

Monday 23 January 2012 – Morning

A2 GCE MATHEMATICS (MEI)

4767 Statistics 2

QUESTION PAPER



Candidates answer on the Printed Answer Book.

OCR supplied materials:

- Printed Answer Book 4767
- MEI Examination Formulae and Tables (MF2)

Duration: 1 hour 30 minutes

Other materials required:

- Scientific or graphical calculator

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found in the centre of the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book.** Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION FOR CANDIDATES

This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets [] at the end of each question or part question on the Question Paper.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **4** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

- 1 Nine long-distance runners are starting an exercise programme to improve their strength. During the first session, each of them has to do a 100 metre run and to do as many push-ups as possible in one minute. The times taken for the run, together with the number of push-ups each runner achieves, are shown in the table.

Runner	A	B	C	D	E	F	G	H	I
100 metre time (seconds)	13.2	11.6	10.9	12.3	14.7	13.1	11.7	13.6	12.4
Push-ups achieved	32	42	22	36	41	27	37	38	33

- (i) Draw a scatter diagram to illustrate the data. [3]
- (ii) Calculate the value of Spearman's rank correlation coefficient. [5]
- (iii) Carry out a hypothesis test at the 5% significance level to examine whether there is any association between time taken for the run and number of push-ups achieved. [6]
- (iv) Under what circumstances is it appropriate to carry out a hypothesis test based on the product moment correlation coefficient? State, with a reason, which test is more appropriate for these data. [3]
- 2 The number of printing errors per page in a book is modelled by a Poisson distribution with a mean of 0.85.
- (i) State conditions for a Poisson distribution to be a suitable model for the number of printing errors per page. [2]
- (ii) A page is chosen at random. Find the probability of
 (A) exactly 1 error on this page,
 (B) at least 2 errors on this page. [5]
- 10 pages are chosen at random.
- (iii) Find the probability of exactly 10 errors in these 10 pages. [3]
- (iv) Find the least integer k such that the probability of there being k or more errors in these 10 pages is less than 1%. [4]
- 30 pages are chosen at random.
- (v) Use a suitable approximating distribution to find the probability of no more than 30 errors in these 30 pages. [5]

- 3 The lifetime of a particular type of light bulb is X hours, where X is Normally distributed with mean 1100 and variance 2000.
- (i) Find $P(1100 < X < 1200)$. [3]
- (ii) Use a suitable approximating distribution to find the probability that, in a random sample of 100 of these light bulbs, no more than 40 have a lifetime between 1100 and 1200 hours. [5]
- (iii) A factory has a large number of these light bulbs installed. As soon as 1% of the bulbs have come to the end of their lifetimes, it is company policy to replace all of the bulbs. After how many hours should the bulbs need to be replaced? [3]
- (iv) The bulbs are to be replaced by low-energy bulbs. The lifetime of these bulbs is Normally distributed and the mean is claimed by the manufacturer to be 7000 hours. The standard deviation is known to be 100 hours. A random sample of 25 low-energy bulbs is selected. Their mean lifetime is found to be 6972 hours. Carry out a 2-tail test at the 10% level to investigate the claim. [8]

[Question 4 is printed overleaf.]



Copyright Information

OCR is committed to seeking permission to reproduce all third-party content that it uses in its assessment materials. OCR has attempted to identify and contact all copyright holders whose work is used in this paper. To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced in the OCR Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download from our public website (www.ocr.org.uk) after the live examination series.

If OCR has unwittingly failed to correctly acknowledge or clear any third-party content in this assessment material, OCR will be happy to correct its mistake at the earliest possible opportunity.

For queries or further information please contact the Copyright Team, First Floor, 9 Hills Road, Cambridge CB2 1GE.

OCR is part of the Cambridge Assessment Group; Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

- 4 Birds are observed at feeding stations in three different places – woodland, farm and garden. The numbers of finches, thrushes and tits observed at each site are summarised in the table. The birds observed are regarded as a random sample from the population of birds of these species that use these feeding stations.

