X202/701

NATIONAL QUALIFICATIONS 2009

MONDAY, 25 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. When the cutting process is in control, timber battens produced at a sawmill have width, W (mm), where $W \sim N(50, 0{\cdot}16)$. Samples of 4 battens are taken every 15 minutes during production. A control chart for the sample means is maintained to monitor the process.

(*a*) Calculate upper and lower three-sigma chart limits and verify that, when the process is in control, the probability that a point lies outwith these limits is 0·0026. **3**

(*b*) Given that the mean changes to 50·5 mm, calculate the probability that a point now lies outwith the limits and estimate the approximate duration of production until the chart signals an out of control situation. **4**

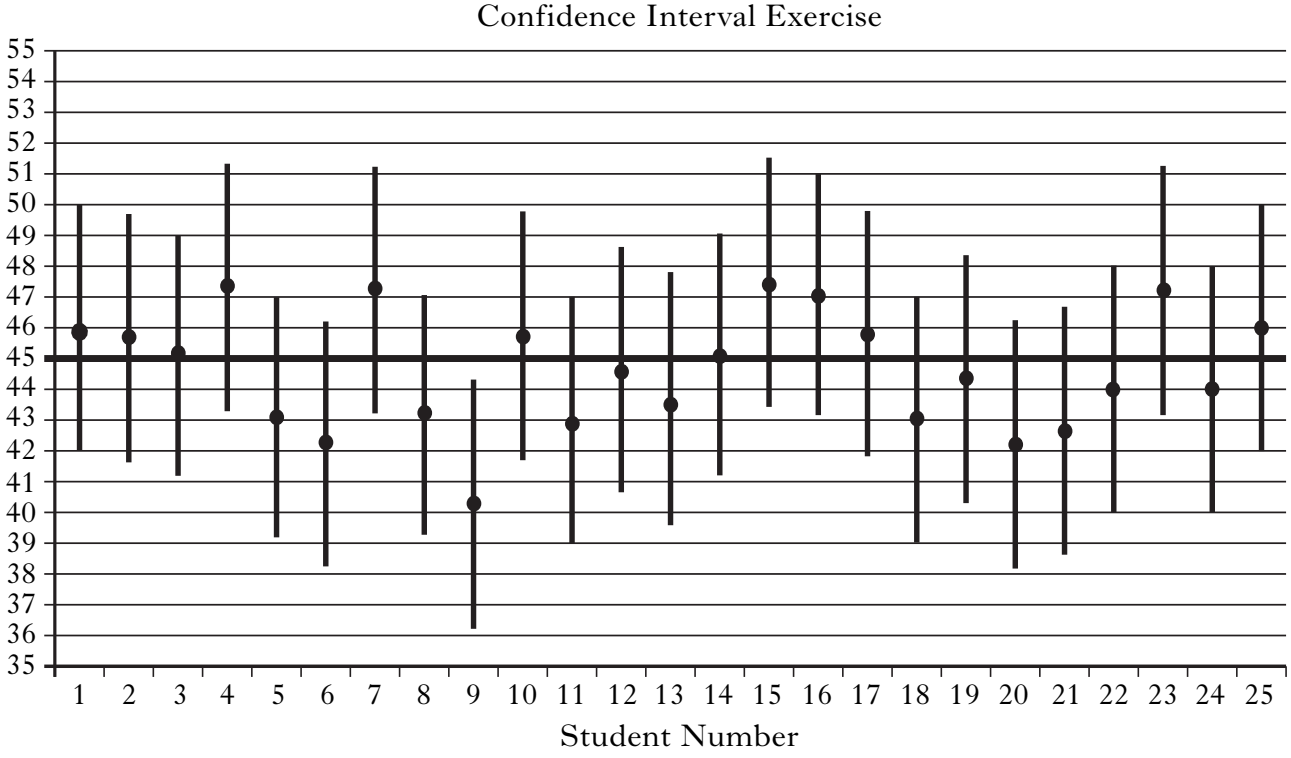
A2. For each of the 25 students in her class, a statistics tutor generated a random sample of 9 observations from a normal distribution with $\sigma = 6$.

