SCOTTISH
QUALIFICATIONS
AUTHORITY

2006 Applied Mathematics

Advanced Higher – Statistics

Finalised Marking Instructions

General Marking Principles

These principles describe the approach taken when marking Advanced Higher 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, is 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, full credit will be given for an alternative valid method.

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. M indicates a method mark, so in question 1, 1M, 1, 1 means a method mark for the product rule (and then a mark for each of the terms). E is shorthand for error. So for example, 2E1, means that a correct answer is awarded 2 marks but that 1 mark is deducted for each error.

Advanced Higher Applied Mathematics 2006
Marking Instructions
Section A – Statistics

A1. Let V denote 'contains a void' and let S denote 'signal'.

(a) $P(V) = 0{\cdot}06;\ P(S) = 0{\cdot}05;\ P(V \text{ and } S) = 0{\cdot}04$ **1**

so that $P(V \text{ and } S) \neq P(V)P(S)$. **1**

(b)

$$P\left(\bar{V} \mid S\right) = \frac{0{\cdot}01}{0{\cdot}01 + 0{\cdot}04} \qquad \mathbf{1}$$

$$= \frac{1}{5} \qquad \mathbf{1}$$

(c) $P\left(\bar{S} \mid V\right) = \dfrac{1}{3}$ **1**

A2. In stratified sampling the population of interest is divided into relatively homogeneous sub-populations or strata and simple random samples are selected from each stratum. **1**

In cluster sampling the population of interest is subdivided into convenient clusters of population members. A simple random sample of clusters is selected and then data is obtained for some or all population members within the clusters selected. **1**

Stratified sampling can lead to more precise population estimates when variability differs widely between the sub-populations. **1**

Cluster sampling can be advantageous when clusters of population members are widely separated geographically. **1**

A3. We require the smallest integer n such that

$$P(X \leqslant n) > 0{\cdot}9 \qquad \mathbf{1}$$

where $X \sim \text{Poi}(1{\cdot}8)$. **1**

x	$P(X = x)$	$P(X \leqslant x)$
0	0·1653	0·1653
1	0·2975	0·4628
2	0·2678	0·7306
3	0·1607	0·8913
4	0·0723	0·9636

1,1

The tabulation indicates that $n = 4$. **1**

The manager should aim to have 4 bottles in stock.

A4. The assumptions required are:–

the chosen companies are independent (or the portfolio is a random sample) **1**

and there is a constant probability $p = 0{\cdot}2$ of a drop in share price **1**

The number of shares which drop will then be

$X \sim \text{Bin}(40,\ 0{\cdot}2)$ **1**

and this can be approximated by

$X \sim \text{N}(8,\ 6{\cdot}4)$ **1**

$$P(X \leqslant 5) \sim P\left(Z \leqslant \frac{5{\cdot}5 - 8}{\sqrt{6{\cdot}4}}\right) \qquad \mathbf{1}$$

$$= P(Z \leqslant -0{\cdot}99)$$

$$= 0{\cdot}1611 \qquad \mathbf{1}$$

Notes: (a) the continuity correction is required.

(b) a direct calculation of Bin (40, 0·2) leading to 0·1613 is acceptable.

A5. All expected frequencies are $1200 \times \frac{1}{6} = 200$. **1**

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

$$x^2 = \frac{(185 - 200)^2}{200} + \frac{(191 - 200)^2}{200} + \frac{(200 - 200)^2}{200} + \frac{(208 - 200)^2}{200} + \frac{(209 - 200)^2}{200} + \frac{(207 - 200)^2}{200}$$

$$= 1{\cdot}125 + 0{\cdot}405 + 0 + 0{\cdot}32 + 0{\cdot}405 + 0{\cdot}245$$

$$= 2{\cdot}5 \qquad \mathbf{1}$$

$$X^2 \sim \chi_5^2$$

i.e. there are 5 degrees of freedom. **1**

As $2{\cdot}5 < 11{\cdot}07$ (5% critical value of chi-squared) **1**

there is no evidence to doubt the fit to a uniform distribution. **1**

A6. (a) For 99% confidence, $z = 2{\cdot}58$. **1**

The required confidence interval is given by:–

$$\hat{p} \pm 2.58\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \qquad \mathbf{1}$$

with $\hat{p} = \frac{3127}{5000}$ and $n = 5000$ **1**

$$= (0{\cdot}608,\ 0{\cdot}643) \qquad \mathbf{1}$$

(b) Since 0·60 does not lie within the confidence interval **1**

there is evidence against the governor's claim. **1**

A7. (a) Assuming that "door-to-needle" time is normally distributed **1**
and the sample is random. **1**

$H_0 : \mu = 50$

$H_1 : \mu < 50$ **1**

The critical region for a 1-tail z-test at the 5% level of significance is $z < -1{\cdot}64$. **1**

