Clear attempt to get **precisely 25** people in the ≥ 60 category (must mention 60)
	…by numbering each list and using a **random number** generator.	E1	1.1	Any sensible sampling method to select 25 people from a list.
1(b)	If 50 questionnaires are sent out, it is very unlikely that she will get 50 responses.	E1	3.1a	oe Likely responses < 50

Question	Scheme	Marks	AO	Notes
1(c)	Send out lots more than 50 questionnaires…	E1	3.1a	> 50 questionnaires
	…and stop analysing responses when Eva has 25 responses from those aged 60 and over, and 25 responses from those aged under 60. **or** …and choose 25 responses from each category using a random process.	E1	3.1a	
	Alternative			
	Ask the chosen members to complete the questionnaire next time they come into the gym…	(E1)		
	…and supervise them as they complete it.	(E1)		Some measure to stop the questionnaire from being taken away.
	Total	**7**		

Question	Scheme	Marks	AO	Notes
2(a)	Dirk could have used a spreadsheet program.	E1	1.1	If mention database functions must mention query
2(b)	$r = 0.967$ (3 s.f.)	B1	1.2	awrt **Actual:** 0.967377…
	There is a strong positive correlation between the number of human births and the number of domesticated chickens.	E1	2.1a	Must see **strong** and **positive**
2(c)	$D = 21\ 100 + 0.0872H$ (3 s.f.)	B1	1.2	awrt 21000 & 0.087 **Actual:** $a = 21\ 061.829\ldots$ $b = 0.087244\ldots$
		B1	1.2	Correct variables and form $D = a + bH$
	SC: Coefficients awrt -177 000 & 10.7 scores B1B0			
	The model would estimate the mean number of domesticated chickens for a country with no human births to be 21 100 [in 2014].	E1	2.1a	Condone omission of **mean** Condone 'we would expect…' oe Condone a instead of 21100
	The model would estimate the mean number of domesticated chickens to increase by 0.0872 for each additional human birth [in 2014].	E1	2.1a	Condone omission of **mean** Condone 'we would expect…' oe Condone b instead of 0.0872 FT their a & b, SC a & b reversed max E1

Question	Scheme	Marks	AO	Notes
2(d)	**Possible explanations** The model is being used to extrapolate. The regression equation will estimate the **mean** number of chickens for countries with 399 human births (not a prediction). The model would estimate far too many chickens for a small population.	E1	3.1a	Any sensible mention of **extrapolation** oe Wrong type of regression equation for predicting a value.
2(e)	Correlation does not imply causation.	E1	3.1b	oe
	It is likely that countries with a high population have both a higher birth rate and a higher population of chickens for food.	E1	3.1b	Connection to third variable, population. Implication that correlation is positive between all variables.
	Alternative Sensible explanation of how one of the factors may influence the other Done clearly in context	(E1) (E1)		
	Total	**10**		

Question	Scheme	Marks	AO	Notes
3(a)(i)	[X = number of page views in a one-minute period]			
	$X \sim Po(2.8)$	M1	2.1a	Poisson distribution used PI
	$P(X = 2) = 0.238$ (3 s.f.)	A1	1.2	awrt
3(a)(ii)	[Y = number of page views in a five-minute period]			
	$Y \sim Po(14)$	M1	1.2	$\lambda = 14$ PI
	$P(Y > 20) = 0.0479$ (3 s.f.)	A1	1.2	awrt 0.048
	NOTE: 0.952 [$P(Y \leq 20)$] or 0.0765 [$P(Y \geq 20)$] seen scores M1A0			

