A-Level

Statistics

SS05

Final Mark scheme

6380

June 2017

Version/Stage: v1.0

Mark schemes are prepared by the Lead Assessment Writer and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation events which all associates participate in and is the scheme which was used by them in this examination. The standardisation process ensures that the mark scheme covers the students' responses to questions and that every associate understands and applies it in the same correct way. As preparation for standardisation each associate analyses a number of students' scripts. Alternative answers not already covered by the mark scheme are discussed and legislated for. If, after the standardisation process, associates encounter unusual answers which have not been raised they are required to refer these to the Lead Assessment Writer.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of students' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Further copies of this mark scheme are available from aqa.org.uk

Key to mark scheme abbreviations

M	mark is for method
m or dM	mark is dependent on one or more M marks and is for method
A	mark is dependent on M or m marks and is for accuracy
B	mark is independent of M or m marks and is for method and accuracy
E	mark is for explanation
√or ft or F	follow through from previous incorrect result
CAO	correct answer only
CSO	correct solution only
AWFW	anything which falls within
AWRT	anything which rounds to
ACF	any correct form
AG	answer given
SC	special case
OE	or equivalent
A2,1	2 or 1 (or 0) accuracy marks
–*x* EE	deduct *x* marks for each error
NMS	no method shown
PI	possibly implied
SCA	substantially correct approach
c	candidate
sf	significant figure(s)
dp	decimal place(s)

No Method Shown

Where the question specifically requires a particular method to be used, we must usually see evidence of use of this method for any marks to be awarded.

Where the answer can be reasonably obtained without showing working and it is very unlikely that the correct answer can be obtained by using an incorrect method, we must award **full marks**. However, the obvious penalty to candidates showing no working is that incorrect answers, however close, earn **no marks**.

Where a question asks the candidate to state or write down a result, no method need be shown for full marks.

Where the permitted calculator has functions which reasonably allow the solution of the question directly, the correct answer without working earns **full marks**, unless it is given to less than the degree of accuracy accepted in the mark scheme, when it gains **no marks**.

Otherwise we require evidence of a correct method for any marks to be awarded.

Q	Solution	Marks	Total	Comments
1(a)(i)	P(X > 100.5) = $\frac{103-100.5}{103-99}$	M1		x/4 or (1 – x/4) where 0 < x < 4
	= $\frac{2.5}{4} = \frac{5}{8}$	A1		o.e. eg 0.625
				alt: P(X > 100.5) = $1 - \frac{100.5-99}{103-99}$ M1 = $1 - \frac{1.5}{4} = \frac{5}{8}$ A1
				alt: f(x) = ¼ , 103 – 100.5 = 2.5 M1 P(X > 100.5) = 2.5×¼ = 0.625 A1
			2	
(a)(ii)	$\text{mean} = \frac{103+99}{2} = 101$	B1		
	variance $= \frac{(4)^2}{12} = \frac{4}{3}$	M1		Must be labelled as variance unless A1 gained
	s.d = 1.15 or $\frac{2}{\sqrt{3}}$	A1		1.15 ~ 1.16
			3	
1(b)	$P(\bar{X} > 100.75) = P(Z > \frac{100.75-101}{\sqrt{\frac{4/3}{50}}})$	B1F		B1F: their variance in a(ii) ÷ by 50 or their s.d in a(ii) ÷√50 identified or used.
	= P(Z > -1.5309..)	M1		M1: standardising, P.I. ; allow 101 – 100.75 in the numerator, must have x/√50 in denominator.
	= 0.937	A1		awfw 0.93 ~ 0.94
			3	
	Total		**8**	