Observed Frequency		Place			
		Farm	Garden	Woodland	Totals
Species	Thrushes	11	74	7	92
	Tits	70	26	88	184
	Finches	17	2	10	29
	Totals	98	102	105	305

The expected frequencies under the null hypothesis for the usual χ^2 test are shown in the table below.

Expected Frequency		Place		
		Farm	Garden	Woodland
Species	Thrushes	29.5607	30.7672	31.6721
	Tits	59.1213	61.5344	63.3443
	Finches	9.3180	9.6984	9.9836

- (i) Verify that the entry 9.3180 is correct.

[2]

The corresponding contributions to the test statistic are shown in the table below.

Contribution		Place		
		Farm	Garden	Woodland
Species	Thrushes	11.6539	60.7489	19.2192
	Tits	2.0017	20.5201	9.5969
	Finches	6.3332	6.1108	0.0000

- (ii) Verify that the entry 6.3332 is correct.

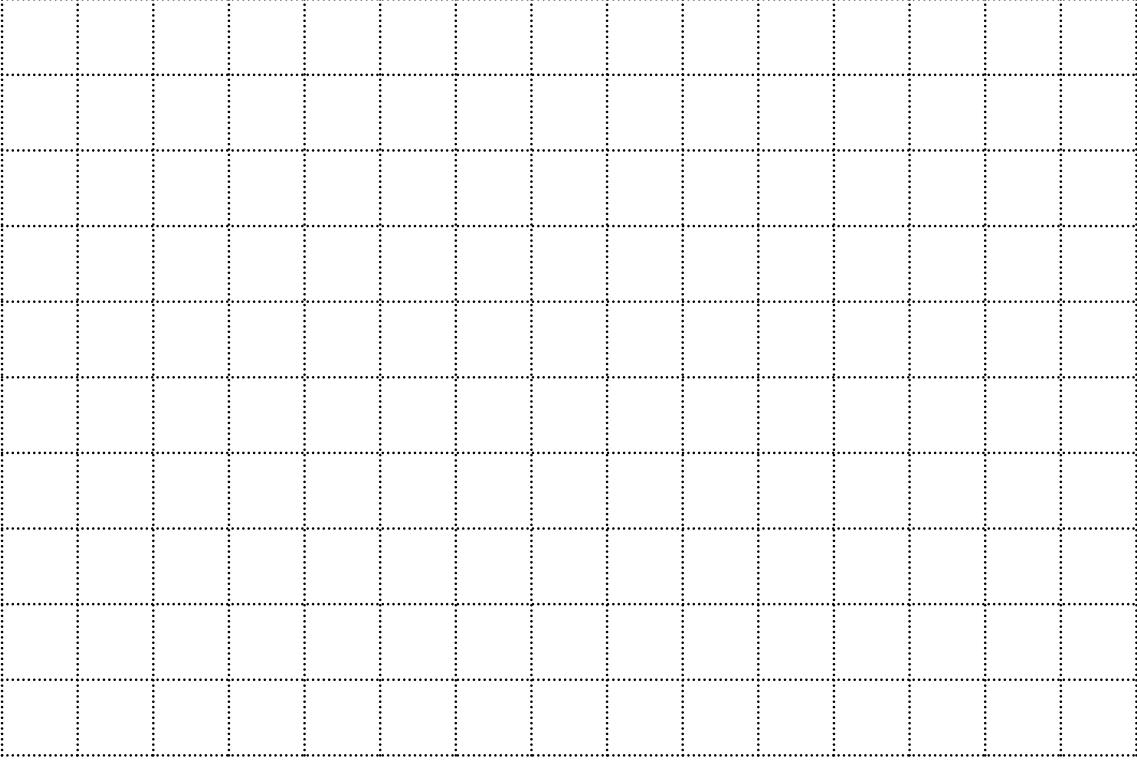
[2]