3(b)	**Exponential method** [W = time between two consecutive page views (mins)]			
	$W \sim Exp(2.8)$	B1	2.1a	PI **Alt:** Time in secs & $W \sim Exp\left(\frac{7}{150}\right)$ awrt 0.0467
	$2{:}25 = 2\frac{5}{12} = \frac{29}{12} = 2.417$ (4 s.f.) mins	B1	1.2	Conversion from m:ss to minutes **Alt:** 145 secs
	$P\left(W > \frac{29}{12}\right) = e^{-2.8\times\frac{29}{12}}$	M1	1.2	PI Use of exponential cdf formula **Alt:** $P(W > 145) = e^{-\frac{7}{150}\times 145}$
	$= 0.001152$ (3 s.f.)	A1	1.2	$= 1.52 \times 10^{-3}$ awfw 0.001150~0.001154
	NOTE: $P(W > 2.25) = e^{-2.8\times 2.25} = 0.0018(363)$ scores B1B0M1A0			

Question	Scheme	Marks	AO	Notes
3(b) continued	**Poisson method**			
	$2:25 = 2\frac{5}{12} = \frac{29}{12} = 2.417$ (4 s.f.) mins	(B1)		Conversion from m:ss to minutes **Alt:** 145 secs
	[V = number of page views in a 2.417 minute (145 second) period]			
	$V \sim Po\left(2.8 \times \frac{29}{12}\right) = Po\left(\frac{174}{25}\right) = Po(6.96)$	(B1ft)		Using their time in mins/secs PI
	$P(V = 0) = e^{-6.96}\frac{(6.96)^0}{0!}$	(M1)		Use of Poisson pdf formula PI
	$= 0.00152$ (3 s.f.)	A1	1.2	$= 1.52 \times 10^{-3}$ awfw 0.00150~0.00154
	NOTE: $2\frac{1}{4} \times 2.8 = 6.3, P(V = 0) = 0.0018(363)$ scores B1B0M1A0			

Question	Scheme	Marks	AO	Notes
3(c)	It does not support Rhodri's suspicion.	E1	2.1b	
	We would expect a pause that long by chance regularly.	E1	2.1a	
	Less than 1 in 1000 (1 in 868) pauses would be of this length on average, and we would expect several thousand (4032) pauses each day. **or** We would expect 4.65 pauses of this length every day, on average. **or** We would expect a pause of this length every 5:10 (or 5.17) hours, on average.	E1	2.1b	Any sound numerical reasoning to back up the argument. Probabilities & figures may be approximate FT calculations using their answer to part (b)
	SC: 'It does support Rhodri's suspicion as the probability is so low' scores E1E1E0			
3(d)	The data is showing a weekly seasonality.	E1	2.1b	oe (e.g. trends, oscillates, dips)
	This may be because fewer people are studying at the weekend, and so will be less likely to be looking for maths articles.	E1	2.1a	Any sensible explanation. Must be in context.
3(e)	The data is showing short-term variation. **or** There are fewer page views in the summer months.	E1	2.1b	oe Accept 'annual seasonality'.
	This may be because fewer people are studying in summer, as it lies between academic years, and so will be less likely to be looking for maths articles.	E1	2.1a	Any sensible explanation. Must be in context.

Question	Scheme	Marks	AO	Notes
3(f)	**Possible explanations** There may have been a recording error in the data. A MOOC may have run where students were asked to look at the website. A bot may have repeatedly accessed the website on one day. Page may be trending on Wikipedia frontpage Poisson distribution may have been mentioned in a news article Lots of people learning about it at school/university	E1, E1	2.1a, 2.1a	E1 for each sensible explanation (max E2)
3(g)	**Possible reasons** The data clearly trends (weekly and annually), so the rate will not be constant. The mean rate may not be accurate, as the outlier described in (c)(iii) may be inflating its value. Page views may not be independent, as one person may suggest another person visit the website.	E1, E1	3.1a, 3.1a	Or e.g. peak in Jan shows not indep E1 for each reason (max E2)
	Total	**19**		