Q	Solution	Marks	Total	Comments
2(a)	$H_0: \sigma = 15.3$ $H_1: \sigma < 15.3$	B1		Both hypotheses o.e.
	$s_{new} = 9.09$	B1		awfw 9.08 ~ 9.09 (9.086453…)
	$\frac{(n-1)s^2}{\sigma^2} = \frac{10 \times 9.086^2}{15.3^2}$	M1		f.t. their s_{new}
	$= 3.53$	A1		awfw 3.52 ~ 3.53 (3.527…)
	c.v $\chi^2_{10} = 3.94$	B1		or p-value 0.0338 (0.0335~0.0339)
	t.s. 3.53 < c.v 3.94 or 0.0338 < 0.05 reject H_0	A1		A1 correct decision.
	Reject H_0. There is evidence at the 5% level to suggest there has been a reduction in the standard deviation of the difference between the actual and estimated delivery times since the purchase of the new software.	E1dep	**7**	Correct conclusion in context dependent previous A1
2(b)	Part a) showed that the differences between the actual and estimated delivery times with the new software are less variable; Gilbert has achieved his aim.	E1dep	**1**	Both “Gilbert has achieved his aim” or “yes” and reason wrt variability required; dep on rejecting H_0 above.
	Total		**8**	

Q	Solution	Marks	Total	Comments
3(a)	Mean = $\frac{1}{\lambda} = \frac{1}{0.0125}$			
	= 80	M1		$\frac{1}{0.0125}$
	Mean lifetime is 80 000 hours	A1		cao s.c. B1 for 80 with no other working shown.
			2	
3(b)(i)	P(T < 100) = $\left(1 - e^{-\frac{100}{80}}\right)$	B1		Using T = 100
	= $(1 - e^{-1.25})$	M1		Use of $F(t)$ = 1- $e^{-0.0125t}$ with their T
	= 0.713	A1		0.71~ 0.72 (0.713495…)
(ii)	P(50 < T < 150) = P(T < 150) – P(T < 50)			
	= $\left(1 - e^{-\frac{150}{80}}\right)$ - $\left(1 - e^{-\frac{50}{80}}\right)$	M1		Subtracting two valid cumulative probabilities o.e.; ft their λ ; must be using T =150 and T = 50; allow small slip.
	= $e^{-0.625}$ - $e^{-1.875}$ = 0.53526 – 0.15335			or 0.84665 – 0.46474
	= 0.382	A1		0.38 ~ 0.39 (0.38190…)
			5	
3(c)	365 days = 365 × 24 hours = 8760			
	P(all bulbs last longer than 8760 hours) = $\left(e^{-\frac{8.76}{80}}\right)^{15}$ = $(0.89628)^{15}$ o.e. eg $e^{-0.0125\times8.76\times15}$	M1,m1		M1;dividing their calculated hours by 1000 and using $e^{-0.0125t}$ m1 raising their probability to the power of 15.
	= 0.193	A1	**3**	0.193 ~ 0.194 (0.1934886…)
	Alt: 365×24×15=131400 **M1**; P(T>131.4) = $e^{-0.0125\times131.4}$ **m1** = 0.193 **A1**			
	Total		**10**	

Q	Solution	Marks	Total	Comments
4(a)	H_0: The Poisson distribution is a suitable model H_1: The Poisson distribution is not a suitable model	B1		At least H_0; alt ; "data fits a Poisson distribution"
	Mean = $\frac{\sum x}{n} = \frac{480}{120} = 4$	B1		Clear attempt to find mean; P.I.
		M1		Method for probabilities ft on their mean
	(table below)	m1		Expected values ft *p* x 120 – must be using Poisson probabilities
		A1		At least 7 expected values correct to 2s.f.

live litter size	no. of litters	prob	E
0	3	0.018	2.198
1	10	0.073	8.792
2	13	0.147	17.583
3	18	0.195	23.444
4	27	0.195	23.444
5	25	0.156	18.755
6	14	0.104	12.503
7	7	0.060	7.145
8	3	0.030	3.572
>8	0	0.021	2.564

