Mark Scheme (Results)

November 2021

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

November 2021
Question Paper Log Number P66657A
Publications Code 9ST0_03_2111_MS
All the material in this publication is copyright
© Pearson Education Ltd 2021

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.

Qu	Scheme	Marks	AO	Notes
1(a)(i)	$\frac{724}{1429} \approx 0.507$	B1	1.2	oe awrt 0.51
1(a)(ii)	$\frac{1317}{1429} \approx 0.922$	B1	1.2	oe awrt 0.92
1(a)(iii)	$\frac{133}{1429} \approx 0.0931$	B1	1.2	oe awrt 0.093
1(a)(iv)	M $\cup$ A = 362 + 112 (= 474)	M1	1.2	PI Union correctly attempted May be seen as probability awrt 0.33
	$\frac{89 + 12}{474}$	M1	1.2	PI Division of P(C) by P(M $\cup$ A) May be seen as awrt $\frac{0.071}{0.33}$
	$\frac{101}{474} \approx 0.213$	A1	1.2	oe awfw 0.212~0.215
1(b)	The **probability** that the whale sighted was an **orca given that** the month was **March or April**.	E1	2.1a	oe Correct definition in context

Qu	Scheme	Marks	AO	Notes
1(c)	$P(F) \times P(W) = \frac{486}{1429} \times \frac{211}{1429}$ $= 0.340 \times 0.148$	M1	1.2	PI Correctly multiplying probabilities awfw $0.34 \times (0.147 \sim 0.148)$
	$= 0.0502 \neq \frac{133}{1429}$ $(=0.0931)$	E1dep	2.1b	**Accept** $\neq \mathrm{P}(W \cap F)$ Correctly comparing probabilities awrt 0.0502~0.0503 **Accept** equivalent sentence in words Dep all figures correct
	Alternative 1			
	$P(F\|W) = \frac{133}{211}$ $(= 0.630)$	(M1)		Find $\mathrm{P}(F\|W)$
	$\neq \frac{486}{1429}$ $(= 0.340)$	(E1)		**Accept** $\neq \mathrm{P}(F)$ Clearly demonstrate, with numerical support, that $\mathrm{P}(F\|W) \neq \mathrm{P}(F)$
	Alternative 2			
	$P(W\|F) = \frac{133}{486}$ $(= 0.274)$	(M1)		Find $\mathrm{P}(W\|F)$
	$\neq \frac{211}{1429}$ $(= 0.148)$	(E1)		**Accept** $\neq \mathrm{P}(W)$ Clearly demonstrate, with numerical support, that $\mathrm{P}(W\|F) \neq \mathrm{P}(W)$

Qu	Scheme	Marks	AO	Notes
1(d)	<table><tr><td>Expected</td><td>Jan</td><td>Feb</td><td>Mar</td><td>April</td></tr><tr><td>orca</td><td>162.1</td><td>168.0</td><td>125.1</td><td>**38.7**</td></tr><tr><td>fin</td><td>69.3</td><td>71.8</td><td>**53.5**</td><td>**16.5**</td></tr><tr><td>pilot</td><td>237.6</td><td>**246.2**</td><td>**183.4**</td><td>**56.7**</td></tr></table>			Results of expected values calculations should be given to 3sf or 2dp if < 1 **Condone** small slip
		M1	1.3	Any correct value in bold box
		A1	1.3	All calculated expected values correct and shown
	<table><tr><td>4.79</td><td>7.29</td><td>10.4</td><td>18.4</td></tr><tr><td>8.49</td><td>52.3</td><td>7.83</td><td>16.5</td></tr><tr><td>0.06</td><td>37.6</td><td>17.5</td><td>33.0</td></tr></table>			(Table of contributions)
	$\frac{(190-162.1)^2}{162.1}+\cdots+\frac{(100-56.7)^2}{56.7}$	M1	1.3	PI Method for contributions correct Give mark for one correct value seen **Condone** small slip
	= 214	A1	1.3	awfw 213~215
1(e)	df = 6	B1	1.3	
	cv = 12.6	A1	1.3	awrt 12.6 May appear elsewhere