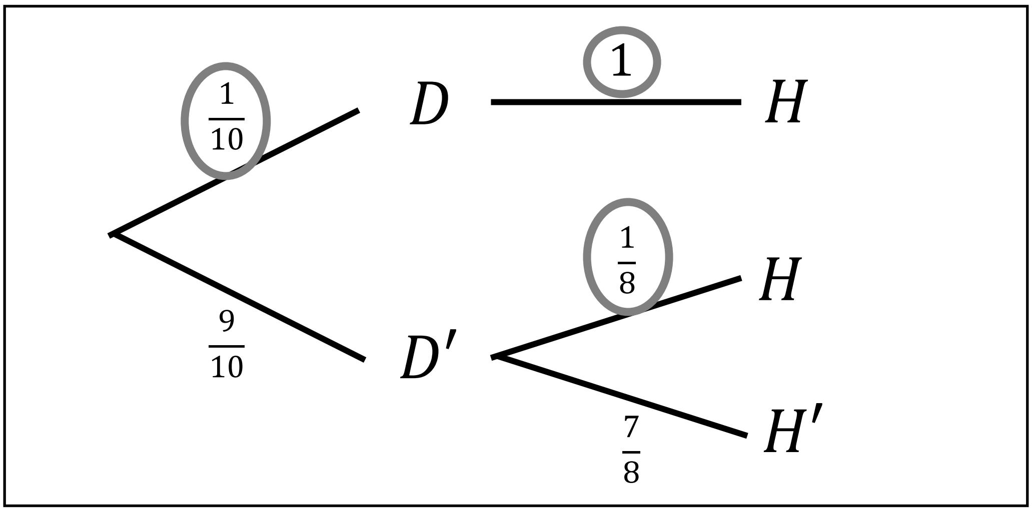
Question	Scheme	Marks	AO	Notes
4	**Example experiment**			
	Kayoko should select a **large** sample of students.			Large (30+) **Accept** simple random **Accept** sensible stratification, provided no hint of disproportionality
	The students should be blocked by GCSE maths grade.			Relevant blocking **Examples** • Level of education • Sex • Age • Subject
	Half given caffeine and one half not given caffeine.			Control group **Alt:** Levels of treatment, but must include control group
	The subjects should each be tested with a mental arithmetic test before the treatment.			Test before treatment (as well as after) or a matched pairs experiment where similar students are compared
	Tests of equivalent difficulty or same test given to control group and main group or 2nd test if paired			
	The subjects should each be given an identical-looking soft drink, which may or may not contain caffeine.			Blind or double-blind **Alt examples:** • Coffee and decaf coffee • Caffeine and placebo tablet
	Wait for caffeine to take effect or for it to run out			

		E1, E1, E1, E1, E1	1.1	1 mark per reason
	Total	**5**		

Question	Scheme	Marks	AO	Notes
5(a)	Run a query with a join between the Car_ID fields in the two tables.	E1	1.1	‘query’ seen
		E1	1.1	‘join’ seen Condone ‘link’
		E1	1.1	Description that the join should go between the two Car_ID fields
	Alternative 1 Transfer tables to a spreadsheet program…	(E1)		
	…and write a macro to check for each Car_ID in both tables (and produce a new table with combined data). **or** …and manually check for each Car_ID in both tables (and copy and paste to make a new table with combined data).	(E1)		Or any viable alternative. **NOTE:** Alternative solutions score E2 max
	Alternative 2 Any other method that would work	(E1)		e.g. entering all data manually into the database
5(b)(i)	The data is not symmetric. **or** The right-hand tail is longer than the left-hand tail.	E1	2.1b	oe
5(b)(ii)	The data is roughly bell-shaped.	E1	2.1b	oe

