X202/13/01

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. A nurse who applied for a promoted post asked the charge nurse on his ward for a reference. He expected to be given a reference that was positive, neutral or negative with respective probabilities 0·6, 0·3 and 0·1. He estimated that with a positive reference from the charge nurse he had probability 0·55 of securing the post, with a neutral reference he had probability 0·25 and that with a negative reference he had probability of only 0·05 of securing the post.

    (*a*) Calculate the probability that he secured the post. **2**

    (*b*) Given that he failed to secure the post, calculate the probability that he was given a positive reference by the charge nurse. **4**

A2. You have decided to visit and interview a random sample of Scottish sixth year pupils in order to research such things as their fitness and lifestyle. When you take a random sample of these pupils, there are a total of 376 secondary schools throughout 8 regions of Scotland.

    (*a*) Give a reason why the taking of a simple random sample of sixth year pupils in Scottish secondary schools would not be readily feasible. **1**

    (*b*) Explain how you would proceed to take a cluster sample. **2**

    (*c*) Explain how you would proceed to take a stratified sample by region. **1**

    (*d*) State an advantage of each of cluster and stratified sampling. **2**

A3. In the Caucasian segment of the population of the USA the blood groups O, A, B and AB occur in the proportions 45%, 40%, 11% and 4% respectively. A random sample of people of Hawaiian ethnic origin included 190 people with group O, 249 with group A, 18 with group B and 10 with group AB. Carry out a goodness-of-fit test and comment. **6**

A4. A gardener grew Midlothian Early potatoes with either compost or National Growmore as fertiliser. The weights of tubers (g) yielded by random samples of 8 plants grown with each of the fertilisers are displayed in the back-to-back stem-and-leaf display with a leaf unit of 1 g.

Compost		Growmore
1	54	
8 5	54	
3 1 1	55	2 3
5	55	5 7
0	56	1 3
	56	6
	57	0

By performing a non-parametric test, determine whether or not the data provide evidence, at the 5% level, that median yields differ for the two fertilisers.

6

A5. In May 2012 an opinion poll commissioned by the Scottish anti-independence campaign found that 33% of a sample of 1004 Scots wished to achieve independence from the United Kingdom. An approximate 95% confidence interval for the population proportion may be written as $\hat{p} \pm M$ where

$M = 1{\cdot}96\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$ may be referred to as the margin of error.

(*a*) Stating any assumption required, obtain an approximate 95% confidence interval for the proportion of the population of Scots who wished to achieve independence.

3

(*b*) Obtain the maximum value of the expression $x(1-x)$ and hence show that the maximum margin of error possible, with 95% confidence, in a poll of sample of n respondents is approximately $1/\sqrt{n}$.

3

(*c*) With no prior knowledge of a population proportion available, determine the sample size necessary to ensure a 95% confidence interval with a margin of error of at most 1%.

2

[Turn over

A6. Automotive components manufactured from an aluminium-silicon alloy have, on average, tensile strength 540 N/mm². A metallurgist believed that the addition of calcium during casting would alter the mean tensile strength. A pilot batch of alloy made with calcium yielded a random sample of components with the following tensile strengths.

531   535   539   535   547   526   534   531   540   537

(*a*) Test, at the 5% level of significance, the null hypothesis $H_0: \mu = 540$ versus the alternative hypothesis $H_1: \mu \neq 540$ and state any assumption required. **5**

(*b*) A 95% confidence interval for the mean tensile strength calculated from the given data is (531·4, 539·6).

State precisely what may be concluded concerning the null hypothesis $H_0: \mu = 540$ from scrutiny of the confidence interval. **2**

A7. A business analyst wishes to model the time shoppers spend in a supermarket, in order to make planning decisions on checkout provision. Her initial model takes the time a shopper spends selecting goods, X minutes, to be uniformly distributed on the interval [10,50] and the time spent checking out, Y minutes, to be uniformly distributed on the interval [5,15].

(*a*) Assuming that X and Y are independent and that the mean and variance of a uniformly distributed random variable on the interval [a,b] are (a + b)/2 and (b–a)²/12 respectively, obtain the mean μ and standard deviation σ of the total time, T minutes, that a customer takes to shop. **4**

For k > 0, Cantelli's inequality states that for any continuous random variable V,

$$P\left(V \geq \mu_V + k\sigma_V\right) \leq \frac{1}{1+k^2} \ .$$

(*b*) Use this inequality to make a statement about the probability that a shopper, selected at random, spends more than an hour shopping. **2**

(*c*) State a reason why X and Y are unlikely to be independent and how you could carry out a check. **2**

A8. A country hosting a major international soccer tournament struck a special commemorative coin to be used by match referees and sold as a souvenir. Concern was expressed prior to the tournament that the coins were biased yielding heads from 55% of tosses. It was decided to select one of the coins at random and toss it 200 times. It was agreed to assess the coin biased, as suspected, if 110 or more heads were obtained and to assess the coin fair otherwise.

