X202/701

NATIONAL QUALIFICATIONS 2011

FRIDAY, 20 MAY
1.00 PM – 4.00 PM

APPLIED MATHEMATICS
ADVANCED HIGHER
Statistics

Read carefully

1. Calculators may be used in this paper.
2. Candidates should answer all questions.

   Section A assesses the Units Statistics 1 and 2
   Section B assesses the Unit Mathematics for Applied Mathematics
3. **Full credit will be given only where the solution contains appropriate working.**
4. A booklet of Statistical Formulae and Tables is supplied for all candidates.

Section A (Statistics 1 and 2)

Marks

Answer all the questions.

A1. The population of a city with one million inhabitants includes 100 known terrorists whose images are stored in a surveillance system database.

A surveillance camera in the system has probability 0·999 of signalling an alarm when a known terrorist is scanned and probability 0·001 of signalling an alarm when anyone else is scanned.

A report claims, mistakenly, that one could be 99·9% sure that a person signalling an alarm via the camera was one of the known terrorists.

(*a*) Stating an assumption required, calculate the correct percentage. **4**

(*b*) State a condition under which the percentage could be 99·9%. **1**

A2. In the manufacture of sterile hypodermic syringes for single use, an international standard specifies that the volume of drug for a nominal 10 ml dose should lie between 9·6 ml and 10·4 ml. The machine that fills the syringes delivers volumes of the drug with standard deviation 0·16 ml. Following the setting up of the machine for a long production run, a quality control check yielded a random sample of 50 syringes with mean drug volume 9·978 ml.

(*a*) Demonstrate that the data provide no evidence that the mean amount of drug being delivered to the syringes differs from the target value of 10·0 ml and indicate why a two-tail alternative hypothesis is appropriate. **5**

(*b*) Calculate the value to which the standard deviation would have to be reduced in order to ensure that 99·9% of syringes contained amounts of drug between the specified limits. **2**

A3. A random sample of 10 patients is taken from a population of patients experiencing difficulties in sleeping. Given a dose of a sleep-inducing drug, the additional hours of sleep gained (x) were:

0·7 –1·6 –0·2 –1·2 –0·1 3·4 3·7 0·8 0·0 2·0

In this context the researchers were interested in the null hypothesis H_0: $\mu = 0$ and alternative hypothesis H_1: $\mu > 0$. On the assumption that the additional hours of sleep gained are normally distributed, a $100(1 - \alpha)\%$ one-sided confidence interval for the population mean is given, in standard notation, by

$$\left(\bar{x} - t_{n-1,1-\alpha}\,\frac{s}{\sqrt{n}}, \infty\right).$$

Obtain the 95% confidence interval in this case and state what may be concluded concerning the hypotheses and the effectiveness of the drug. **5**

Marks

A4. The salaries of employees of a global IT company have mean £25 000 with standard deviation £1500. In the coming financial year, the company will pay each employee a Christmas bonus of £1000 and an incentive bonus of 10% of current salary.

(*a*) Calculate the mean and standard deviation of individual employee earnings from the company in the coming year. **3**

(*b*) Estimate the probability that the total earnings in the coming year of a random sample of 36 employees would exceed £1 000 000. **4**

A5. From 1858 onwards, Charles Darwin published results of a series of experimental studies into the reproductive biology of various plant species. He contended that cross-pollination of plants yielded offspring that were taller than those produced by self-pollination. He obtained the data below on the heights of pairs of plants, each pair being grown under identical conditions and with one member of each pair, chosen at random, being bred via cross-pollination and the other being bred via self-pollination.

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

Use the normal approximation to the binomial distribution to estimate the p-value for an appropriate hypothesis test. Evaluate the evidence from the data in support of Darwin's contention. **7**

A6. Every year in Scotland each police authority completes a statistical return that provides details of the circumstances for all the accidents in its own area, involving injury, between a motor vehicle and a bicycle, over the past five years.

(*a*) Given that you have access to a database containing the details of all 793 accidents involving injury between a motor vehicle and a bicycle in one such area over the past five years, outline the steps involved in selecting a random sample of 120 accidents for detailed study. **2**

From a random sample of 120 accidents it was established that, of those cyclists wearing helmets, 33 sustained a head injury and 64 did not. Amongst those not wearing helmets, 16 sustained a head injury.