Question	Scheme	Marks	AO	Notes
5(c)	[C = Maintenance costs for a car $C \sim N(511.36, 168.65^2)$ V = Maintenance costs for a van $V \sim N(885.12, 232.78^2)$]			
	$X = \sum_{i=1}^{10} C_i + \sum_{i=1}^{4} V_i$	B1	1.2	**or** $C + C + \ldots + V + V\ldots$ PI **Condone** $10C + 4V$
	$\mu = 10 \times 511.36 + 4 \times 785.12$ = 8654 (0 d.p.)	B1	1.2	awrt
	$\sigma = \sqrt{10 \times 168.65^2 + 4 \times 192.78^2}$	M1	1.2	Attempt to find σ or σ^2
	= 708 (0 d.p.)	A1	1.2	$\sigma =$ awrt 708 **or** $\sigma^2 =$ awrt 501 000
	So $X \sim N(8654, 708^2)$	M1	1.2	Total costs normally distributed with their μ and σ above PI
	$P(X > 10000) = 0.0286$ (3 s.f.) **Note:** 0.242 implies B1B1M1M1	A1	1.2	awfw 0.028~0.029

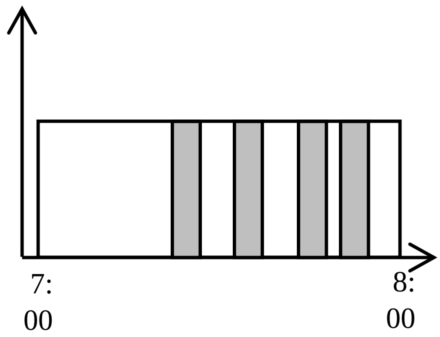
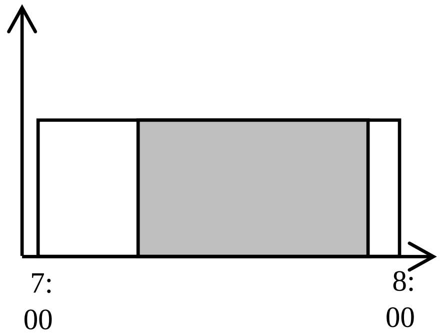
Question	Scheme	Marks	AO	Notes
5(d)	**Possible reasons** The cars and vans will be brand new, so they will not be a random sample. Some of the cars are likely to be the same model, so they will not be a random sample. Inflation is likely to have made maintenance costs higher since 2018. Maintenance costs may be unusually high/low in the area of the new branch. Different brands of cars/vans may have different repaid costs Figures may not have been recorded reliably.	E1, E1	3.1a, 3.1a	E1 for each reason (Max E2)
	Total	**13**		

Question	Scheme	Marks	AO	Notes
6	[D = The coin is a double-header H = Three heads on coin tosses]			
				
	$P(D) = \frac{1}{10}$	B1	1.2	PI
	$P(H\|D) = 1$	B1	1.2	PI
	$P(H\|D') = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$	B1	1.2	PI
	Using Bayes' theorem (or from tree diagram),			
	$P(D\|H) = \frac{1\times\frac{1}{10}}{1\times\frac{1}{10}+\frac{1}{8}\times\frac{9}{10}} = \frac{\frac{1}{10}}{\frac{1}{10}+\frac{9}{80}}$	M1	1.2	Clear attempt at Bayes' theorem PI
	$= \frac{8}{17} = 0.471$ (to 3 s.f.)	A1	1.2	awrt
	Total	**5**		

Question	Scheme	Marks	AO	Notes
7(a)	Number of correct cards $X \sim B\left(1850, \frac{1}{5}\right)$			
	Approximate to $Y \sim N(370, 296)$	B1	2.1b	Normal approx. stated or clearly used
		B1	1.2	Mean = 370 cao
		B1	1.2	Variance = 296 cao or SD = awrt 17.2
	$P(X \geq 558) \approx P(Y > 557.5)$	M1	1.2	**PI** Use of 557.5 (or 558) **Alt** z = awrt 10.9
	$= 5.88 \times 10^{-28}$	A1	1.2	**Actual:** 5.87695×10^{-28} (awrt 5.9) **no continuity correction:** 4.27×10^{-28} (awrt 4.3) **Binomial prob**: Scores max of M1A1 for 2.19×10^{-25} (awrt 2.2)
7(b)	**Dice**			
	It is approximately…	E1	2.1b	Indication that the probability is not exact. • Accept 'more than' or 'less than'
	…the probability of rolling a die 31 times and getting a six every time. **or**	B1	1.2	• 30~32 seen • Number on die specified

