X202/13/01

NATIONAL
QUALIFICATIONS
2013

TUESDAY, 14 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. (*a*) Give an example of both a random and a non-random method of sampling from a population and state an advantage and a disadvantage of each. **4**

In chain-referral, or "snowball" sampling, used in social research, the researcher first identifies a member of the population of interest to include in the sample and to interview. The first member is then asked to refer the interviewer to a second member of the population for inclusion in the sample and so on.

(*b*) State a disadvantage of this type of sampling. **1**

A2. The random variable X has the binomial distribution with parameters 25 and 0·2.

(*a*) Write down an expression for $P(X<4)$ and evaluate it correct to four decimal places. **2**

(*b*) Use a normal approximation to estimate the required probability, both with and without a continuity correction. Comment in relation to the widely quoted rule of thumb for the reliability of this approximation. **5**

A3. The standard score for a random variable X is defined as $Z = \dfrac{X - \mu}{\sigma}$ where $\mu = E(X)$ and $\sigma^2 = V(X)$.

(*a*) Use the laws of expectation and variance to determine the mean and standard deviation of Z. **3**

In national examinations in Mathematics and Music the mean marks were 50 and 65 respectively and the standard deviations 10 and 15 respectively. A student scored 60 in Mathematics and 80 in Music.

(*b*) Calculate this student's standard scores and comment. **2**

Marks

A4. Weight gain in lambs of a breed of sheep reared on a particular diet regime is known to have mean 160 g/day and standard deviation 24 g/day. A random sample of lambs of this breed were fed the diet supplemented by fish meal and the weight gains were:

218, 201, 143, 184, 172, 193, 163, 216, 127, 163, 156 and 173 g/day.

(*a*) Stating any assumptions required, use a z-test to determine whether or not the data provide any evidence that the fish meal supplement leads to a change in mean weight gain for the breed. **5**

(*b*) Alternatively, using a student's t-distribution, a 95% confidence interval for the mean weight gain of lambs of this breed, fed with the supplemented diet, is (158·0, 193·5).

Comment on this result and also on the use of a t-interval. **2**

A5. At the Vienna General Hospital in 1843 there were 5799 births in the two maternity clinics. Maternal mortality due to puerperal fever was of major concern to one of the physicians, Dr Ignaz Semelweiss. Classification of the births yielded the following contingency table.

		Clinic in which birth took place	
		No. 1	No. 2
Maternal outcome	Death from puerperal fever	274	164
	No death from puerperal fever	2786	2575

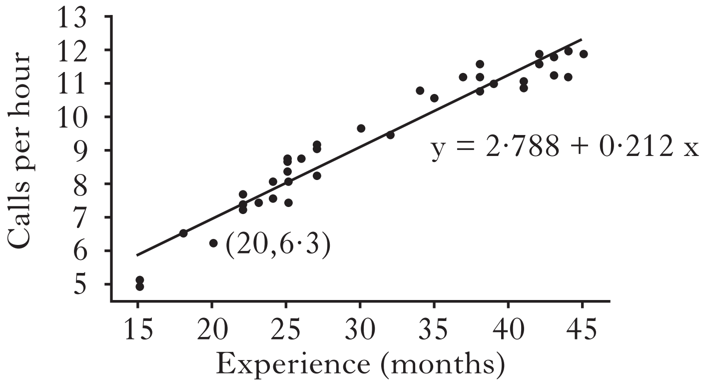
(*a*) Carry out a formal test of association between maternal outcome and clinic in which the birth took place. **4**

(*b*) Summarise your findings, including numerical information, in a concise form that could be understood by someone with no knowledge of statistics. **2**

[Turn over

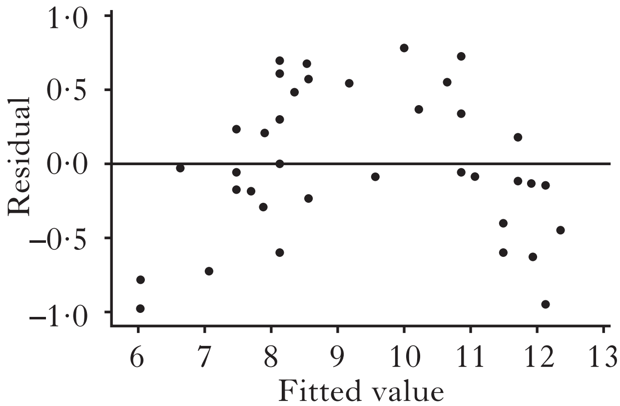
Marks

A6. At a large call centre the resource management team carried out an investigation of the relationship between experience of staff, as measured by months in post, and productivity, as measured by completed calls per hour over a month. The data for a random sample of 40 staff are displayed in the scatter plot below, annotated with the equation of the least squares regression line and the coordinates of one of the data points.