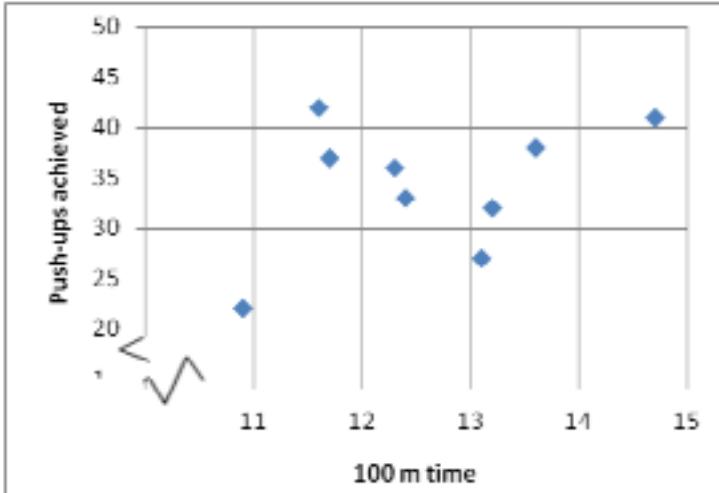
- (iii) Carry out the test at the 1% level of significance.

[7]

- (iv) For each place, use the table of contributions to comment briefly on the differences between the observed and expected distributions of species.

[6]

1 (i)	
1 (ii)	<hr/>

Question		Answer	Marks	Guidance																						
1	(i)	 <table border="1"><caption>Data points from scatter plot</caption><thead><tr><th>100 m time</th><th>Push-ups achieved</th></tr></thead><tbody><tr><td>11.5</td><td>22</td></tr><tr><td>11.8</td><td>37</td></tr><tr><td>12.0</td><td>35</td></tr><tr><td>12.2</td><td>33</td></tr><tr><td>12.5</td><td>42</td></tr><tr><td>12.8</td><td>38</td></tr><tr><td>13.0</td><td>27</td></tr><tr><td>13.5</td><td>32</td></tr><tr><td>14.0</td><td>39</td></tr><tr><td>14.8</td><td>41</td></tr></tbody></table>	100 m time	Push-ups achieved	11.5	22	11.8	37	12.0	35	12.2	33	12.5	42	12.8	38	13.0	27	13.5	32	14.0	39	14.8	41	G1 G2,1,0 [3]	<p>G1 For axes suitably labelled with some indication of linear scale provided.</p> <p>G2 for points plotted correctly. G1 if 8 points plotted correctly. G0 if two or more incorrectly plotted/omitted points.</p> <p>Special Case SC1 for points visibly correct on axes where no indication of scale has been provided.</p> <p>Allow axes reversed</p>
100 m time	Push-ups achieved																									
11.5	22																									
11.8	37																									
12.0	35																									
12.2	33																									
12.5	42																									
12.8	38																									
13.0	27																									
13.5	32																									
14.0	39																									
14.8	41																									