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{42 - 50}{12/\sqrt{10}} = -2{\cdot}11. \quad \mathbf{1}$$

Since $-2{\cdot}11 < -1{\cdot}64$ the null hypothesis would be rejected **1**
so the data provide evidence of reduced mean "door-to-needle" time. **1**

(b) The p-value of the test is $P(Z < -2{\cdot}11) = 0{\cdot}0174$ **1**
Since the p-value is less than 0·05, the conclusion is confirmed, **1**

(c) For example: data on survival rates of patients; recovery time; compare with other treatment centres. **1**

A8. (a) The process mean $\mu = 10{\cdot}0$ **1**

$$\mathrm{C_p} = 1 \Rightarrow \frac{\mathrm{USL} - \mathrm{LSL}}{6\sigma} = 1 \quad \mathbf{1}$$

$$\Rightarrow \sigma = \frac{0{\cdot}1}{6} = 0{\cdot}01667 \quad \mathbf{1}$$

$$P(9{\cdot}95 < \text{length} < 10{\cdot}05) = P\left(\frac{9{\cdot}95 - 10{\cdot}00}{0{\cdot}01667} < Z < \frac{10{\cdot}05 - 10{\cdot}00}{0{\cdot}01667}\right) \quad \mathbf{1}$$

$$= P(-3 < Z < 3)$$

$$= 0{\cdot}9974 \quad \mathbf{1}$$

P(Failure to meet specifications)

$$= 1 - 0{\cdot}9974 = 0{\cdot}0026$$

$$= 2600 \ \text{PPM} \quad \mathbf{1}$$

(b) $P(\text{Length} > \text{USL}) = 600/10^6 = 0{\cdot}0006$ **1**

$$\Rightarrow \Phi\left(\frac{10{\cdot}05 - 10{\cdot}00}{\sigma}\right) = 0{\cdot}9994 \quad \mathbf{1}$$

$$\Rightarrow \frac{0{\cdot}05}{\sigma} = 3{\cdot}24 \quad \mathbf{1}$$

$$\Rightarrow \sigma = 0{\cdot}0154 \quad \mathbf{1}$$

A9. (a) There are $n = 6$ Dry and $m = 8$ Wet days. **1**

Back-to-back stem and leaf display of the data

$$\begin{array}{r|c|l} \text{dry} & & \text{wet} \\ 7 & 4 & 5 \\ 20 & 5 & 144 \\ 975 & 5 & 6 \\ & 6 & 011 \end{array}$$ **1**

(b) $H_0 : \eta_W = \eta_D$

$H_1 : \eta_W > \eta_D$ **1**

Duration	45	47	50	51	52	54	54	55	56	57	59	60	61	61
Weather	W	D	D	W	D	W	W	D	W	D	D	W	W	W
Rank	1	2	3	4	5	6·5	6·5	8	9	10	11	12	13·5	13·5

Rank sum for Dry

$$W = 2 + 3 + 5 + 8 + 10 + 11 = 39$$ **1**

$$W - \tfrac{1}{2}n(n + 1) = 39 - 21 = 18$$ **1**

$$P(W - \tfrac{1}{2}n(n + 1) < 18)$$

$$= \frac{737}{3003}$$

$$= 0{\cdot}245$$ **1**

Since this value exceeds 0.05 the null hypothesis cannot be rejected at the 5% level of significance. **1**

Thus the driver's hunch is not supported by the data. **1**

A10. (a)

$$t = \frac{b}{s/\sqrt{S_{xx}}} = \frac{0{\cdot}851}{4{\cdot}895/\sqrt{15684{\cdot}4}} = 21{\cdot}8$$ **1**

The critical region (33 d.f.) is $|t| > 3{\cdot}611$. **1**

Since t lies in the critical region,

the null hypothesis is rejected at the 0·1% significance level. **1**

(b) $Y_i = 152{\cdot}193 + 0{\cdot}851 \times 1165 = 1143{\cdot}6$ **1**

The required prediction interval is given by:–

$$Y_i \pm ts\sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}}$$ **1**

$$= 1143{\cdot}6 \pm 2{\cdot}733 \times 4{\cdot}895\sqrt{1 + \frac{1}{35} + \frac{(1165 - 1149{\cdot}6)^2}{15684{\cdot}4}}$$ **1**

$$= 1143{\cdot}6 \pm 2{\cdot}733 \times 4{\cdot}895 \times 1{\cdot}0216$$

$$= 1143{\cdot}6 \pm 13{\cdot}7 \quad \text{or} \quad (1129{\cdot}9,\ 1157{\cdot}3)$$ **1**

(c) The prediction interval contains the desired temperature of 1150° so it would be reasonable to advise pouring. **1**

(d) Since $(x_i - \bar{x})^2$ is minimised when the values in brackets are equal, the required value is 1149·6. **1**

Section B – Mathematics for Applied Mathematics

B1.

