2011 Applied Mathematics

Advanced Higher – Statistics

Finalised Marking Instructions

General Marking Principles

These principles describe the approach taken when marking Advanced Higher Applied Mathematics papers. For more detailed guidance please refer to the detailed Marking Instructions.

1 The main principle is to give credit for the skills demonstrated and the criteria met. Failure to have a correct method may not preclude a candidate gaining credit for their solution.

2 The answer to one part of a question, even if incorrect, can be accepted as a basis for subsequent dependent parts of the question.

3 The following are not penalised:

- working subsequent to a correct answer (unless it provides firm evidence that the requirements of the question have not been met)
- legitimate variation in numerical values / algebraic expressions.

4 Full credit will only be given where the solution contains appropriate working. Where the correct answer might be obtained by inspection or mentally, credit may be given.

5 Sometimes the method to be used in a particular question is explicitly stated; no credit will be given where a candidate obtains the correct answer by an alternative method.

6 Where the method to be used in a particular question is not explicitly stated in the question paper, full credit is available for an alternative valid method. (Some likely alternatives are included but these should not be assumed to be the only acceptable ones.)

In the detailed Marking Instructions which follow, marks are shown alongside the line for which they are awarded. There are two codes used M and E. The code M indicates a method mark, so in question B1(a), 1M means a method mark for the product rule.

Advanced Higher Applied Mathematics 2011
Statistics Solutions

A1. (a) An assumption required is that all members of the population are equally likely to pass before the camera. **1**

$$\text{P(Terrorist} \mid \text{Alarm)} = \text{P}(T \mid S)$$

$$= \frac{\text{P}(S \cap T)}{\text{P}(S)} = \frac{\text{P}(T \cap S)}{\text{P}(S)} = \frac{\text{P}(T)\,\text{P}(S \mid T)}{\text{P}(T)\,\text{P}(S \mid T) + \text{P}(\bar{T})\,\text{P}(S \mid \bar{T})}$$ **1**

$$= \frac{(100/1\,000\,000)\,0{\cdot}999}{(100/1\,000\,000)\,0{\cdot}999 + (999900/1000000)\,0{\cdot}001}$$ **1**

$$= \frac{0{\cdot}0000999}{0{\cdot}0000999 + 0{\cdot}0009999} = 0{\cdot}091 \rightarrow 9{\cdot}1\%$$ **1**

{other methods acceptable}

(b) The percentage would be 99·9% if half the population consisted of terrorists. **1**

A2. (a)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{9{\cdot}978 - 10{\cdot}000}{\frac{0{\cdot}16}{\sqrt{50}}} = -0{\cdot}97$$ **1**

With null and alternative hypotheses

$H_0 : \mu = 10$ and $H_1 : \mu \neq 10$ **1**

the critical region for 5% significance is $|z| > 1{\cdot}96$. **1**

Since –0·97 lies outwith the critical region H_0 cannot be rejected so there is no evidence that the mean amount delivered differs from the target. **1**

A two-tail test is appropriate as a mean below 10 could result in patients receiving less of the drug than intended with less benefit and a mean above 10 could result in patients receiving more of the drug than intended with potential harmful effects. **1**

(b)

$$\frac{10{\cdot}4 - 10{\cdot}0}{\sigma} = 3{\cdot}29$$ **1**

$$\Rightarrow \sigma = \frac{0{\cdot}4}{3{\cdot}29} = 0{\cdot}12$$

The standard deviation would have to be reduced to 0·12. **1**

A3.

$$\bar{x} = 0{\cdot}75 \qquad s = 1{\cdot}789$$ **1**

$$t_{9,0{\cdot}95} = 1{\cdot}833$$ **1**

$$\left(\bar{x} - t_{n-1,1-\alpha}\frac{s}{\sqrt{n}}, \infty\right) = \left(0{\cdot}75 - 1{\cdot}833\frac{1{\cdot}789}{\sqrt{10}}, \infty\right)$$

$$= (-0{\cdot}287, \infty)$$ **1**

Since this interval includes 0 the null hypothesis cannot be rejected. **1**

Thus there is no evidence that the drug increases the average number of hours sleep in patients who experience difficulty in sleeping. **1**