Qu	Scheme	Marks	AO	Notes
1(f)	H_0: Whale species and month are **independent**. H_1: Whale species and month are **not independent**.	B1	1.3	oe Both hypotheses
	214 > 12.6	M1ft	2.1a	oe Stated comparison of *their* ts and cv ft ts from part (d) as long as the two method marks gained **Condone** "the ts is in the cr"
	(Reject H_0) There is **sufficient evidence** that the **species** of a whale, sighted in the Southern Ocean, is **not independent** of the **month** of the year.	E1dep	2.1a	oe Conclusion in context Dep correct ts and cv
	There were far more fin whales sighted in February than expected.	E1	2.1b	Mark for correct comment about fin whales in February Ignore further comments about long finned pilot whales - more in April than expected and less in February than expected

Qu	Scheme	Marks	AO	Notes
1(g)	(No,) whales are likely to travel in family groupings (of the same whale species).			oe Sensible reason for a whale species to group together such as family or diet or environment.
	(No,) whales might choose not to enter an area dominated by another species of whale.			oe Sensible reason for whale species to avoid one another. **Accept** use of equivalent word to species such as "type" or "kind" of whale.
	(No,) you may count the same whale more than once			
		B1	3.1b	Any one sensible reason for a lack of independence in this context.
	Total	**20**		

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td rowspan="6">2(a)</td><td>H_0: $\eta_d = 0$
H_1: $\eta_d \neq 0$</td><td>B1</td><td>1.3</td><td>oe
Two-tailed hypotheses
Accept μ_d</td></tr>
<tr><td><table>
<tr><th>Country</th><th>d</th><th>rank</th></tr>
<tr><td>Austria</td><td>-0.07</td><td>1</td></tr>
<tr><td>Indonesia</td><td>0.08</td><td>2.5</td></tr>
<tr><td>Panama</td><td>0.08</td><td>2.5</td></tr>
<tr><td>Cuba</td><td>-0.12</td><td>4</td></tr>
<tr><td>Israel</td><td>-0.13</td><td>5</td></tr>
<tr><td>Costa Rica</td><td>0.3</td><td>6</td></tr>
<tr><td>UAE</td><td>0.7</td><td>7</td></tr>
<tr><td>Côte d'Ivoire</td><td>0.73</td><td>8</td></tr>
<tr><td>Guinea</td><td>0.77</td><td>9</td></tr>
<tr><td>Guatemala</td><td>1.14</td><td>10</td></tr>
<tr><td>Afghanistan</td><td>1.73</td><td>11</td></tr>
</table></td><td></td><td></td><td></td></tr>
<tr><td></td><td>M1</td><td>1.3</td><td>Correct ranking</td></tr>
<tr><td>T = 10 or T = 56</td><td>A1</td><td>1.3</td><td>At least one rank total correct</td></tr>
<tr><td>cv = 11 (or 55)</td><td>B1</td><td>1.3</td><td>either cv correct</td></tr>
<tr><td>10 < 11 or 56 > 55
(Reject H_0)</td><td>M1</td><td>2.1b</td><td>Correct ts compared to correct tail cv</td></tr>
<tr><td>Significant evidence that total fertility rates by country changed on average between 2000-2005 and 2010-2015.</td><td>E1dep</td><td>2.1a</td><td>Conclusion in context.
Dep correct ts compared to correct tail cv</td></tr>
</table>

Qu	Scheme	Marks	AO	Notes
2(b)	(The figures tell us about individual countries only.) We do not know how much each country contributes to the world total fertility rate.	B1	3.1b	oe **Accept** "We would need to know how many (childbearing) women are in each country (of the world)." **Condone** relevant comments on population
				Do not accept "we only have a sample of 11 countries". (Having these figures for **all** of the countries in the world would still not be sufficient.)

Qu	Scheme	Marks	AO	Notes
2(c)	**Practical difficulties** Comment on bias [Not exhaustive]			
	It is difficult to collect accurate records for a representative sample or census (of women and births).			oe Comment on the general difficulty of obtaining a representative sample or census (**either**)
	If a sample is not representative, bias will be introduced.			
	The data will be biased towards women who are easy to reach in a census.			May see specific practical areas of bias, as described below.
	The required data might be impractical to collect from mothers who live in rural locations, far from healthcare facilities.			oe Identification of appropriate practical difficulty for example poverty, war, lack of governmental structures, lack of healthcare, distance from healthcare or other social and political issues
	The data might have a bias towards mothers who live in urban areas (who may have more or fewer babies than those in rural areas).			