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 2 & 3 & 1 & 0 & 1 & 0 \\ 2 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right)$$

$$\to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -2 & 0 & 1 \end{array}\right)$$ **M1, 2E1**

$$\text{So } A^{-1} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ -2 & 0 & 1 \end{pmatrix}.$$

Other valid methods of obtaining A^{-1} will be accepted.

$$\begin{array}{ccccccc} x & + & y & & & = & 1 \\ 2x & + & 3y & + & z & = & 2 \\ 2x & + & 2y & + & z & = & 1 \end{array}$$

$$A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^{-1}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ -2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$ **M1,1**

so $x = 0, y = 1, z = -1$.

B2.

$$y = \ln(1 + \sin x)$$

$$\frac{dy}{dx} = \frac{\cos x}{1 + \sin x}$$ **M1,1**

$$\text{so } \frac{d^2y}{dx^2} = \frac{(1 + \sin x)(-\sin x) - \cos x \cos x}{(1 + \sin x)^2}$$ **M1,1**

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

$$= \frac{-1}{(1 + \sin x)}.$$

B3.

$$S_n = \tfrac{1}{6}n(n + 1)(2n + 1)$$ **1**

$$S_{2n+1} = \tfrac{1}{6}(2n + 1)(2n + 2)(4n + 3)$$ **1**

$$2^2 + 4^2 + \ldots + (2n)^2 = 4(1^2 + 2^2 + \ldots + n^2)$$

$$= \tfrac{2}{3}n(n + 1)(2n + 1)$$ **1**

B4.

$$\cos^2 y \frac{dy}{dx} = y$$

$$\int \frac{dy}{y} = \int \sec^2 x \; dx \qquad \textbf{M1}$$

$$\text{so} \qquad \ln y = \tan x + c. \qquad \textbf{1,1}$$

When $y = 2, x = 0$ giving $c = \ln 2$. **1**

Hence $\ln y - \ln 2 = \tan x$, i.e. $\ln \frac{1}{2}y = \tan x$

$$\Rightarrow \; y = 2e^{\tan x}. \qquad \textbf{1}$$

B5. $1 + x^2 = u \Rightarrow x\,dx = \frac{1}{2}\,du$ so **1**

$$\int \frac{x^3}{\sqrt{1 + x^2}}\,dx = \int \frac{(u - 1)}{\sqrt{u}} \frac{1}{2}\,du \qquad \textbf{1}$$

$$= \frac{1}{2} \int \left(u^{1/2} - u^{-1/2}\right) du \qquad \textbf{1}$$

$$= \frac{1}{3} u^{3/2} - u^{1/2} + c \qquad \textbf{1}$$

$$= \frac{1}{3}\left(1 + x^2\right)^{3/2} - \left(1 + x^2\right)^{1/2} + c \qquad \textbf{1}$$

$$= \frac{1}{3}\left(x^2 - 2\right)\sqrt{1 + x^2} + c$$

B6. (a)

$$\int_0^1 x\,e^{2x}\,dx = \left[x \int e^{2x}\,dx - \int \tfrac{1}{2}\,e^{2x}\,dx\right]_0^1 \qquad \textbf{M1, 1}$$

$$= \left[\tfrac{1}{2}\,x\,e^{2x} - \tfrac{1}{4}\,e^{2x}\right]_0^1 \qquad \textbf{1}$$

$$= \tfrac{1}{2}\,e^2 - \tfrac{1}{4}\,e^2 + \tfrac{1}{4} = \tfrac{1}{4}\left(e^2 + 1\right) \qquad \textbf{1}$$

(b)

$$\int_0^1 x^2\,e^{2x}\,dx = \left[x^2 \int e^{2x}\,dx\right]_0^1 - \int_0^1 2x.\tfrac{1}{2}\,e^{2x}\,dx \qquad \textbf{1}$$

$$= \left[\tfrac{1}{2}x^2 e^{2x}\right]_0^1 - \int_0^1 x\,e^{x}\,dx \qquad \textbf{1}$$

$$= \left[\tfrac{1}{2}\,e^2 - 0\right] - \tfrac{1}{4}\left(e^2 + 1\right) = \tfrac{1}{4}\left(e^2 - 1\right) \qquad \textbf{1}$$

(c)

$$\int_0^1 \left(3x^2 + 2x\right) e^{2x}\,dx = 3\int_0^1 x^2\,e^{2x}\,dx + 2\int_0^1 x\,e^{2x}\,dx \qquad \textbf{1}$$

$$= \tfrac{3}{4}\left(e^2 - 1\right) + \tfrac{2}{4}\left(e^2 + 1\right) \qquad \textbf{1}$$

$$= \tfrac{1}{4}\left(5e^2 - 1\right)$$

[END OF MARKING INSTRUCTIONS]