A4. (a) Earnings = Salary + Incentive bonus + Christmas bonus

$Y = 1{\cdot}1S + 1000$ **1**

$E(Y) = E(1{\cdot}1S + 1000) = 1{\cdot}1 \times E(S) + 1000 = 28\,500$ **1**

$V(Y) = 1{\cdot}1^2 \times V(S) = 2\,722\,500 \Rightarrow SD(Y) = 1650$ **1**

(b) For total earnings of 36 to exceed 1 000 000 the mean earnings would have to exceed $\frac{1\,000\,000}{36} = 27\,778.$ **1**

$P(\bar{Y} > 27\,778) = P\left(Z > \frac{27\,778 - 28\,500}{\frac{1650}{\sqrt{36}}}\right)$ **1,1**

$= P(Z > -2{\cdot}63)$

$= 0{\cdot}9957$ **1**

A5.

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cross	23·5	12·0	21·0	22·0	19·1	21·5	22·1	20·4	18·3	21·6	23·3	21·0	22·1	23·0	12·0
Self	17·4	20·4	20·0	20·0	18·4	18·6	18·6	15·3	16·5	18·0	16·3	18·0	12·8	15·5	18·0
Sign	+	–	+	+	+	+	+	+	+	+	+	+	+	+	–

$H_0 : \eta_d = 0$, $H_1 : \eta_d > 0$, and the distribution under H_0 is **1**

$B(15, 0{\cdot}5) \sim N(7{\cdot}5, 1{\cdot}94^2)$ **1**

The difference in height of each pair is positive in 13 cases and negative in the other 2. **1**

$$P(B \leqslant 2) = P\left(Z \leqslant \frac{2{\cdot}5 - 7{\cdot}5}{1{\cdot}94}\right) = P(Z \leqslant -2{\cdot}58) = 0{\cdot}005$$ **1,1**

Since the p-value is less than 0·01 the null hypothesis is rejected at the 1% level **1**
of significance so Darwin's contention is strongly supported by the data. **1**

A6. (a) Number the accidents from 1 to 793 and **1**
generate 120 different random integers from 1 to 793. **1**

(b) H_0: there is no association between wearing a helmet and avoiding a head injury. **1**
H_1: there is an association.

	No head injury	*Head injury*
No helmet	7 (13·6)	16 (9·4)
Helmet	64 (57·4)	33 (39·6)

1

Table plus observed frequencies
Expected frequencies in brackets

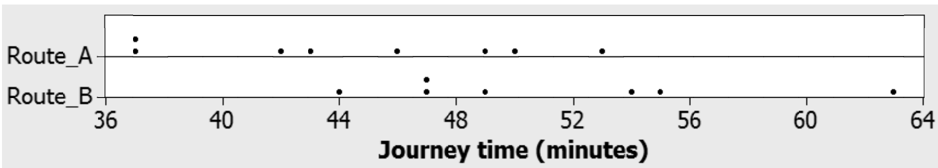
$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(7 - 13{\cdot}6)^2}{13{\cdot}6} + \frac{(16 - 9{\cdot}4)^2}{9{\cdot}4} + \frac{(64 - 57{\cdot}4)^2}{57{\cdot}4} + \frac{(33 - 39{\cdot}6)^2}{39{\cdot}64}$$

$$\approx 3{\cdot}20 + 4{\cdot}63 + 0{\cdot}76 + 1{\cdot}10 = 9{\cdot}69$$ **1**

The critical value of chi-squared with one degree of freedom
at the 1% level of significance is 6·635 so, since 9·7 exceeds this value, **1**
evidence is provided that helmets reduce the risk of sustaining a head injury. **1**

A7. (a)



Scales, labels, accuracy **1,1,1**

(b) The ranks for the Route B times are:

1 2 3 6.5 8·5 8·5 11

with rank sum 40·5. **1**

H_0 : Median on route A = Median on route B
H_1 : Median on route A < Median on route B **1**

$$W - \tfrac{1}{2}n(n+1) = 40{\cdot}5 - 28 = 12{\cdot}5$$ **1**

$$\mathrm{P}\left(W - \tfrac{1}{2}n(n+1) < 12{\cdot}5\right)$$

$$\approx \frac{267}{6435} \text{ (interpolation expected)}$$ **1**

$$= 0{\cdot}041$$

Since this p-value is less than 0.05 the null hypothesis is
rejected at the 5% level of significance. **1**
Thus the data provide evidence in support of the claim made by commuter's wife. **1**

A8. (a)

$$S_{xx} = 28$$
$$S_{xy} = -7{\cdot}76$$
$$b = \frac{S_{xy}}{S_{xx}} = -0{\cdot}2771 \qquad \mathbf{1}$$
$$a = \bar{y} - b\bar{x} = 5{\cdot}5086$$
$$y = 5{\cdot}5086 - 0{\cdot}2771x \qquad \mathbf{1}$$

$$S_{yy} = 215{\cdot}67 - \frac{38{\cdot}56^2}{7} = 3{\cdot}2595$$
$$SSR = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 3{\cdot}2595 - \frac{(-7{\cdot}76)^2}{28} = 1{\cdot}10887 \qquad \mathbf{1}$$
$$\Rightarrow s = \sqrt{\frac{1{\cdot}10887}{5}} = 0{\cdot}4709 \qquad \mathbf{1}$$

$$t = \frac{b}{\frac{s}{\sqrt{S_{xx}}}} = \frac{-0{\cdot}2771}{\frac{0.4709}{\sqrt{28}}} = -3{\cdot}11 \qquad \mathbf{1}$$

with 5 degrees of freedom.
The critical region for 5% significance is $|t| > 2{\cdot}571$ so
since $-3{\cdot}11$ lies in this region the null hypothesis of a non-zero slope is rejected; the data provide evidence of a non-zero slope. **1**

(b)

$$\hat{Y}_i \pm t_{n-2,1-\frac{1}{2}\alpha}s\sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}}$$

$$(5{\cdot}5086 - 0{\cdot}2771 \times 6) \pm 2{\cdot}571 \times 0{\cdot}4709\sqrt{1 + \frac{1}{7} + \frac{(6 - 0)^2}{28}} \qquad \mathbf{1}$$

$$3{\cdot}846 \pm 1{\cdot}887 = 1{\cdot}959,\ 5{\cdot}733 \qquad \mathbf{1}$$

We are 95% confident that the interval $(1{\cdot}959,\ 5{\cdot}733)$ contains the Extent of Sea Ice in 2011. **1**
But, since this involves extrapolation, we are assuming that the relationship continues to be linear. **1**

A9. (a)

$$\text{Mean} = \frac{(713 \times 0 + 299 \times 1 + 66 \times 2 + 16 \times 3 + 1 \times 4)}{1095}$$

$$= \frac{483}{1095} = 0{\cdot}44 \qquad \textbf{1}$$

Expected frequency of one homicide per day

$$= 1095 \times \frac{e^{-0{\cdot}44} \times 0{\cdot}44^1}{1!} = 310{\cdot}3 \qquad \textbf{1,1}$$

Expected frequency of 5 or more homicides per day

$$= 1095 - (705{\cdot}2 + 310{\cdot}3 + 68{\cdot}3 + 10{\cdot}0 + 1{\cdot}1)$$

$$= 0{\cdot}1 \qquad \textbf{1}$$

(b) She must have used frequency categories
0, 1, 2 and **3 or more** so that no expected frequency **1**
was less than 5. Degrees of freedom would then be
calculated as 4 − No. of parameters estimated −1 = 2. **1**

The critical value of chi-squared for significance level 5%
is 5·991. Since 3·58 is less than this value there is no **1**
evidence against the data fitting a Poisson distribution. **1**