The sample given to the first student was:

43·0 47·3 39·4 35·3 54·3 38·2 47·8 57·4 50·7.

(*a*) Obtain a 95% confidence interval for the population mean μ. **3**

(*b*) The tutor displayed the confidence intervals obtained by all the students as vertical line segments in the plot below and revealed to the students that μ was in fact 45.



Determine the proportion of students whose samples yielded confidence intervals that captured the population mean and comment. **2**

A3. In 1973 the statistician Frank Anscombe published four artificial sets of bivariate data which are displayed below.

Data Set 1	
x	y
10	8·04
8	6·95
13	7·58
9	8·81
11	8·33
14	9·96
6	7·24
4	4·26
12	10·84
7	4·82
5	5·68

Data Set 2	
x	y
10	9·14
8	8·14
13	8·74
9	8·77
11	9·26
14	8·10
6	6·13
4	3·10
12	9·13
7	7·26
5	4·74

Data Set 3	
x	y
10	7·46
8	6·77
13	12·74
9	7·11
11	7·81
14	8·84
6	6·08
4	5·39
12	8·15
7	6·42
5	5·73

Data Set 4	
x	y
8	6·58
8	5·76
8	7·71
8	8·84
8	8·47
8	7·04
8	5·25
19	12·50
8	5·56
8	7·91
8	6·89

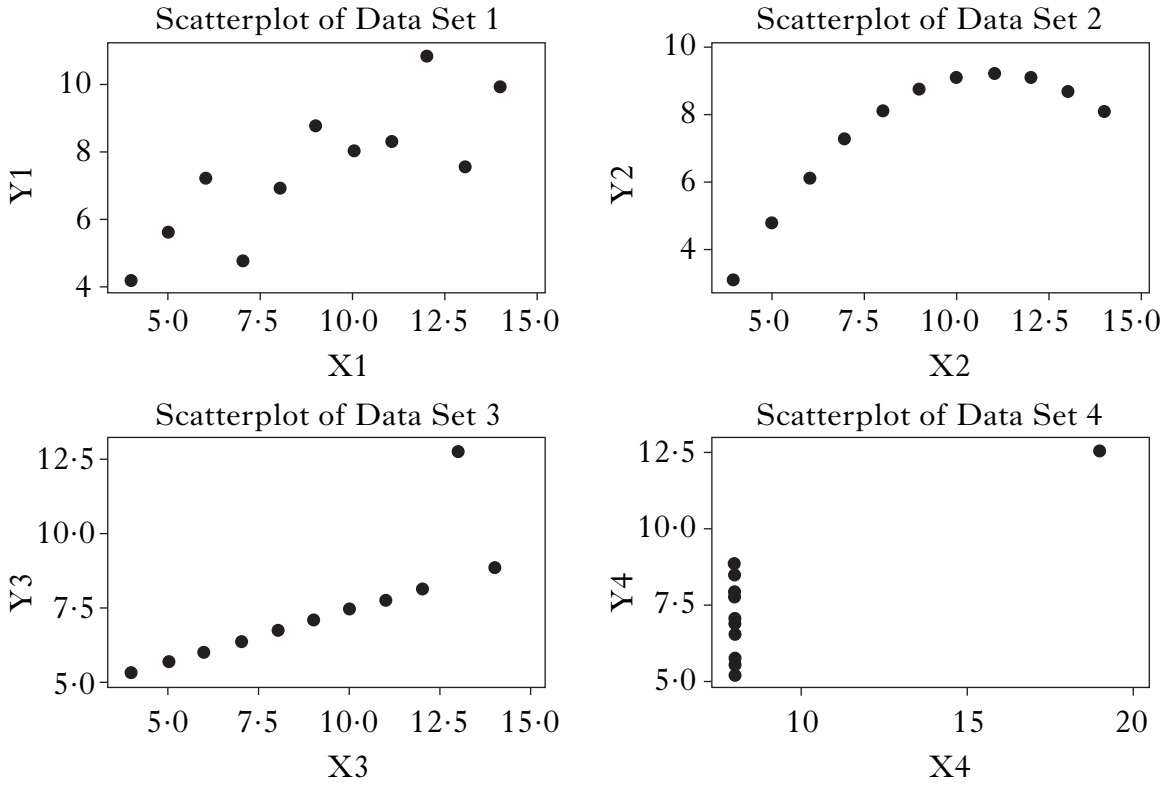
(Source: Graphs in Statistical Analysis. *American Statistician*, **27**, pp. 17–21.)