(*a*) Calculate the residual for the data point given. **2**

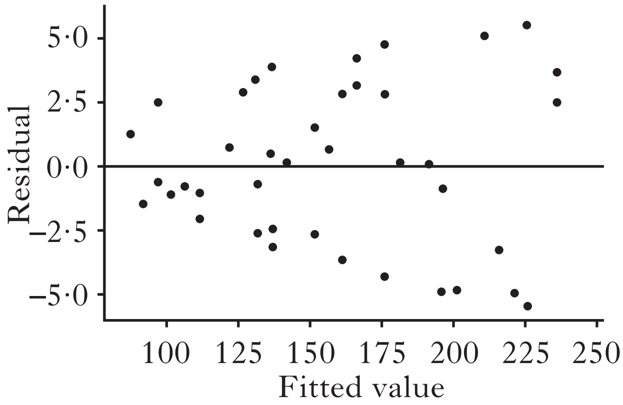
The plot of residuals against fitted values is shown below.



(*b*) State why this plot indicates a problem with the fitted model. How might the model be improved? **2**

A6. (continued)

In another investigation the team obtained the following residual plot.



(*c*) State the assumption of the standard linear model that appears to be violated in this case and what action might be taken to deal with it. **2**

A7. A computer system is subject to random attacks from both Attila and Berserker malware programs. The mean number of attacks per day from each of these is λ and the two types of attack occur independently of each other.

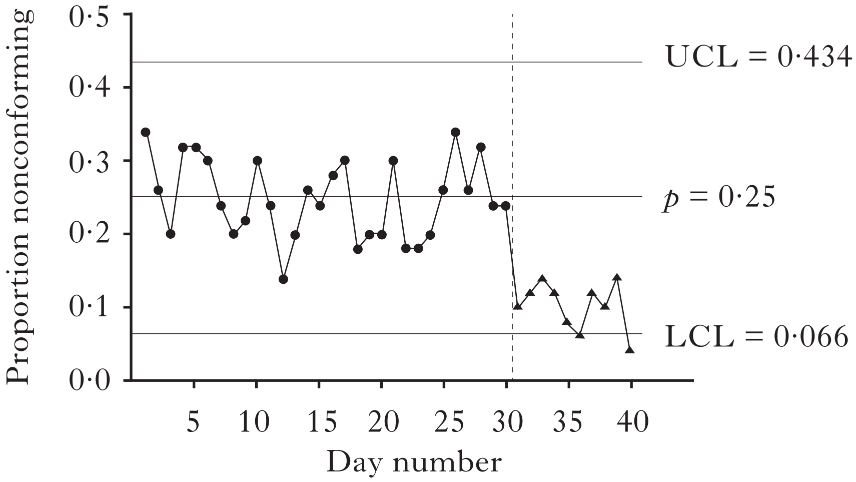
(*a*) State the distribution of the number of attacks per day from each type of malware and show that over a period of n days the expected number of days on which attacks occur is $n(1 - e^{-2\lambda})$. **3**

(*b*) Show that $\dfrac{1-e^{-\lambda}}{1-e^{-2\lambda}} = \dfrac{1}{1+e^{-\lambda}}$ and hence show that on a day when the system was subjected to attack the probability that it was due to just one of the two types of malware is $\dfrac{2}{e^{\lambda}+1}$. **5**

[Turn over

A8. (a) Using the variance of the random variable $X \sim B(n, p)$, obtain the standard deviation of the proportion of successes X/n, justifying your method. **3**

During a 30-day period of manufacturing In-Plane Switching displays for tablet computers, the proportion of nonconforming displays at final inspection was estimated to be 0·25. The proportions nonconforming in random samples of 50 taken daily over that period of 30 days are displayed in the control chart shown (circular symbols).



(b) Confirm the values of the 3-sigma control chart limits. **2**

At the end of the 30-day period, modifications were made to the manufacturing process. Data for the subsequent 10-day period are also displayed in the control chart (triangular symbols).

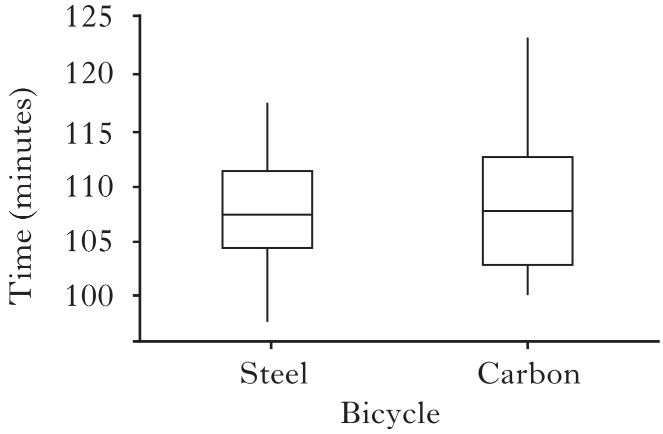
(c) Explain how the chart provides evidence of improvement. **3**

Following the modification, the proportion of nonconforming displays was estimated to be 0·10. The process manager wished to continue to monitor the process with a view to further reducing the proportion of nonconforming displays.