(*b*) Display the data in a contingency table. Carry out an appropriate statistical test on this data set and report your findings in terms of the benefit to cyclists of wearing a helmet. **5**

[Turn over

Marks

A7. A commuter travels to work by car using one of two routes, A and B, and his wife claims that route A is generally faster than route B. For three weeks the commuter selected the route to take by tossing a coin, using route A if the outcome was a head and route B otherwise. Journey times (minutes) were:

Route A	46	37	53	50	42	43	49	37
Route B	54	63	44	49	47	55	47	

(*a*) Display the data. **3**

(*b*) Use the Mann-Whitney table to investigate whether or not his wife's claim is supported by the data. **6**

A8. The National Snow and Ice Data Center at the University of Colorado, USA, published the following data on the Extent of Sea Ice (y), in millions of square kilometres, for the years 2002–2008.

Year	2002	2003	2004	2005	2006	2007	2008
Extent (y)	5·96	6·15	6·04	5·57	5·89	4·28	4·67

The variable x is defined by x = (Year–2005).

(*a*) Assuming that a scatterplot of the data shows that a linear fit is appropriate, obtain an equation of the regression line of y on x and show that the slope of this line differs significantly from zero. **6**

(*b*) Calculate a 95% prediction interval for Extent of Sea Ice in 2011. Give an interpretation and comment. **4**

Marks

A9. In London, on 1095 days between April 2004 and March 2007, the observed number of homicides per day are given in the table. Also given are some expected frequencies for a fitted Poisson distribution.

Homicides	0	1	2	3	4	5 or more
Observed	713	299	66	16	1	0
Expected	705·2		68·3	10·0	1·1	

(*a*) Calculate, correct to two decimal places, the mean number of homicides per day and hence the two missing expected frequencies. **4**

(*b*) In analysing the data, a statistician computed a chi-squared goodness-of-fit test statistic to be 3·58 with 2 degrees of freedom. Outline the steps she would have taken in order to obtain these values and state, with justification, what may be concluded concerning the fit of the Poisson distribution. **4**

(*c*) Other statisticians used the above data to calculate 95% prediction limits of 137 and 187 for the total number of homicides in 2008. Calculate the expected number of homicides in 2008 on the basis of the given data and, given that there were 152 homicides in London in 2008 (which was a leap year), comment on whether or not there is any evidence that the city was becoming safer. **3**

[*END OF SECTION A*]

[Turn over for Section B on *Pages six and seven*

Section B (Mathematics for Applied Mathematics)

Marks

Answer all the questions.

B1. Differentiate the following functions, simplifying where possible:

(*a*) $f(x) = \dfrac{1+\sin x}{1+2\sin x}$, $0 \le x \le \pi$; **3**

(*b*) $g(x) = \ln(1+e^{2x})$. **2**

B2. Given $A = \begin{pmatrix} 1 & -2 \\ 3 & 0 \end{pmatrix}$, obtain A^{-1}.

Given $AB = \begin{pmatrix} -4 & -3 \\ 6 & -3 \end{pmatrix}$, find the matrix B. **5**

B3. A curve is defined by the equations

$$x = 5\cos t \qquad \text{and} \qquad y = 3\sin t, \qquad (0 \le t < 2\pi).$$

Find the gradient of the curve when $t = \dfrac{\pi}{6}$. **4**

B4. Find the value of N for which $\displaystyle\sum_{r=1}^{N} r = 210$. **3**

Evaluate $\displaystyle\sum_{r=1}^{N} r^2$ for this value of N. **2**

Marks

B5. Use the substitution $u = \ln x$ to obtain $\int \frac{2}{x \ln x} dx$, where $x > 1$. **4**

B6. At any point (x, y) on a curve C, where $x \neq 0$, the gradient of the tangent is $4 - \frac{3y}{x}$.
Given that the point $(1, 3)$ lies on C, obtain an equation for C in the form $y = f(x)$. **9**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]

[BLANK PAGE]