X202/701

NATIONAL QUALIFICATIONS 2005

MONDAY, 23 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.

SCOTTISH QUALIFICATIONS AUTHORITY

Section A (Statistics 1 and 2)

Answer all the questions.

Marks

A1. An internet user classifies incoming e-mail messages as either spam (unsolicited commercial e-mail) or genuine. Let S represent the event that an incoming e-mail is spam and let C represent the event that an incoming e-mail contains the text string "click here". It is estimated that 40% of incoming messages are genuine, that $P(C \mid S) = 0{\cdot}75$ and that $P(C \mid \overline{S}) = 0{\cdot}05$. The user sets up a screening system for incoming e-mails whereby any message containing the text string "click here" is placed in a junk folder and all other messages are placed in the inbox folder. Estimate the probability that a message placed in the junk folder is genuine. **5**

A2. A nutritional expert in a country defines the 95% normal range for Body Mass Index (BMI) to be 18·0 to 26·0 (kg/m^2). She also considers people with BMI greater than 28·0 to be obese. Assuming that BMI is a normally distributed random variable in the country and that the normal range is symmetrically placed about the mean, calculate:

(*a*) the mean and standard deviation of the distribution; **4**

(*b*) the proportion of the population the expert would consider obese. **3**

A3. In a *Scotsman* article entitled "Sweet-toothed men are Britain's true chocoholics" published in July 2004, it was reported that "more than half (53%) of adults regularly snack on chocolate". The survey on which the article was based involved a sample of 1000 men and 531 members of the sample indicated that they snacked regularly on chocolate.

(*a*) Obtain an approximate 99% confidence interval for the population proportion of men who snack regularly on chocolate and comment on the quotation from the article. **4**

(*b*) State two potential problems in attempting to obtain accurate information on snacking habits from a representative sample of 1000 British men. **2**

(*c*) State an advantage and a disadvantage of carrying out the survey by telephone. **2**

A4. Following complaints from customers, an electronics manufacturing company initiated an improvement project with a view to reducing the mean cycle time for order processing (days). Analysis of the records for a large number of orders indicated that the cycle time for order processing could be adequately modelled by the $N(29, 3^2)$ distribution. Following completion of the project, the cycle times for a sample of orders were:

28 31 27 24 22 27 31 24 30 23.

Stating two assumptions required, evaluate the evidence that the project has been successful in reducing mean cycle time. **7**

Marks

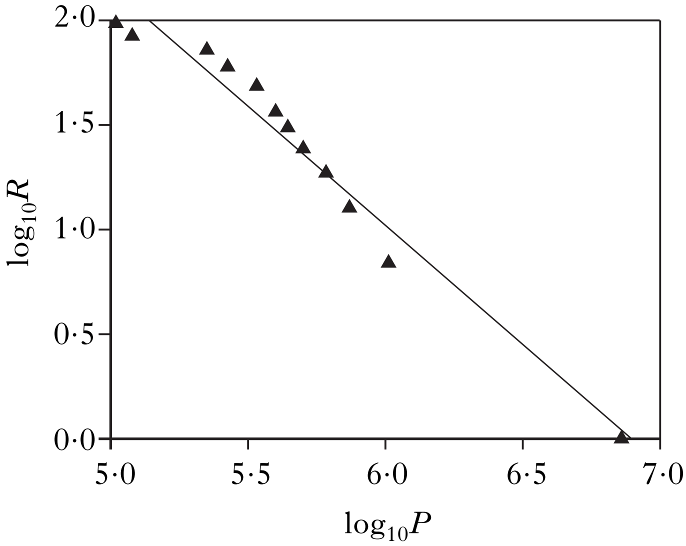
A5. The table below gives the 1994 population (P) and rank of population (R) of a sample of cities in the USA.

City	*Population* (P)	*Rank* (R)
New York	7 322 564	1
Detroit	1 027 974	7
Baltimore	736 014	13
Washington	606 900	19
New Orleans	496 938	25
Kansas	434 829	31
Virginia	393 089	37
Toledo	332 943	49
Arlington	261 721	61
Baton Rouge	219 531	73
Hialeah	117 647	85
Bakersfield	103 093	97

Let Y denote $\log_{10}R$ and X denote $\log_{10}P$. According to Zipf's Law, for the cities in a country, Y and X are related by an equation of the form $Y = k - X$ where k is a constant. The diagram shows annotated output from a least squares regression analysis of the data carried out using a statistical software package.

The equation of the regression line is

$$\log_{10}R = 7{\cdot}846 - 1{\cdot}138 \log_{10}P.$$



Given that the sum of squared residuals is 0·12454, that $S_{xx} = 2{\cdot}58309$ and that a 100 $(1 - \alpha)$% confidence interval for the slope parameter β is given by

$$b \pm t_{n-2,1-\frac{1}{2}\alpha} \frac{s}{\sqrt{S_{xx}}},$$

obtain a 95% confidence interval for the slope parameter and state whether or not Zipf's law is supported by the result. **6**

Marks

A6. A multiple-choice test has 12 questions, each of which is allocated four possible answers with only one being correct in each case. In order to pass the test a candidate has to answer at least 8 questions correctly. Calculate the probability that a candidate who answers all the questions passes, given that:

(*a*) the candidate has no knowledge of the topic; **3**

(*b*) the candidate has sufficient knowledge to discount two of the four answers to each question but has to guess between the remaining two alternatives. **2**

A7. During the Second World War, scientists divided an area of south London up into 576 squares, each with side of length 0·5 km. The table below gives the actual number of squares receiving 0, 1, 2, 3, . . . impacts of flying bombs and some of the expected numbers obtained by fitting a Poisson distribution.