The data sets are such that in all four cases:

$$\sum(x_i - \bar{x})^2 = 110 \cdot 00; \qquad \sum(y_i - \bar{y})^2 = 41 \cdot 27; \qquad \sum(x_i - \bar{x})(y_i - \bar{y}) = 55 \cdot 01$$

(*a*) Calculate the product moment correlation coefficient and test the null hypothesis that the corresponding population value is zero. **5**

(*b*) Consider the scatter plots below. For which data set is it appropriate to calculate the product moment correlation coefficient? **1**



Marks

A4. A random sample of 600 males aged 50 to 60 years was taken from a community and cross-classified by the level of their serum cholesterol (below or above 7·0 mmol/L) and the presence or absence of heart disease. The data are given in the table below.

	Present	*Absent*	
< 7·0 mmol/L	28	425	453
≥ 7·00	21	126	147
	49	551	600

A researcher used a statistical software package to carry out a chi-squared test of association between heart disease and serum cholesterol and obtained a p-value of 0·002.

(*a*) State the null and alternative hypotheses for the test and the conclusion you would make. **3**

(*b*) Calculate the value of the chi-squared test statistic. Use the statistical tables provided to obtain the narrowest possible interval in which the p-value must lie. **4**

A5. A pet food manufacturer claims that a new type of dog biscuit relieves the symptoms of arthritis in 80% of elderly dogs. A reporter for a consumer magazine decides to arrange a trial involving 100 elderly dogs with arthritis and to accept the manufacturer's claim if 75 or more dogs are deemed to have benefited.

Use a normal approximation to estimate the probability that the claim will be:

(*a*) rejected when the relief probability is actually 0·8; **4**

(*b*) accepted when the relief probability is actually 0·7. **2**

Marks

A6. The random variable X assumes only the values 0 and 1, with probability p that it takes the value 1. Legal documents for house purchase are classified as either sound (S), in which case $X = 0$, or deficient (D) when $X = 1$. The results for a random sample of 10 documents are shown below.

Document	1	2	3	4	5	6	7	8	9	10
Status	S	D	S	S	D	S	S	S	S	D
x	0	1	0	0	1	0	0	0	0	1

(*a*) Explain, using the given sample, how the proportion of deficient documents in a sample may be regarded as the mean of the sample of X values. **2**

(*b*) The mean and variance of X are p and pq respectively (where $q = 1 - p$). For random samples of n documents, the sample proportion is $\frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n$.
Derive expressions for the mean and variance of the sample proportion. **4**

(*c*) State the additional information provided by the central limit theorem on the proportion of deficient documents in random samples of size n, where n is large. **1**

A7. Ten seeds of a rare plant were sown in two types of compost, A and B, with four randomly chosen seeds planted in compost A and the remaining six in compost B. The heights (mm) of the seedlings, eight weeks after planting, were as follows.

Compost A 149 155 147 153
Compost B 169 163 165 159 151 157

(*a*) Rank the 10 heights in ascending order and show that the sum of the ranks for A is 12. **2**

(*b*) Show that the number of subsets of {1, 2, 3, 4, 5, 6, 7, 8, 9, 10} that could potentially have provided ranks for A is 210. **1**

(*c*) List all of those subsets which provide a sum of ranks for A less than or equal to 12.