1	(ii)		<table border="1"> <thead> <tr> <th>Athlete</th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th><th>I</th></tr> </thead> <tbody> <tr> <td>100 metre time</td><td>13.2</td><td>11.6</td><td>10.9</td><td>12.3</td><td>14.7</td><td>13.1</td><td>11.7</td><td>13.6</td><td>12.4</td></tr> <tr> <td>Push ups achieved</td><td>32</td><td>42</td><td>22</td><td>36</td><td>41</td><td>27</td><td>37</td><td>38</td><td>33</td></tr> <tr> <td>Rank 100m</td><td>7</td><td>2</td><td>1</td><td>4</td><td>9</td><td>6</td><td>3</td><td>8</td><td>5</td></tr> <tr> <td>Rank Push-ups</td><td>3</td><td>9</td><td>1</td><td>5</td><td>8</td><td>2</td><td>6</td><td>7</td><td>4</td></tr> <tr> <td>d</td><td>4</td><td>-7</td><td>0</td><td>-1</td><td>1</td><td>4</td><td>-3</td><td>1</td><td>1</td></tr> <tr> <td>d^2</td><td>16</td><td>49</td><td>0</td><td>1</td><td>1</td><td>16</td><td>9</td><td>1</td><td>1</td></tr> </tbody> </table> <p> $\Sigma d^2 = 94$ $r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 94}{9 \times 80} = 1 - 0.783$ $= 0.217$ (to 3 s.f.) [allow 0.22 to 2 s.f.] </p>	Athlete	A	B	C	D	E	F	G	H	I	100 metre time	13.2	11.6	10.9	12.3	14.7	13.1	11.7	13.6	12.4	Push ups achieved	32	42	22	36	41	27	37	38	33	Rank 100m	7	2	1	4	9	6	3	8	5	Rank Push-ups	3	9	1	5	8	2	6	7	4	d	4	-7	0	-1	1	4	-3	1	1	d^2	16	49	0	1	1	16	9	1	1	M1 M1 A1 M1 A1 [5]	For ranking (allow all ranks reversed for either or both categories) NB No ranking or re-allocation of pairs scores 0/5 For d^2 For Σd^2 M1 for method for r_s used A1 f.t. for $ r_s < 1$ Allow 13/60 or $r_s = 1 - 1.217 = -0.217$ with reversed ranks Push-up times ranked from highest (1 st) to Lowest (9 th) gives $\Sigma d^2 = 146$ which leads to -0.217. Allow both A marks.
Athlete	A	B	C	D	E	F	G	H	I																																																																		
100 metre time	13.2	11.6	10.9	12.3	14.7	13.1	11.7	13.6	12.4																																																																		
Push ups achieved	32	42	22	36	41	27	37	38	33																																																																		
Rank 100m	7	2	1	4	9	6	3	8	5																																																																		
Rank Push-ups	3	9	1	5	8	2	6	7	4																																																																		
d	4	-7	0	-1	1	4	-3	1	1																																																																		
d^2	16	49	0	1	1	16	9	1	1																																																																		

	A	B	C	D	E	F	G	H	I
100m	7	2	1	4	9	6	3	8	5
Push up	7	1	9	5	2	8	4	3	6
d	0	1	-8	-1	7	-2	-1	5	-1
d^2	0	1	64	1	49	4	1	25	1

1	(iii)	<p>H_0: no association between 100m time and number of push-ups achieved in the population of long distance runners</p> <p>H_1: some association between 100m time and number of push-ups achieved in the population of long distance runners</p> <p>Two tail test critical value at 5% level is 0.7000</p> <p>Since $0.217 < 0.7000$, there is insufficient evidence to reject H_0,</p>	B1	B1 for H_0 in context (not x & y)
			B1	<p>B1 for H_1 in context (not x & y)</p> <p>SC1 for both correct but no context provided</p>

B1 for population SOI
NB $H_0 H_1$ not ito ρ
Do not condone the use of the word ‘correlation’ in place of ‘association’. Population’ should be mentioned to award B1, unless clear, unambiguous alternative wording is used.

B1 for ± 0.7000 (or ± 0.6000 only if H_1 indicates a 1-tailed test is intended)

M1 for sensible comparison with c.v leading to a conclusion seen, provided $|r_s| < 1$
NOTE The comparison can be in the form of a diagram as long as it is clear and unambiguous.

Sensible comparison: e.g. $-0.217 > -0.7000$ is ‘sensible’ whereas $-0.217 < 0.7000$ is ‘not sensible’. Allow $-0.7000 < 0.217 < 0.7000$
Reversed inequality sign e.g. $0.217 > 0.7000$ etc gets max M1 A0.