Count x	*No of squares with x impacts*	*Expected no of squares with x impacts*
0	229	227·3
1	211	
2	93	
3	35	
4	7	7·1
5	1	1·3
6 or more	0	0·1

(*a*) Calculate, to 2 decimal places, the mean number of impacts per square and hence obtain the missing expected numbers. **5**

(*b*) Carry out a test of goodness-of-fit. **6**

(*c*) The scientists concluded from the result of the goodness-of-fit test that the flying bombs were not fitted with a sophisticated guidance system. Explain why this conclusion was reached. **1**

A8. Archaeologists determined the silver content (%Ag) of random samples of coins discovered in Cyprus from both the first and second mintages of the reign of King Manuel I. The data are given in the table below.

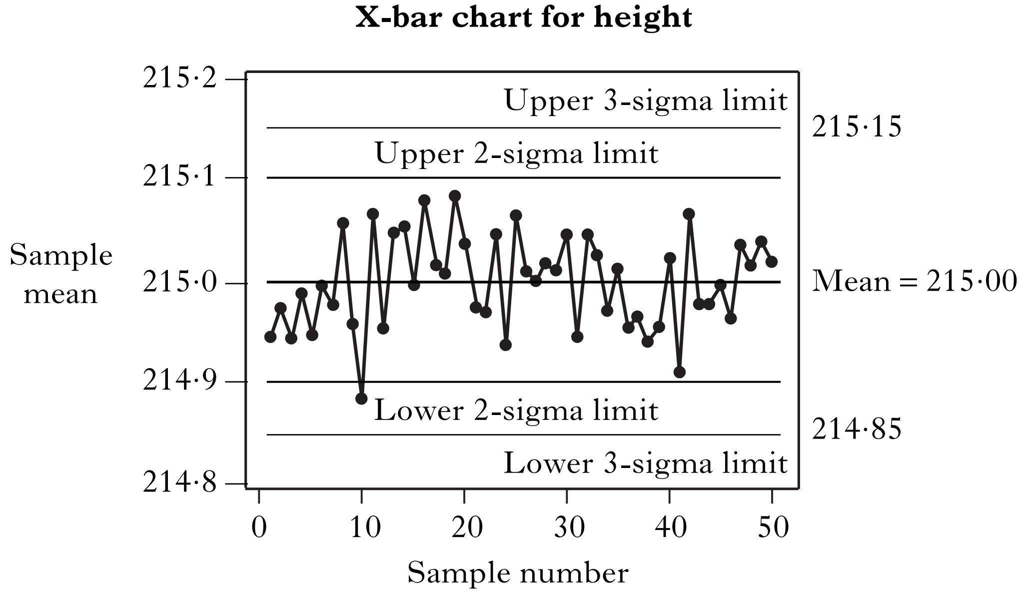
	Silver Content (%Ag)
First Mintage	5·9 6·8 6·4 7·0 6·6 7·7 7·2 6·9 6·2
Second Mintage	6·9 9·0 6·6 8·1 9·3 9·2 8·6

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

(*b*) Carry out an appropriate statistical test to compare the medians of the silver content for the two mintages. **7**

Marks

A9. The height of glass bottles is an important quality characteristic in a high-speed bottling plant as it is very costly to extract a shattered bottle during the manufacturing process. In order to monitor bottle height, a manufacturer takes random samples of four bottles from the production line every 15 minutes and measures their heights. The mean height is then plotted on a Shewhart chart for the sample mean. The chart below was created during a production run for bottles with a design height of 215 mm.



(*a*) Given that the estimated process mean and standard deviation were 215·00 and 0·10, confirm the 3-sigma limits displayed on the chart. **2**

(*b*) State, with justification, whether the quality manager would be satisfied with the above chart for the production run. **1**

(*c*) Given that during the production run the height was $N(215{\cdot}00,\ 0{\cdot}10^2)$, calculate the probability that the use of 3-sigma limits would lead to a "false alarm" signal for a special cause of variation. Calculate the probability of obtaining a point below the 2-sigma lower limit and state whether or not the occurrence of such a point should be of concern to the quality manager. **5**

[*END OF SECTION A*]

[Turn over for Section B on *Page six*

Section B (Mathematics for Applied Mathematics)

Marks

Answer all the questions.

B1. Differentiate, and simplify as appropriate,

(*a*) $f(x) = \exp(\tan \frac{1}{2}x)$, where $-\pi < x < \pi$, **3**

(*b*) $g(x) = (x^3 + 1)\ln(x^3 + 1)$, where $x > 0$. **3**

B2. Given that $A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}$, show that $A^2 - A = kI$ for a suitable value of k, where I is the 2×2 unit matrix. **3**

B3. A curve is defined by the parametric equations $x = 5t^2 - 5$, $y = 3t^3$.

Find the value of t corresponding to the point $(0, -3)$ and calculate the gradient of the curve at this point. **2, 3**

B4. Expand and simplify $\left(2a - \frac{3}{a}\right)^4$. **3**

B5. Express $\dfrac{x^2 + 3}{x(1 + x^2)}$ in partial fractions. **3**

Hence obtain $\displaystyle\int_{1/2}^{1} \frac{x^2 + 3}{x(1 + x^2)}\,dx$. **3**

B6. (*a*) Given the differential equation

$$\sin x \frac{dy}{dx} - 2y\cos x = 0,$$

find the general solution, expressing y explicitly in terms of x. **4**

(*b*) Find the general solution of

$$\sin x \frac{dy}{dx} - 2y\cos x = 3\sin^3 x.$$ **5**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]

[BLANK PAGE]