Qu	Scheme	Marks	AO	Notes
2(c) cont.	**The required data records may be more practical to collect as a country becomes more developed.**			oe Identification of reason for potential bias between times
	The 2010-2015 data may be more accurate than the 2000-2005 data causing a bias.			oe Time bias clearly stated
	Women may not know their age.			
	Women who do not know their age may have a different distribution of number of children born, introducing bias.			
	Women may lie about their age.			
	Women likely to lie about their age may have a different distribution of number of children born, introducing bias.			
		E1	**3.1a**	**Sensible practical difficulty**
		E1	3.1a	Clear identification of a possible bias in the data due to **practical difficulty stated by candidate**
		E1dep	3.1a	Answer fully in context and no errors dep on previous two E1 marks
	Total	**10**		

Qu	Scheme	Marks	AO	Notes
3(i)	$G \sim \mathrm{N}(3151, 551^2)$	M1	1.2	PI Use of correct μ and σ
	$\mathrm{P}(G < 4000) = 0.938$	A1	1.2	awrt 0.938
3(ii)	$(\mu =) 6742$	B1	1.2	PI Mean
	$\sigma^2 = 596^2 + 596^2$	M1	1.2	PI Variance method
	$\sigma^2 = 710\ 432$ or $\sigma = 843$	A1	1.2	PI Variance or sd correct awrt σ = 842~843
	$A_1 + A_2 \sim \mathrm{N}(6742, 710\ 432)$ $\mathrm{P}(A_1 + A_2 < 6500) = 0.387$	A1	1.2	Correct probability awrt 0.386~0.387
				awrt 0.386~0.387 nms gains all marks

Qu	Scheme	Marks	AO	Notes
3(iii)	$A - 1.2G > 0$	M1	2.1b	PI oe correct formulation of combination **Accept** any correct rearrangement seen such as: $A > 1.2G$ $1.2G - A < 0$
	$\mu = -410.2$	B1	1.2	PI Mean awrt -410 **Accept** awrt 410
	$\sigma^2 = 596^2 + 1.2^2 \times 551^2$	M1	1.2	PI Method for variance
	$\sigma^2 = 792401.44$ or $\sigma = 890.2$	A1	1.2	PI Variance or sd correct awrt σ = 890
	$A - 1.2G \sim \mathrm{N}(-410.2, 792\ 401)$ $\mathrm{P}(A - 1.2G > 0) = 0.322$	A1	1.2	Correct probability awrt 0.322~0.323
				awrt 0.322~0.323 nms gains all marks
	Total	**11**		

Qu	Scheme	Marks	AO	Notes
4(a)	H_0: $\pi = 0.32$ H_1: $\pi < 0.32$	B1	1.3	oe both **Condone** p
	$X \sim B(200, 0.32)$			
	Approximate to $Y \sim N(64, 43.52)$	M1	2.1b	PI Normal approx. stated or clearly used
		A1	1.3	PI Mean correct
		A1	1.3	PI Variance correct awrt 43.5 or SD awrt 6.6
	$z = \frac{41.5 - 64}{\sqrt{43.52}}$	M1	1.3	PI Use of continuity correction in probability **or** calculation of test statistic e.g. $P(Y \leq 41.5)$
	$z = -3.41 < -1.64$	A1	1.3	Correct z-value awfw $-3.49 \sim -3.48$ **and** compared to -1.64 **Accept** positive z compared to correct tail cv **or** Correct p-value awfw 0.000324~0.000326 **(**0.000244~0.000246 no cc) **and** compared to 5%
	(Reject H_0.) There is **significant evidence**… … to suggest that a smaller proportion of his employees feel that flexible working is discouraged by their manager than the proportion for the UK as a whole.	E1dep	2.1a	or ...to support Sal's belief. Dep ts and cv correct or p-value correct (with or without CC) and compared to 5% Correct conclusion in context

Qu	Scheme	Marks	AO	Notes
4(a) cont.	**Alternative solution**			
	$H_0: \pi = 0.32$ $H_1: \pi < 0.32$	(B1)		oe both
	$\hat{p} = \frac{41}{200}$ $(= 0.205)$	(M1)		Finding $\hat{p}$
	$z = \frac{0.205 - 0.32}{\sqrt{0.32 \times 0.68 \div 200}}$	(A1)		Denominator correct oe
		(A1)		numerator correct, may be reversed
		(M1)		PI Attempt to use correct formula using correct $\hat{p}$
	$z = -3.49 < -1.64$	(A1)		Correct *z*-value awfw -3.48 ~ -3.49 **and** compared to -1.64 **Accept** positive *z* compared to correct tail cv **or** Correct *p*-value awfw 0.000244~0.000246 **and** compared to 5%
	(Reject H_0.) There is **significant evidence**… … to suggest that a smaller proportion of his employees feel that flexible working is discouraged by their manager than the proportion for the UK as a whole.	(E1dep)		**or** ...to support Sal's belief. Dep ts and cv correct or *p*-value correct and compared to 5% Correct conclusion in context
4(b)	Because the (theoretical) mean = 64 > 10 **and** 136 > 10	E1	3.1a	oe **Accept** *np* **or** μ = 64 > 10 **Accept** *np* **or** μ > any number larger than 10 Must mention "mean" **or** "average" **or** *np* **or** μ **Condone** 64 > 10

