Mark Scheme (Results)

Summer 2022

Pearson Edexcel GCE
In Statistics (9ST0)
Paper 03: Statistics in Practice

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 2022
Question Paper Log Number 69434
Publications Code 9ST0_03_2206_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)	The group that only receives the **standard** balance training.	B1	1.1	Accept “the second group”
1(b)	So that any difference (in outcome) found between the two groups can be more confidently attributed to the video game.			oe Accept “To reduce the likelihood that some factor other than the video game results in a difference (in outcome) between the two groups” or “Reduces experimental error”
	To avoid the **bias** that could arise if patients were not assigned randomly.			oe Reference specifically to removing **bias**.
		E1	3.1a	Either Do not accept “to make the experiment more fair” or similar.
1(c)	Because Giovanni knows which patients are assigned to each group.			oe
	Because the patients know if they are playing the video game or not			
		B1	1.1	Either

Question	Scheme	Marks	AO	Notes
1(d)	**List not exhaustive**			
	Both the control and experimental groups **show improvement** in their (average Berg balance) scores at the **end** (T1) of the experiment.			Accept "Both groups have better (average) balance at the end of the experiment"
	Both the control and experimental groups **show improvement** on their **initial** (average Berg balance) scores **one month after** (T2) the experiment.			Must specify that this is an improvement on **initial** (average) scores or balance.
	The **control group's improvement** in average (Berg balance) score at the end of the experiment **does not remain** one month after (T2) the experiment			Accept "The control group's better average balance is not maintained one month after the experiment"
	The **experimental group's improvement** in average (Berg balance) score at the end of the experiment **was maintained** one month after (T2) the experiment			
	The control and experimental groups have **different distributions** of (Berg balance) scores at the **start** (T0) of the experiment.			This is **not** distinct from comments on the differences in spread or average of the scores at the start of the experiment. (41 vs 45)
	The **control group** has a slightly **lower average** (Berg balance) score at the **start** (T0) of the experiment than the experimental group.			Accept "The control and experimental groups have similar average scores at the start of the experiment."
	The **experimental group** has a higher median balance score than the control group at **all stages**			

Question	Scheme	Marks	AO	Notes
1(d) cont.	The experimental group has **two outliers**			
	The **experimental group** has a **bigger range** of (Berg balance) scores at the **start** (T0) of the experiment than the control group.			
	The **experimental group** has a **smaller** range of (Berg balance) scores at the **end** (T1) of the experiment than the control group.			
	The **experimental group** has a **smaller IQR** of (Berg balance) scores at the **start** (T0) of the experiment than the control group.			
	Control group **one month after the experiment** (T2) has the **same median** as the experimental group had at the **start** of the experiment (T0)			
	The **spread decreases** in the **experimental group** from the end of the experiment (T1) to one month after the experiment (T2).			
	Experimental group are **positively skewed**			
		E1, E1, E1, E1	1.1, 1.1, 1.1, 1.1	One mark for each **distinct** correct comment up to a maximum of 4 marks Maximum 3 if no context attempted **Accept** median or average but **not** mean throughout – penalise only **once**

Question	Scheme	Marks	AO	Notes
1(e)	**Advantages**			
	The box and whisker plots clearly show the differences between the distributions (of the Berg balance scores) before and after the therapies.			oe
	The box and whisker plots clearly show the differences (of the Berg balance scores) between the distributions of the control and experimental groups.			oe control v experimental or e.g. easy to compare averages
	It can show the scores at 3 different times on the same diagram			
	Uses statistics that are not distorted by outliers and shows outlying observations clearly.			
	Easy to compare groups			
		E1	3.1a	Any advantage

Question	Scheme	Marks	AO	Notes
1(e) cont.	**Disadvantages**			
	The meaning of box and whisker plots may need to be explained to a non-specialist audience.			oe Specialist knowledge is required to understand box and whisker plots.
	The box and whisker plot only provides 5 statistics.			oe Limits the information provided
	Doesn't tell us how large the sample was			
	Exact values are difficult to read			
		E1	3.1a	Any disadvantage
	Total	**9**		

Question	Scheme	Marks	AO	Notes
2(a)	$\frac{1}{(13-5)} = 0.125$	E1	2.1a	oe May be stated in words: "Total probability/area is one. The base of the rectangle is 8. One divided by 8 is 0.125" Accept "1/8 = 0.125"
2(b)	9 (working) hours	B1	1.2	
2(c)	0	B1	2.1a	oe "zero"
2(d)	$((7-5) \times 0.125 =)\ 0.25$	B1	1.2	
2(e)	$\frac{3 \times 0.125}{1 - "0.25"}$	M1ft	1.2	PI 0.375/0.75 ft on candidate's 0.25 Other appropriate methods allowed.
	0.5	A1	1.2	oe
	Total	**6**		