(c) The expected number of homicides in 2008 is: -

$$366 \times 0{\cdot}44 = 161 \qquad \textbf{1}$$

Although the observed number of 152 is less than the
expected number, it still lies within the prediction limits **1**
so the city would not appear to be becoming safer. **1**

[END OF SECTION A]

AH Applied Mathematics 2011

Section B Solutions

B1. (a)

$$f(x) = \frac{1 + \sin x}{1 + 2\sin x}$$

$$f'(x) = \frac{\cos x(1 + 2\sin x) - (1 + \sin x)2\cos x}{(1 + 2\sin x)^2}$$ **1M,1**

$$= \frac{\cos x(1 + 2\sin x - 2 - 2\sin x)}{(1 + 2\sin x)^2}$$

$$= \frac{-\cos x}{(1 + 2\sin x)^2}$$ **1**

(b)

$$g(x) = \ln\left(1 + e^{2x}\right)$$

$$g'(x) = \frac{2e^{2x}}{1 + e^{2x}}$$ **1M,1**

B2.

$$A^{-1} = \frac{1}{6}\begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix}$$ **1 for determinant**

1 for elements

$$B = A^{-1}(AB)$$ **1M**

$$= \frac{1}{6}\begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix}\begin{pmatrix} -4 & -3 \\ 6 & -3 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 12 & -6 \\ 18 & 6 \end{pmatrix}$$ **1**

$$= \begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}$$ **1**

B3.

$$x = 5\cos t \Rightarrow \frac{dx}{dt} = -5\sin t$$ **1**

$$y = 3\sin t \Rightarrow \frac{dy}{dt} = 3\cos t$$ **1**

$$\frac{dy}{dx} = \frac{\frac{dy}{dt}}{\frac{dx}{dt}} = \frac{3\cos t}{-5\sin t}$$ **1**

$$= -\frac{3\sqrt{3}}{5} \text{ when } t = \frac{\pi}{6}.$$ **1**

$$\approx -1{\cdot}04$$

B4.

$$\sum_{r=1}^{N} r = \frac{N(N+1)}{2} = 210 \qquad \textbf{1M}$$

$$N^2 + N - 420 = 0 \Rightarrow (N-20)(N+21) = 0 \qquad \textbf{1}$$

$$\Rightarrow N = 20. \qquad \textbf{1}$$

$$\sum_{r=1}^{N} r^2 = \frac{N(N+1)(2N+1)}{6} \qquad \textbf{1}$$

$$= \frac{N(N+1)}{2}\frac{(2N+1)}{3}$$

$$\sum_{r=1}^{20} r^2 = 2870. \qquad \textbf{1}$$

B5.

$$u = \ln x \Rightarrow du = \frac{dx}{x} \qquad \textbf{1}$$

$$\int \frac{2}{x\ln x}\,dx = \int \frac{2}{u}\,du \qquad \textbf{1}$$

$$= 2\ln u + c \qquad \textbf{1}$$

$$= 2\ln(\ln x) + c \qquad \textbf{1}$$

B6.

$$\frac{dy}{dx} = 4 - \frac{3y}{x} \qquad \textbf{1M}$$

$$\frac{dy}{dx} + \frac{3}{x}y = 4 \qquad \textbf{1}$$

The integrating factor is

$$e^{\int \frac{3}{x}dx} = e^{3\ln x} \qquad \textbf{1}$$

$$= e^{\ln(x^3)} = x^3 \qquad \textbf{1}$$

Multiplying gives

$$x^3\frac{dy}{dx} + 3x^2y = 4x^3 \qquad \textbf{1}$$

$$\frac{d}{dx}(x^3y) = 4x^3$$

$$x^3y = \int 4x^3\,dx \qquad \textbf{1}$$

$$= x^4 + c$$

$$y = x + \frac{c}{x^3} \qquad \textbf{1}$$

As (1,3) is on the curve,

$$3 = 1 + \frac{c}{1^3} \Rightarrow c = 2 \qquad \textbf{1}$$

so an equation for the curve is:

$$y = x + \frac{2}{x^3} \qquad \textbf{1}$$