Qu	Scheme	Marks	AO	Notes
4(c)	Sal probably has access to technology that would allow him to work out the exact Binomial probability (for $n = 200$, $p = 0.32$)	E1	3.1a	oe Explanation referring to modern technology **Accept** such answers as: "He could use a spreadsheet to get the right probability" "A graphical calculator could work out the real answer." "There are websites that can calculate any binomial probability for you." **Condone** "Sal has asked an independent organisation and is not doing any tests himself"
	Total	**9**		

<table>
<tr><th>Qu</th><th>Scheme</th><th>Marks</th><th>AO</th><th>Notes</th></tr>
<tr><td>5(a)</td><td>Randomised Block (design)</td><td>B1</td><td>1.1</td><td></td></tr>
<tr><td>5(b)</td><td>
<table>
<tr><th>Source</th><th>SS</th><th>df</th><th>MS</th><th>F</th></tr>
<tr><td>ph</td><td>349.61</td><td>1</td><td>349.61</td><td>216.92</td></tr>
<tr><td>Light</td><td>77.52</td><td>2</td><td>38.76</td><td>24.05</td></tr>
<tr><td>Error</td><td>3.22</td><td>2</td><td>1.61</td><td></td></tr>
<tr><td>Total</td><td>430.35</td><td>5</td><td></td><td></td></tr>
</table>
</td><td></td><td></td><td>MS_{light} 38.7~39
MS_{error} 1.50~1.65
F_{ph} 211~233
F_{light} 23~26</td></tr>
<tr><td></td><td>$SS_{ph} = \frac{130.2^2}{3} + \frac{84.4^2}{3} - \frac{214.6^2}{6}$
$SS_{light} = \frac{66.4^2}{2} + \frac{66.5^2}{2} + \frac{81.7^2}{2} - \frac{214.6^2}{6}$</td><td>M1</td><td>1.3</td><td>PI
SS method, either ph or light level.
Accept awrt 350 or 78</td></tr>
<tr><td></td><td>$SS_{error} = SS_T - SS_{ph} - SS_{light}$</td><td>M1ft</td><td>1.3</td><td>ft their values as long as result is not negative.</td></tr>
<tr><td></td><td>df = 1, 2, 2, 5</td><td>B1</td><td>1.3</td><td>All df correct and in correct row</td></tr>
<tr><td></td><td>$MS_{ph} = \frac{349.61}{1}$
$MS_{light} = \frac{77.52}{2}$
$MS_{error} = \frac{3.22}{2}$</td><td>M1ft</td><td>1.3</td><td>PI
Any MS calculated correctly using their non-negative values</td></tr>
<tr><td></td><td>$F_{ph} = \frac{349.61}{1.61}$
$F_{light} = \frac{38.76}{1.61}$</td><td>M1ft</td><td>1.3</td><td>PI
Either F calculated correctly using their non-negative values</td></tr>
<tr><td></td><td></td><td>A1</td><td>1.3</td><td>All values correct</td></tr>
</table>

Qu	Scheme	Marks	AO	Notes
5(c)	[H_0: $\mu_i = \mu_j$ for all i, j H_1: $\mu_i \neq \mu_j$ for some i, j]			Hypotheses not required
	ts = 24.05 cv = 19.0 (< 24.05)	B1	1.3	Correct cv **or** *p-value* = 0.0399 (< 5%) awrt 0.037~0.042
	There is **significant evidence** that (wollemia) seedlings grown in different **light levels** have **different average** levels of **chlorophyll** in their leaves.	E1dep	2.1a	oe in context. Correctly linking light level and chlorophyll level of leaves. **Accept** "it seems that more light means more chlorophyll" for this mark only Must contain element of doubt dep on B1 and cv or *p-value* used correctly
	The **high** light level appears to produce the **healthiest** wollemia (trees) seedlings.	E1dep	2.1b	oe in context. Must specify the "high" or "highest" light level only May see both E marks in one sentence