Under the null hypothesis that A and B yield eight-week seedlings with equal median heights, write down the probability that the rank sum for A is less than or equal to 12. **3**

(*d*) State, with justification, the decision you would reach on the null hypothesis in relation to the alternative hypothesis that the median for A is less than that for B. Interpret your decision. **3**

[Turn over

Marks

A8. A manufacturer of laptop computers uses batteries from three different suppliers. The batteries are sent to a central store with no record of the suppliers. Supplier A provides 60% of the batteries, of which 3% fail during the warranty period. Supplier B provides 30% of the batteries with a 1% failure rate and the corresponding figures for supplier C are 10% and 2%.

(*a*) Given that a battery fails during the warranty period, calculate the probability that it came from supplier A. **4**

(*b*) Of the first five batteries to fail under warranty, find the probability that exactly 3 of them came from supplier A. **3**

(*c*) During a particular financial year, the manufacturer incurred a total replacement cost of £200 000 for batteries failing during the warranty period. Calculate an appropriate breakdown of this cost among the three suppliers. **3**

A9. An experiment was conducted in order to test the claim that durability was identical for two types of tyre. A tyre of type A and a tyre of type B were fitted on the front wheels of each one of 10 cars, the choice of left or right wheel for each type being made at random. The cars were each driven by a different driver on public roads for a month. Wear measurements (mm) for each tyre are recorded in the table below.

Car	1	2	3	4	5	6	7	8	9	10
Tyre A	1·94	2·13	2·11	1·92	1·98	2·05	2·00	2·03	2·12	1·94
Tyre B	1·99	2·15	2·11	1·98	2·00	2·04	2·03	2·11	2·15	1·98

(*a*) Explain why the designers of the experiment allocated a tyre of each type to each driver rather than fitting five cars with front tyres of type A and the other five with front tyres of type B. **1**

(*b*) Given that the differences in wear between the 10 pairs of tyres have mean 0·032 and standard deviation 0·026, analyse the data, at the 1% level of significance, using a Student's *t*-test, stating one assumption required. **6**

(*c*) Carry out an alternative analysis of the data and comment. **4**

[*END OF SECTION A*]

Section B (Mathematics for Applied Mathematics)

Marks

Answer all the questions.

B1. Obtain the binomial expansion of $\left(b-\frac{2}{b}\right)^5$ and simplify the expression. **4**

B2. Obtain $\int_0^{\pi/3} \cos^5 x \sin x \, dx$ by using the substitution $u = \cos x$ or otherwise. **4**

B3. A particle moves along a curve in the x-y plane. The curve is defined by the parametric equations

$$x = t^2 + 1, \qquad y = 1 - 3t^3,$$

where t is the time elapsed since the start.

Find $\frac{dy}{dx}$ in terms of t and hence obtain an equation of the tangent to the curve when $t = 2$. **5**

B4. Determine k such that the matrix $\begin{pmatrix} 1 & 1 & 0 \\ 0 & k-2 & -1 \\ 1 & 2 & k \end{pmatrix}$ does not have an inverse. **4**

B5. An industrial scientist finds that the differential equation

$$t\frac{dx}{dt} - 2x = 3t^2$$

models a production process.

Find the general solution of the differential equation. **5**

Hence find the particular solution given $x = 1$ when $t = 1$. **1**

[Turn over for Question B6 on *Page eight*

Marks

B6. Given $f(x) = x \tan 2x$ for $-\frac{\pi}{4} < x < \frac{\pi}{4}$, obtain an expression for $f'(x)$ and show that $f''(x) = 4 \sec^2 2x \, (1 + 2x \tan 2x)$. **2,3**

Hence find the exact value of $\int_0^{\pi/6} \frac{1 + 2x \tan 2x}{\cos^2 2x} \, dx$. **4**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]