Also, if the c.v. comes from the p.m.c.c. table (0.6664 for 2-tailed test and 0.5822 for a 1-tailed test) award max M1 A0.

		i.e. conclude that there is not enough evidence to show association between the 100m times and the number of push-ups achieved.	A1 [6]	A1ft for correct conclusion in context. Follow through their r_s NOTE 2-tailed test with correct c.v. must be used to award final A1. Use of a 1-tailed test: max B1 B0 B1 B1 (for 0.6000) M1 A0. i.e. max 4/6. Where hypotheses are reversed, lose first two B1 marks and final A1. max 3/6
1	(iv)	<p>It is appropriate to carry out a hypothesis test based on the product moment correlation coefficient when the underlying population has a bivariate Normal distribution.</p> <p>The scatter diagram does not appear to be roughly elliptical so the Spearman coefficient is more appropriate</p>	E1 E1 E1dep [3]	<p>Do not accept ‘both Norrnally distributed’</p> <p>Allow reasonable alternatives e.g. in this case, one variable is discrete so pmcc invalid.</p> <p>E1 dependent on previous E1</p>
2	(i)	<p>Errors have a uniform average rate of occurrence</p> <p>and occur randomly and independently</p>	E1 E1 [2]	<p>E1 must refer to ‘errors’ not ‘events’, ‘data’ or ‘conditions’.</p> <p>Condone ‘constant/fixed average/mean rate/per page’ but not ‘constant average’, ‘constant rate’ or ‘uniform rate’, etc.</p> <p>Allow large n and small p if both defined</p> <p>E1 for randomly and independently</p> <p>If ‘errors’ not referred to then SC1 if otherwise correct. Condone ‘the number of errors’</p>

2	(ii)	(A)	$P(X=1) = e^{-0.85} \frac{0.85^1}{1!} = 0.3633$	M1 A1	M1 for attempt to find $P(X=1)$ either by Poisson p.d.f. or use of tables. A1 CAO 3s.f. for answers which round to 0.363 www NOTE If $P(X \leq 1)$ used for final answer, award M0A0. Interpolation gives $0.79065 - 0.42795 = 0.3627$
		(B)	$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) = 1 - e^{-0.85} \frac{0.85^0}{0!} - e^{-0.85} \frac{0.85^1}{1!} \\ &= 1 - 0.4274 - 0.3633 = 0.2093 \end{aligned}$		M1 M1 A1 [5]
2	(iii)		New $\lambda = 10 \times 0.85 = 8.5$ $P(\text{Exactly 10 in 10 pages}) = 0.7634 - 0.6530 = 0.1104$ Or $= e^{-8.5} \frac{8.5^{10}}{10!} = 0.1104$	B1 M1 A1 [3]	B1 for 8.5 M1 for $P(X=10)$ calculation using $\lambda = 8.5$ CAO Allow 0.110 and 0.11 www Award M1 only if $\lambda = 8.5$ used

2	(iv)	<p>So $P(k - 1 \text{ or less in 10 pages}) > 99\%$</p> <p>From tables $P(X \leq 15) = 0.9862$, $P(X \leq 16) = 0.9934$</p> <p>$P(X \geq 16) = 1 - P(X \leq 15) = 0.0138 > 1\%$ $P(X \geq 17) = 1 - P(X \leq 16) = 0.0066 < 1\%$</p> <p>$P(k \text{ or more in 10 pages}) < 1\% \text{ means}$ $k - 1 = 16, k = 17$</p>	M1 A1 A1 A1 [4]	<p>M1 for $P(X \leq k - 1) > 0.99$ seen, or evidence of a search for values > 0.99 from cumulative Poisson tables seen.</p> <p>A1 for finding either one of 0.9862 and 0.9934 (or either one of 0.0138 and 0.0066) A1 for both (3s.f.)</p> <p>A1 CAO for $k = 17$ SC1 for evidence of a search for values > 0.9 from cumulative Poisson tables seen. Or for $k = 17$ with no supporting evidence seen.</p>
2	(v)	<p>Mean number in 30 pages = $30 \times 0.85 = 25.5$</p> <p>Using Normal approx. to the Poisson, $X \sim N(25.5, 25.5)$</p> $\begin{aligned} P(X \leq 30) &= P\left(Z \leq \frac{30.5 - 25.5}{\sqrt{25.5}}\right) \\ &= P(Z < 0.9901) = \Phi(0.9901) \\ &= 0.8389 \end{aligned}$	B1 B1 B1 M1 A1 [5]	<p>For Normal approx attempted.(SOI)</p> <p>For correct parameters (SOI)</p> <p>For correct continuity correction.</p> <p>For correct structure with their parameters CAO (Do not FT wrong or omitted CC) Allow 0.839</p>
3	(i)	$\begin{aligned} P(1100 < X < 1200) &= P\left(\frac{1100 - 1100}{\sqrt{2000}} < Z < \frac{1200 - 1100}{\sqrt{2000}}\right) \\ &= P(0 < Z < 2.236) \\ &= \Phi(2.236) - 0.5 \\ &= 0.9873 - 0.5 \\ &= 0.4873 \end{aligned}$	M1 M1 A1 [3]	<p>For standardising M0 if ‘continuity correction’ applied</p> <p>For for correct structure</p> <p>A1 CAO do not allow 0.4871, 0.48713, 0.48745 or 0.4875</p>