Question	Scheme	Marks	AO	Notes
3(a)	(n = 22) t = 2.080	B1	1.3	t-value Condone z=1.96
	$54.8 \pm 2.080\left(\frac{35.4}{\sqrt{22}}\right)$	M1ft	1.3	Formula correct ft on B1
	(39.1, 70.5)	A1	1.3	awfw 39~39.1, 70.4~70.5
3(b)	The confidence intervals do not overlap.	M1	2.1b	Comparison of candidates CI from (a) with given CI
	So there is significant evidence to support Klazine's belief.	E1dep	2.1b	oe dep M1 Evidence supports Klazine's belief
	Involving a child in preparing their own meal affects what they choose to eat at that meal (this can be seen for salad)	E1dep	2.1a	dep on M1 Response can be more specific, e.g. "children who help prepare their own meal eat more salad" or "children who do not prepare their own meal eat less salad"
				Award E marks independently from one another. May be seen in one sentence. Disregard references to health.

Question	Scheme	Marks	AO	Notes
3(c)	C signifies child prepared with parent D signifies parent prepared alone			Can be the other way around with a negative ts and cv
	H_0: $\mu_C - \mu_D = 10$ H_1: $\mu_C - \mu_D > 10$	B1	1.3	oe both, subscripts clearly defined
	$s_p^2 = \dfrac{(25-1)50.1^2 + (22-1)51.3^2}{25+22-2}$	M1	1.3	PI
	= 2567	A1	1.3	PI awrt 2560~2570 or s_p = awfw 50.5~50.7
	$ts = \dfrac{(110.5-89.7)-10}{\sqrt{2567\left(\frac{1}{25}+\frac{1}{22}\right)}}$	M1	1.3	PI Numerator or denominator correct (ignoring -10)
		M1dep	1.3	PI Fully correct numerator with candidates s_p^2
	Critical value method: $10 + 1.679 \times \sqrt{219.344} = 34.866$			1.679 scores M1 Formula scores M1dep
	= 0.729	A1	1.3	ts awfw 0.728~0.730 or 34.9
	cv = 1.679 or p-value = 0.235	B1	1.3	Any of 1.679, 0.235, 20.8 selected for comparison
	(0.729<1.679, No significant evidence to reject H_0) There is insufficient evidence that the children who prepared their meal with a parent ate over 10 grams more cauliflower, on average, than the children who did not.	E1dep	2.1a	oe 0.235>0.05 or 20.8<34.9 in context. Must contain element of doubt. dep on whole test correct

Question	Scheme	Marks	AO	Notes
3(d)	The (population) **variance** of the **weight** of cauliflower **eaten** by **children who prepare their meal with a parent** should be very **similar** to the (population) **variance** of the **weight** of cauliflower **eaten** by **children whose parent prepared their meal alone.**	B1	1.3	oe Variances equal/similar/close in context. Accept "Variance of cauliflower should be the same for children who prepare and those that don't."
	(The sample variances/sds are very close so) there is no evidence that the assumption is not valid	E1dep	2.1b	oe Clear indication that the assumption is valid. Dep "variances similar" assumption stated. Not dep gaining B1 mark
3(e)	The t-distribution techniques used in (a) and (c) can be (validly) applied to small samples.			oe Accept "t-test can be used with small samples"
	Populations might not be normally distributed (and n is not large enough for CLT to apply)			Do not accept CLT doesn't apply by itself
		E1	3.1a	Either
	Using larger sample sizes will increase the power of the test (to detect a given difference in means).			oe Accept "large sample means more likely to find a small difference" or "large sample means less likely to make a Type II error" or "more reliable"
	Would use z instead of t			
		E1	3.1a	Either

3(f)	The width of the confidence intervals should decrease.	B1	2.1b	oe Confidence intervals narrower. Condone smaller
	Total	**19**		

Question	Scheme	Marks	AO	Notes
4(a)	$\left(\sqrt{2.8} =\right) 1.67$	B1	1.2	awrt 1.67
4(b)	0.222	B1	1.2	awfw 0.222~0.223
4(c)	$\lambda = 8.4$	M1	1.2	PI rescaling
	$(1 - 0.399 =)\ 0.601$	A1	1.2	awfw 0.600~0.602
4(d)	Exponential (distribution)	B1	2.1b	
	with parameter 2.8	B1dep	2.1b	$\lambda = 2.8$ **or** **mean** = 1/2.8 = 0.357 *must* state that this is the mean Dep on previous B1
				Accept Exp(2.8) for both marks
4(e)	$\left(\frac{1}{2.8} =\right) = \frac{5}{14} = 0.357$ (years)	B1	1.2	oe