	…the probability of rolling 31 dice and all land on a six.	E1	2.1b	**Alt:** If single die used, accept ‘correct guesses’ oe **Alt:** 31~33 dice with all landing on ‘same number’ Clear implication that the die is **only** rolled 30~32 times (or **only** 30~32 dice are rolled) **e.g.** ‘31 sixes in a row’ scores E0
	Lottery It is less likely…	(E1)		or ‘lower probability’ oe
	…than winning the Lotto jackpot from a single play on each of three specific draws.	(B1)		Three draws specifically. $\left(\frac{1}{14000000}\right)^3 =$ 3.64×10^{-22} **Do not accept** ‘winning three times in one draw’ oe
		(E1)		• ‘single play’ or ‘single set of numbers’ specified e.g. ‘winning on three Lotto draws’ scores E0 • Draws must be **specific**, not general e.g. ‘winning three draws in a row’ scores E0
7(c)	I disagree with the psychologist’s conclusion…	E1	3.1b	

	Possible reasons …as Hubert Pearce was not supervised, and so he could have cheated. …as the experiment was not run under properly controlled conditions.	E1	3.1a	Correct criticism of the experiment
	Alternative I agree with the psychologist as the probability of this happening by chance is so tiny.	(E1)		This solution scores E1 only
	Total	**10**		

Question	Scheme	Marks	AO	Notes
8(a)(i)		M1	1.1	Rectangular distribution diagram with (a least one) train journey clearly indicated **or** 4 trains between 7:00 and 8:00 **or** $\frac{n\times 5}{60}$ **or** 20 (mins) seen
	$\frac{4\times 5}{60}=\frac{1}{3}$	A1*	1.2	oe Must be convinced. No gaps in argument for full marks.
8(a)(ii)		M1	1.1	Rectangular distribution diagram with single block clearly indicated
	$\frac{38}{60}=\frac{19}{30}=0.633$ (3 s.f.)	A1	1.2	awrt
	Alternative $4\times 10=40$ mins			
	-2 mins overlap	(M1)		Clear indication that overlap has been considered **or** 37/38 (mins) seen.
	$\frac{38}{60}=\frac{19}{30}=0.633$ (3 s.f.)	(A1)		Awrt
8(b)	[Let X = Number of days the passenger has to wait for less than 5 minutes]			

	$X \sim B\left(5, \frac{1}{3}\right)$	M1ft	2.1a	or $X \sim B\left(5, \frac{2}{3}\right)$ if X defined using 'more than 5 minutes' ft their p in (a)(i)
	$P(X \geq 2) = \frac{131}{243} = 0.539$ (3 s.f.)	A1	1.2	awrt
8(c)	The passenger waits less than 10 minutes if arriving during 7:15-7:53	M1ft	1.2	PI **Note:** This mark may be awarded if seen in (a)(ii) Ft their times from (a)(ii)
	$P(15 \leq T \leq 53)$	M1	1.2	PI Accept clearly labelled diagram with correct shaded region.
	$= 0.922$ (3 s.f.)	A1	1.2	awfw 0.922~0.925
	Note: True probability ≈ 0.924 It is anticipated that candidates will discard other time periods due to being outside $\mu \pm\ 3\sigma$			

8(d)	**Possible assumptions** Comment on validity **The trains are all on time** This is unlikely to be valid as lots of trains are late in the uk **There is only one train route from Godalming to London** Validity would depend on the setup of the railway lines	**E1** E1dep	**2.1a** 3.1a	**or** this is likely to be valid as most trains are on time **or** this is valid as Godalming only has a single line through it. **Correct assumption** Relevant comment on validity. Dep on previous E1
	Total	**11**		

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