Q	Solution	Marks	Total	Comments
	Combining classes	m1		Combining first 2 groups (E < 5)
				Last group ≥ 8
		m1		
		m1		Correct attempt at and summation of $(O\text{-}E)^2/E$ values- independent of previous m1,m1- at least one value correct to 2sf.

live litter size	no. of litters	E	(O - E)^2	(O-E)^2/E
≤1	13	10.99	4.04	0.368
2	13	17.58	21.00	1.195
3	18	23.44	29.64	1.264
4	27	23.44	12.65	0.539
5	25	18.76	39.00	2.079
6	14	12.50	2.24	0.179
7	7	7.15	.021	0.003
≥8	3	6.14	9.83	1.603
Total				7.23

4(b)	chi-squared test statistic = 7.23	A1		Awfw 7.15 ~ 7.25
	df = 8 – 1 – 1 = 6	B1ft		Their no. of classes - 2
	c.v. = 10.645 > t.s.	B1		cao or p -value = 0.300 (3sf) > 0.1
	Accept H_0. There is no significant evidence that the Poisson distribution is not a suitable model for the number of live births per litter of wild dogs living in the Canadian wilderness.	E1dep	12	Correct conclusion in context dependent on correct cv or p-value and correct ts
	This is not a random sample of the live litter sizes of all wild dogs so it might be a biased sample. **or** Some of the litters of puppies are likely to be coming from the same or related mothers so we might suspect that the numbers of live births in different litters are not always independent. **or** The number of live births per litter might not happen at a constant average rate. This rate might be affected by, for example, the weather or disease present when the litter was born. **or** The variance of the number of live births per litter is 3.36 and the mean is 4. These should be at least approximately the same for a Poisson distribution. **or** The Poisson distribution theoretically can range up to ∞. This is not possible for the size of a litter of dogs.	E1, EI	2	OE Any two reasonable answers in context wrt the sample chosen or the requirements of a Poisson model Must be in context eg mention of dogs/puppies/births etc.. NOT data
	Total		**14**	

Q	Solution	Marks	Total	Comments
5(a)	$s_x = 0.582$, $s_y = 0.683$ or $s_x^2 = 0.339$, $s_y^2 = 0.466$	B1		either; 0.58 ~ 0.59 (0.58216..);0.68 ~ 0.69 (0.6827…) 0.33 ~ 0.34 (0.33892…);0.46 ~ 0.47 (0.4661..)
	$H_0: \sigma_x^2 = \sigma_y^2$ $H_1: \sigma_x^2 \neq \sigma_y^2$	B1		both hypotheses
	t.s. F = $\frac{0.683^2}{0.582^2}$ = 1.375	M1 A1		1.32 ~ 1.42
	c.v. $F_{8,5} = 6.76$ or p-value = 0.755 > 0.05	B1 ,B1		B1; d.f. (may be implied by correct c.v.) B1; c.v. NB F= 4.82 or 4.817 gains B1 for d.f. **alt**: B1 p-value 0.74 ~ 0.78 (0.7552..) B1 correct comparison with 0.05.
	1.375 < 6.76 accept H_0 , no evidence at the 5% level that the variances of the weights differ .	E1dep	7	conclusion in context, dep on A1 for t.s. and B1 for c.v. or p-value
5(b)	$H_0: \mu_x = \mu_y$ $H_1: \mu_x > \mu_y$	B1		both hypotheses
	$\bar{x} = 3.88$ $\bar{y} = 3.49$	B1		either 3.88 ; 3.48 ~3.50 (3.4911…)
	$S_p{}^2 = \dfrac{5 \times 0.582^2 + 8 \times 0.683^2}{13} = 0.417$	M1		M1: use of correct formula (0.417...; $s_p = 0.6459..$)
	t.s. = $\dfrac{3.88-3.49}{\sqrt{0.417\left(\frac{1}{6}+\frac{1}{9}\right)}}$	M1 M1		numerator (accept (3.49 – 3.88)) denominator (ft on s_p^2 if M1 earned)
	= 1.142	A1		1.13 ~ 1.15 (± to agree with their numerator)
	c.v. $t_{13} = \pm 1.35$ or p - value = 0.137 > 0.1	B1		±1.35 or p-value= 0.13 ~ 0.14(0.1371…) compared with 0.1
	1.142 < 1.35 accept H_0	A1dep		Dep on A1 for t.s. , B1 for c.v. or p -value and correct signs. (accept -1.35 < -1.142)
	<u>Evidence</u> at the 10% level that Emily's suspicion is <u>not supported</u>.	E1dep	9	Conclusion in context dep on previous A1dep oe e.g. Alt: <u>insufficient evidence</u> at 10% level that babies born in the summer months are <u>on average heavier</u> than those born at other times of the year.