Question	Scheme	Marks	AO	Notes
4(f)	Use of memoryless property	M1	2.1a	PI by correct working May be stated or demonstrated by candidate clearly disregarding the wind turbine history e.g. "P(X < 0.5)" with no conditional probability used.
	$1 - e^{-2.8\times 0.5}$	M1ft	1.2	PI $1 - e^{-1.4}$ Accept ft of candidate's λ
	0.753	A1	1.2	awrt 0.753
	Alternative			
	Use of Poisson distribution with λ = 1.4	(M1)		PI
	1 – $P(X = 0)$	(M1)		PI 1 – 0.247
	0.753	(A1)		awrt 0.753

Question	Scheme	Marks	AO	Notes
4(g)	$\frac{84}{6 \times 5} = 2.8$	B1	1.2	oe working Full calculation must be demonstrated. May state in words.
4(h)	A wind turbine may not fail at a constant average rate as older turbines may be more likely to fail.			Challenge to the assumption that failure rate for a wind turbine is constant.
	The failure of a wind turbine may not be a random event because it may be caused by weather.			Challenge to the assumption that wind turbines fail at random.
	The failure of a particular wind turbine may not be independent of the failure of another wind turbine as one may fall onto another.			Challenge to the assumption that wind turbines fail independently.
	Petra's calculation of λ made the assumption that all of the wind turbines fail at the same average rate per year. This might not be true.			Challenge to the assumption that **all wind turbines fail** at the **same rate**.
	Petra based her value of 2.8 on a limited amount of data			Challenging Petra's value of 2.8 e.g. When a wind turbine has failed it can't fail again until repaired
	More than one turbine could fail at exactly the same time			
	There would be an upper limit to the number of failures in one year			
		E1, E1, E1	3.1b, 3.1b, 3.1b	Any three **distinct** answers from the above, in context of wind turbine failure. Max E1E0E0 if no context
	Total	**14**		

Question	Scheme	Marks	AO	Notes
5(a)	(The distribution has an approximate) bell shape	E1	2.1a	oe Accept "(Distribution is) monomodal" **or** "unimodal" **or** "One clear peak (in the distribution)"
5(b)	The distribution has a (positive) skew.			Skew **or** not symmetrical
	Too much of the distribution is **in** the **tails** (for it to be a normal distribution).			oe Accept "(distribution/shape) **too triangular**" or (distribution has) **tails** (that are) **too large**" or "(distribution has) **high kurtosis**"
		E1	2.1a	**Either** May use calculations that show an equivalent argument but the point being made must be clear.

Question	Scheme	Marks	AO	Notes
5(c)	$s = P(80 < X < 100)$ or $t = P(100 < X < 120)$	M1	1.2	PI oe Clear attempt to find either probability using correct normal distribution.
	$s = 0.2789$ or $t = 0.3406$	A1	1.2	**Either** *s* or *t* correct *s* awfw 0.278~0.280 t awfw 0.340~0.342
	$250 \times s$ or $250 \times t$	M1ft	1.2	Either PI ft candidate's *s* or *t*
	$u = 69.73$ and $v = 85.15$ or 85.14	A1	1.2	**Both** *u* and *v* correct *u* awfw 69.5~70 *v* awfw 85~85.5 Correct to 1 or more dp
5(d)	They should pool $140 < x \leq 160$ with $160 < x$ because the expected frequency is less than 5.	E1	1.3	oe Accept "pooling because $1.7 < 5$"

Question	Scheme	Marks	AO	Notes
5(e)	H_0: The normal distribution is a suitable model H_1: The normal distribution is not a suitable model	B1	1.3	oe Hypotheses, both
	df = 3 cv = 7.815 or *p*-value = 0.00455	B1	1.3	cv or awfw *p*-value = 0.004~0.005
	($13.04 > 7.815$ or $0.00455 < 0.05$) Reject H_0	M1dep	2.1b	Comparison dep B1
	There is sufficient evidence to conclude that the normal distribution is **not** a suitable model for film running times.	E1dep	2.1a	Correct conclusion in context dep B1M1
	Total	**11**		

Question	Scheme	Marks	AO	Notes
6(a)	(Sign test)	B1	2.1b	PI sign test clearly used or stated Implied by use of binomial e.g. $P(X \leq 4)$
	[X = number of rounds with more blue winners than red winners] $X \sim B(20,0.5)$	B1	2.1a	PI use of B(20,0.5) Condone n=21
	H_0: p = 0.5 H_1: p $\neq$ 0.5	B1	1.3	oe Condone 1-tail
	$P(X \leq 4)$	M1	1.3	oe PI Attempt to calculate p-value Accept $P(X \geq 16)$ Or Attempt to find critical region
	= 0.00591 or Critical region is $X \leq 5$ as p=0.021	A1	1.3	awrt 0.0059
	0.00591 < 0.025 or 4<5) Reject H_0	M1dep	2.1b	Comparison Condone compared to 0.05 if 1-tailed hypotheses Dep M1
	There is significant evidence that wearing red affects the success of (male) combatants.	E1dep	2.1a	Dep A1M1 Condone “There appears to be an advantage of wearing red (over blue) for the combatants.”