(d) Show that, with sample size 50, a negative value would now be obtained for the lower chart limit. Determine the minimum sample size that would yield a non-negative lower limit and state why such a lower limit is desirable. **4**

A9. On the days between mid-January 2010 and mid-July 2010 when he cycled to work, Dr Jeremy Groves randomly allocated either his steel frame bicycle or his carbon frame bicycle for his journey. The time the bicycle was moving for the 27-mile round trip was recorded using a bicycle computer. Analysis of his data was published in the Royal Statistical Society magazine *Significance*. The data is used with permission of Dr Groves who believed that the lighter carbon frame would lead to shorter times. He made 30 journeys on the steel bicycle and 26 on the carbon one and boxplots of the times are shown below. *Marks*

(*a*) Comment on the boxplots. **1**



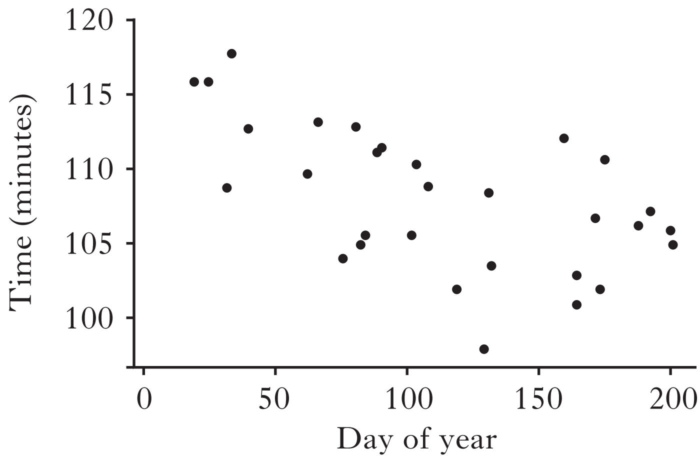
The data were ranked with the shortest time being allocated rank 1 and the rank sum for the steel bicycle was 856·5.

(*b*) Perform a formal test to evaluate the data for any evidence of a difference in median times. **5**

A non-parametric test of the null hypothesis that the variances of time for the two bicycles are equal, with a two-sided alternative, yielded a p-value of 0·198.

(*c*) Explain the relevance of this to the test performance in (*b*). **2**

A scatter plot of time versus day of the year for the steel bicycle is shown below, the sample product moment correlation coefficient being –0·589.



(*d*) Show that there is strong evidence of a non-zero population correlation coefficient and comment in relation to the test in (*b*). **4**

[*END OF SECTION A*]

[Turn over for Section B on *Pages eight and nine*

Section B (Mathematics for Applied Mathematics)

Marks

Answer all the questions

B1. Given that $y = \sin(e^{5x})$, find $\dfrac{dy}{dx}$. **2**

B2. Matrices are given as

$$A = \begin{pmatrix} 4 & x \\ 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 5 & 1 \\ 0 & 1 \end{pmatrix} \quad C = \begin{pmatrix} y & 3 \\ -1 & 2 \end{pmatrix}.$$

(*a*) Write $A^2 - 3B$ as a single matrix. **2**

(*b*) (i) Given that C is non-singular, find C^{-1}, the inverse of C. **2**

(ii) For what value of y would matrix C be singular? **1**

B3. Use integration by parts to obtain

$$\int \frac{\ln x}{x^3}\,dx$$

where $x > 0$. **4**

B4. (*a*) State $\displaystyle\sum_{r=1}^{n} r$ and $\displaystyle\sum_{r=1}^{n} r^3$ in terms of n.

Hence show that

$$\sum_{r=1}^{n} \left(r^3 - 3r\right) = \frac{n(n+1)(n-2)(n+3)}{4}.$$ **4**

(*b*) Use the above result to evaluate $\displaystyle\sum_{r=5}^{15} \left(r^3 - 3r\right)$. **2**

Marks

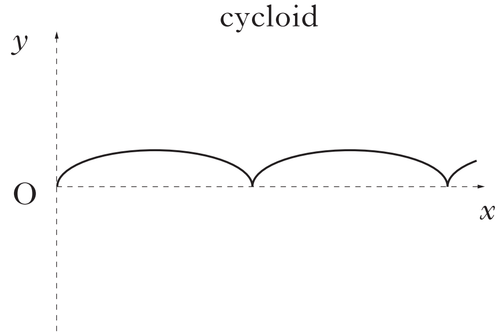
B5. Find the general solution of the differential equation

$$\frac{1}{x}\frac{dy}{dx} + 2y = 6, \; x \neq 0.$$ **6**

B6. The cycloid curve below is defined by the parametric equations

$$x = t - \sin t, \; y = 1 - \cos t.$$

cycloid



(*a*) Find $\dfrac{dy}{dx}$ in terms of t. **2**

(*b*) Show that the value of $\dfrac{d^2y}{dx^2}$ is always negative, in the case where $0 < t < 2\pi$. **5**

(*c*) A particle follows the path of the cycloid where t is the time elapsed since the particle's motion commenced.
Calculate the speed of the particle when $t = \dfrac{\pi}{3}$. **2**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]

[BLANK PAGE]

[BLANK PAGE]