5(c)(i)	H_0 has been accepted so a Type II error might have been made.	E1		stating eg Type II error is when H_0 is accepted – H_0 must be accepted in (b) s.c. B1 for Type I error etc… if H_0 is rejected in (b)
(ii)	It was incorrectly concluded that there was no difference between the mean weight of babies born in the summer months and the mean weight of babies born at other times of the year. or Mean weight of babies born in the summer months is greater than the mean weight of babies born at other times of the year. or H_0 incorrectly accepted; Emily's suspicion should have been justified.	E1dep, E1 dep	**3**	E1 explanation of a Type II error eg H_0 accepted when it should have been rejected ; dep E1 in c(i) E1 all correct – must have "mean " or "average" or "Emily's suspicion"
	Total		**19**	

Q	Solution	Marks	Total	Comments
6(a)(i)	Range is 218 – 194 = 24 6 × s.d = 15 24 > 15	M1 A1	**2**	M1 : value seen for range A1: value seen for at least 6 x 2.5 = 15 <u>and</u> correct comparison with 24 or: 24 / 2.5 = 9.6 > (at least) 6 and comment. should be no evidence of calculation of mean or sd
(a)(ii)	s = 8.02 or s^2 = 64.3 $\chi_8^2 = 1.646$, 20.090 98% CI for variance Upper limit $\frac{8\times 8.017^2}{1.646}$ = 312.38 Lower Limit $\frac{8\times 8.017^2}{20.090}$ = 25.59 98% CI for s.d 5.06 < σ < 17.7	B1 B1 M1 m1 A1	**5**	8.01 ~ 8.02 or 64.2 ~ 64.3 (8.017…) both 1.64 ~ 1.65 , 20.0 ~ 20.1 either limit – ft on their χ^2 values square root both ; 5.05~ 5.08 , 17.6 ~ 17.8
(a)(iii)	2.5 lies <u>below CI</u> ; evidence to support Ronan's deduction	E1ft E1dep	**2**	E1 ft correct comparison with their CI in (ii) E1 correct conclusion dep A1 and E1ft Accept "yes, it does"
6(b)(i)	$\bar{x} = 208$ t_8 = 1.86 90% CI: 208 ±1.86×$\frac{8.017}{\sqrt{9}}$ 203 < μ < 213	B1 B1 M1,m1 A1 both	**5**	207 ~ 208 (207.55…) M1; use of their $\frac{s}{\sqrt{9}}$; m1 correct method for interval ft their $\bar{x}$, and $\frac{s}{\sqrt{9}}$, must have correct t value. 201 ~ 203, 211 ~ 213 (202.59, 212.53) NMS: Values must be in range 202.5 ~ 203; 212.5 ~ 213
(b)(ii)	200 is below the lower limit of CI so <u>mean</u> content of pots is at least 200ml Some pots in Ronan's sample contain less than 200ml. or ; Large standard deviation so expect some pots to contain less than 200ml.	E1 ft E1	**2**	Comment in context based on correct comparison of 200 with their CI in b(i) - must include "mean" or "average" Any sensible and correct comment based on sample data ignore comments about validity of amount given on label
	Total		**16**	