Question	Scheme	Marks	AO	Notes
6(b)	$H_0: \mu_d = 0$ $H_1: \mu_d \neq 0$	B1	1.3	oe Hypotheses re: population means e.g. $\mu_{red} = \mu_{not\ red}$ Condone 1-tail
	d = 0.5, 0.5, 1, 0.5, 1.5	M1	1.3	PI Attempt at differences (signs may be all negative)
	$\bar{d} = 0.8 \quad s_d = 0.4472$	A1ft	1.3	PI mean and sd of their differences
	$ts = \dfrac{0.8}{\left(\dfrac{0.4472}{\sqrt{5}}\right)}$	M1	1.3	PI calculation of ts may be negative Allow their mean/sd or $0 \pm 2.776 \times \frac{0.4472}{\sqrt{5}}$ which scores the B1 for cv
	t = 4.00	A1	1.3	awrt 4.00 or $\frac{\bar{d}}{\frac{0.4472}{\sqrt{5}}} = 2.776$
	(df = 4) cv = 2.776	B1	1.3	Either correct cv (ignore sign) or p-value or p-value = 0.0161 awfw 0.0159~0.0162 or $\bar{d} = \pm 0.555$ **Condone** 1-tail cv = 2.132 if 1-tailed hypotheses
	4 > 2.776 Reject H_0.	M1dep	2.1b	PI or -4 < -2.776 or 0.0161 < 0.025 or 0.8>0.555 dep A1B1

				Correct comparison
	There is **evidence** at the 5% significance level that the **success** of a football team is **affected** by wearing **red**; (teams appear to do **better** when **wearing red**.)	E1dep	2.1a	oe full explanation in context required for E1 mark dep A1B1

SC Two sample t test max B1M0A0M1A0B1M1E0

$\bar{d}_{red} = 0.4, \bar{d}_{not\ red} = -0.4$

$s_p = 0.65$

t=1.94

Question	Scheme	Marks	AO	Notes
6(c)	**Possible criticisms (not exhaustive)**			
	Both (a) and (b) have very small samples			oe Or either sample small
	There is no evidence of randomisation in (b).			oe Or neither sample was selected at random. "Football teams weren't randomly assigned to wear red or not red"
	Fights in (a) not independent as same fighter may fight in red and in blue			
	Additional factors not taken into account e.g. relative strength of teams, ability of combatant			Blocking factors could be mentioned here
	(a) only considers 2004			
	(b) only considers Europe			
	All combat rounds not equally weighted			
	(a) only tests red against blue…			
	…whereas (b) tests against many different colours, so the two tests are not providing consistent conclusions.			
	The t-test [in (b)] may not be appropriate because the (differences) may not be normally distributed.			oe
		E1, E1, E1, E1	3.1a, 3.1a, 3.1a, 3.1a,	Any four **distinct** correct comments.

Question	Scheme	Marks	AO	Notes
6(d)	The data in (a) and (b) is **only about men's** sport (so the results may not be applicable to women's sport).			oe Accept "all data only about men"
	The tests were about **football and combat** sports so results might **not** extend to **netball**			
	In both (a) and (b) it was found that the players **wearing red did better** (so it may be helpful to Charlottes team also).			oe Accept "tests show playing in red helps"
	Both of the studies in (a) and (b) were **very small** (so it might not be worth paying for red kit without further evidence).			oe reference to limited size of evidence Accept "sample sizes are too small"
	Neither study was conclusive that the effect found was due to wearing red. **Effects could have been due to colours other than red** (so it might not be worth paying for new kit without further evidence).			oe reference to interpretation of evidence Accept "no evidence found of causation/causal effect of red on success"
	The **assumptions** of the tests used in this research **may not be true**(so Charlotte shouldn't spend money on new kit without better evidence).			oe
	The tests in (a) and (b) **were two tailed**, so we only have evidence that red is a difference not an improvement			
		E1, E1	3.1b, 3.1b	Any two distinct comments
SC if concluded that there was no difference in (a) or (b) may earn one mark for a relevant comment				
	Total	**21**		

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