(*a*) If the coin is actually fair, use an appropriate approximation to estimate the probability that it is assessed biased in the way suspected. **6**

(*b*) Similarly, estimate the probability that a biased coin is assessed to be fair. **3**

(*c*) State a step that could be taken to reduce the risks of incorrect assessment of a coin as biased or unbiased. **1**
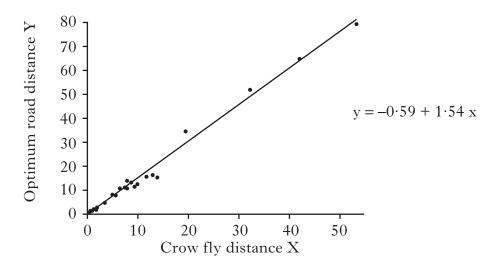
[Turn over

A9. During the run-up to the 2012 Olympic Games, the Royal Statistical Society Centre for Statistical Education created the teaching and learning resource *Getting to the Point in 2012*. The resource may be used to obtain both the direct (crow fly) distance (X) and the optimum road distance (Y) between locations.

It was used to determine both distances from each of a random sample of 25 Scottish schools to the nearest location at which the passing of the Olympic Torch could be viewed. For example, for Crieff High School the crow fly distance to view the torch at Kinross was 19·4 miles whereas the optimum distance by road was 34·7 miles. A scatter plot of the data together with the least squares regression line of y on x is displayed below.

$$y = -0.59 + 1.54\ x$$

The data provide very strong evidence that the slope parameter differs significantly from zero.

(*a*) Give an interpretation of the slope parameter estimate of 1·54 in terms of optimum road distance per additional mile of crow fly distance. **1**

(*b*) Suggest a reason why a similar analysis of data for 25 schools in an urban area yielded a smaller value of the slope parameter. **1**

(*c*) The Scottish data gave $\bar{x} = 10.55$, $S_{xx} = 4251.662$ and $s^2 = 4.669$.

Calculate both a 95% prediction interval and a confidence interval for the optimum road distances for Scottish schools with crow fly distances of 40 miles.

Give an interpretation of both intervals. **7**

(*d*) Analysis of the Scottish data yields a p-value of 0·297 for a two-tailed test of the null hypothesis that the intercept parameter α in the linear model is zero.

State the implication of this and why it is to be expected. **2**

[*END OF SECTION A*]

Answer all the questions

B1. Find the gradient of the tangent to the curve

$$y = 2x\sqrt{x-1}$$

at the point where $x = 10$.   **4**

B2. Matrices are given as

$$A = \begin{pmatrix} 1 & 3 & 4 \\ k & 0 & -1 \\ 5 & 3 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & -10 & 2 \\ -3 & 9 & 0 \\ 0 & -2 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 2 & -6 \\ 1 & 1 & -2 \\ 2 & 2 & -1 \end{pmatrix}.$$

(*a*) Calculate $A + B$.   **1**

(*b*) Find the determinant of A.   **2**

(*c*) Calculate BC.   **1**

(*d*) Describe the relationship between B and C.   **2**

B3. Find the exact value of $\displaystyle\int_0^{2\pi} x\sin 3x \; dx$.   **5**

B4. Evaluate $\displaystyle\sum_{r=1}^{80} 3r^2$.   **2**

B5. (*a*) Write down and simplify the binomial expansion of $(e^x + 2)^4$.   **3**

(*b*) Hence obtain $\displaystyle\int (e^x + 2)^4 \, dx$.   **2**

[Turn over for Question B6 on *Page eight*

B6. A flu-like virus starts to spread through the 20 000 inhabitants of Dumbarton.

The situation can be modelled by the differential equation

$$\frac{dN}{dt} = \frac{N(20\,000 - N)}{10\,000},$$

where N is the number of people infected after t days and $0 < N < 20\,000$.

(*a*) How many people are infected when the infection is spreading most rapidly?　　**1**

(*b*) Express $\dfrac{10\,000}{N(20\,000 - N)}$ in partial fractions and show that

$$\ln\left(\frac{N}{(20\,000 - N)}\right) = 2t + C, \text{ for some constant } C.$$　　**5**

Initially there were 100 people infected.

(*c*) Show that $N = \dfrac{20\,000\,e^{2t}}{199 + e^{2t}}$.　　**4**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]