Qu	Scheme	Marks	AO	Notes
5(d)	Yes	B1dep	3.1a	Dep attempt at correct reason Must unambiguously agree
	Possible Explanations			List not exhaustive
	Because soil pH accounted for much of the variance in the chlorophyll levels.			oe
	Significant evidence that different soil pH also implies different average chlorophyll level			oe conclusion of a hypothesis test: cv = (18.5 <) 216.92 or *p-value* awfw 0.0046~0.0049 (< 5%)
	Plants growing in 4.5 pH soil had **more chlorophyll** than those growing in 6.5 pH soil			
		E1	3.1b	Any suitable reason
5(e)	**Possible explanations**			
	So that she could be more sure that any differences found between groups were due to soil pH and/or light levels.			
	To reduce variance in the results due to factors other than soil pH or light level.			
	To control for confounding/other factors such as weather.			
	To reduce experimental error.			
		E1	3.1a	Correct reason given No context needed
		E1	3.1b	Fully correct in context
				Do not accept "To make the experiment fair." oe for either mark.

Qu	Scheme	Marks	AO	Notes
5(f)	External factors that Catherine has not considered might change (how light affects) the health of wollemia seedlings.			
	or			
	The conditions that Catherine has created in the lab for wollemia seedlings might never occur in the wild.			
		E1	3.1b	Conditions in the wild/outside of the laboratory are different from those in the laboratory
		E1	3.1b	Fully correct in context
	Total	16		

Qu	Scheme	Marks	AO	Notes
6(a)	A statistic is a **numerical property of a sample**	E1	1.1	oe **Accept** "a number calculated from a sample"
	and is **a function of only the values in the sample** (and contains no unknown parameters)	E1	1.1	oe Clarification that a statistic is calculated from only the values in the sample **Accept** "using **just** the numbers in the sample"
6(b)	<table><tr><td>200g</td><td></td></tr><tr><td>192g</td><td>✓</td></tr><tr><td>4.91g</td><td>✓</td></tr></table>	B1	1.1	cao Must indicate both correct answers and not the incorrect answer unambiguously
6(c)	t = (±)2.201	B1	1.3	PI Correct t-value **Accept** either sign
	$192 \pm (t) \times 4.91$	M1ft	1.2	PI Method for CI ft incorrect t-value or z-value
	CI: (181.2, 202.8)	A1	1.2	Both correct to 1dp awrt 181 and 202~203
				NMS with correct CI scores full marks

Qu	Scheme	Marks	AO	Notes
6(d)	200g is within the 95% confidence interval	M1ft	2.1a	oe **Accept** "200 is in CI" ft their (c)
	So there is no significant evidence to doubt the manufacturer's claim (that the average weight of biscuits in a packet is 200g)	A1ft	2.1b	oe **Accept** "so the claim is supported" Must not express certainty ft their (c)
	Total	**8**		

Qu	Scheme	Marks	AO	Notes
7				**Accept** "countryside" for rural and "city" for urban throughout **Do not accept** marks for repetition of the values from **Fig 8** in a sentence without contextual meaning throughout
	Median			
	There is **greater education deprivation**, on **average**, in **urban** areas **than rural** areas (of Northern Ireland)	B1	2.1a	oe correct *comparison* of medians
	IQR			
	There is a **greater spread** of **levels of education deprivation** in **urban** areas **than rural** areas (of Northern Ireland)	B1	2.1a	oe Correct *comparison* of IQRs **Accept** "interquartile range of education deprivation levels higher for urban than rural areas" for B1

Qu	Scheme	Marks	AO	Notes
7 cont.	**SRCC**			
	There is a stronger link between education deprivation and **income** deprivation in urban areas than rural areas.			oe Comparison **Accept** reference to correlation/association **Accept** implied causality for B mark only
	There is (also) a stronger link between education deprivation and **employment** deprivation in urban areas than rural areas.			
		B1	2.1a	Any of the above
		E1	2.1b	Style mark for median/IQR/SRCC For correct comments written in appropriate non-technical style
	Education deprivation is greater and has stronger links to other types of deprivation in **urban** areas than **rural** areas of **Northern Ireland**.	E1	2.1b	oe Conclusion **in context** summing up the contrasting situation in rural and urban areas of NI
		E1	2.1b	Style mark for conclusion
	Total	**6**		

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