3	(ii)	<p>Use Normal approx with</p> $\mu = np = 100 \times 0.4873 = 48.73$ $\sigma^2 = npq = 100 \times 0.4873 \times 0.5127 = 24.98$ $P(X \leq 40) = P\left(Z \leq \frac{40.5 - 48.73}{\sqrt{24.98}}\right)$ $= P(Z \leq -1.647) = 1 - \Phi(1.647) = 1 - 0.9502$ $= 0.0498$	B1 B1 M1 A1	<p>An appropriate Normal approximation must be used.</p> <p>B1 B1 for μ & σ^2 ft their answer to part (i) provided that a normal approximation is appropriate from their part (i).</p> <p>B1 for continuity correction 40.5</p> <p>M1 correct structure using appropriate Normal approximation</p> <p>CAO 3s.f.</p> <p>NOTE Using B(100, 0.4873) gives 0.0494. which gets 0/5</p> <p>SC If p small enough to justify a Poisson approximation, e.g. 0.05, then B1 for Poisson used, B1 ft for parameter, M1 for structure, M1 attempt at summation, A0</p>
3	(iii)	<p>From tables $\Phi^{-1}(0.01) = -2.326$</p> $\frac{k-1100}{\sqrt{2000}} = -2.326$ $k = 1100 - 2.326 \times \sqrt{2000}$ $k = 996$	B1 M1 A1	<p>± 2.326 seen</p> <p>M1 correct equation as seen or equivalent</p> <p>CAO Allow 996.0</p>

3	(iv)	<p>$H_0: \mu = 7000$; $H_1: \mu \neq 7000$</p> <p>Where μ denotes the population mean lifetime of low energy bulbs</p> <p>Test statistic = $\frac{6972 - 7000}{100/\sqrt{25}} = \frac{-28}{20} = -1.4$</p> <p>Lower 10% level 2 tailed critical value of $z = -1.645$</p> <p>$-1.4 > -1.645$ so not significant.</p> <p>There is not sufficient evidence to reject H_0</p> <p>There is insufficient evidence to conclude that the manufacturer is wrong.</p>	B1 B1 B1 M1 A1	<p>B1 For use of 7000 in hypotheses B1 For correct hypotheses given in terms of μ (not p or x, etc. unless letter used is clearly defined as population mean.) If hypotheses are reversed lose second B1 and final A1 B1 for definition of μ. M1 calculation of test statistic with a divisor of $100/\sqrt{25}$. Condone numerator reversed. CAO for -1.4 Allow $+1.4$ for A1 only if this is later compared with $+1.645$</p> <p>B1 For -1.645 Must be negative unless it is clear that absolute values are being used. NB: FT a 1-tail test (c.v. = -1.282)</p> <p>M1 M1 for a sensible comparison leading to a conclusion.</p> <p>A1 For correct conclusion in context FT only their test statistic if c.v. correct and both M marks earned. <u>Critical Value Method</u> $7000 - 1.645 \times 100 \div \sqrt{25}$ gets M1B1 = 6967.1 gets A1 $6972 > 6967.1$ gets M1 for sensible comparison A1 still available for correct conclusion in words & context</p> <p><u>Confidence Interval Method</u> CI centred on 6972 + or $-1.645 \times 100 \div \sqrt{25}$ gets M1 B1 = (6939.1, 7004.9) gets A1 contains 7000 gets M1 A1 still available for correct conclusion in words & context</p>
---	------	---	----------------------------	--

					<p><u>Probability Method</u> Finding $P(\text{sample mean} < 6972) = 0.0808$ gets M1 A1 B1 $0.0808 > 0.05$ gets M1 for a sensible comparison if a conclusion is made. $0.0808 < 0.10$ gets M1 A0 unless using one-tailed test A1 still available for correct (one-tailed) conclusion in words & context. Condone $P(\text{sample mean} > 6972) = 0.9192$ for M1 but only allow A1 B1 if later compared with 0.95, at which point the final M1 and A1 are still available</p>	
			[8]		<p><u>One-tailed test</u> Max B1 B0 B1 M1 A1 B1 (for $\text{cv} = -1.282$) M1 A0</p>	
4	(i)		Expected frequency = $29/305 \times 98$ = 9.3180	M1 A1 [2]	M1 for row total \times column total \div grand total NB Answer given	
4	(ii)		Contribution = $(17 - 9.3180)^2 / 9.3180$ = 6.3332	M1 A1 [2]	M1 for valid attempt at $(f_o - f_e)^2 \div f_e$ NB Answer given	
4	(iii)		H ₀ : no association between place and species. H ₁ : some association between place and species. Test statistic $X^2 = 136.185$ Refer to X ₄ ² Critical value at 1% level = 13.28 Result is significant There is (strong) evidence to suggest association between place and species.	B1 M1 A1 B1 B1 B1 E1 [7]	B1 for hypotheses in context NB if H ₀ H ₁ reversed, or 'correlation' mentioned, do not award first B1 or final E1 M1 for sum of contributions A1 for test statistic. Allow 136.1847 For 4 deg of freedom (seen) CAO For cv – no FT from here if wrong/omitted For significant E1 for correct conclusion in context	

4	(iv)	<p>Farm Contribution of 11.6539 implies that there are (far) fewer thrushes than expected. Also more finches than expected and more tits.</p> <p>Garden Contribution of 60.7489 implies that there are far more thrushes than expected. Contribution of 20.5201 implies that there are (far) fewer tits than expected. Also fewer finches than expected.</p> <p>Woodland Contribution of 19.2192 implies that there are (far) fewer thrushes than expected. Contribution of 9.5969 implies that there are (far) more tits than expected. Number of finches is as expected.</p>	E2,1,0	<p>For each place - award E2 for an explicit reference to a value from the contributions table with a correct* corresponding comment, provided that no further incorrect statements are made regarding that place. *See table below Wording needs to be clear. Do not accept 'different to expected' or 'positive/negative association'</p> <p>Award E1 either for an explicit/implicit reference to a value from the contributions table with a correct corresponding comment accompanied by one or more incorrect comment for that place</p> <p>Allow alternative wording. e.g. 'more thrushes were expected' in place of 'there were fewer thrushes than expected'</p>
			E2,1, 0	

	Farm	Garden	Woodland
Thrushes	11.6539 (far) fewer than expected	60.7489 Far more than expected	19.2192 (far) fewer than expected
Tits	2.0017 As expected/ more than expected	20.5201 (far) fewer than expected	9.5969 (far) more than expected
Finches	6.3332 (far) more than expected	6.1108 fewer than expected	0.0000 as expected