# X202/13/01

NATIONAL QUALIFICATIONS 2015

THURSDAY, 30 APRIL
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*) In order to study television-viewing patterns across all households with a television in a country, a carefully chosen panel of private homes is to be created. These homes are to be selected to be representative of all households with a television across the country. State the type of sampling involved. **1**

(*b*) Explain carefully how you would take a systematic sample of 8 households, from a street that included 120 households living in houses numbered 1 to 120, as part of the investigation. State a problem that would arise were you to proceed in precisely the same way to take a sample of 8 households from a street that included 125 households living in houses numbered 1 to 125. Suggest an alternative method that could be used to take a sample. **4**

**A2.** Fasteners used by a clothing manufacturer are delivered in batches from a supplier with factories F1 and F2. Factory F1 produces proportion $p_1$ defective and factory F2 proportion $p_2$ defective and half the batches are from each factory. During acceptance sampling of batches a random sample of $n$ fasteners from a batch, for which the factory that produced it is not known, is inspected and found to include $x$ defective fasteners.

(*a*) Show that the probability that the batch was produced at F1 is given by:

$$\frac{p_1^x (1-p_1)^{n-x}}{p_1^x (1-p_1)^{n-x} + p_2^x (1-p_2)^{n-x}}$$ **4**

(*b*) Comment on the value of this probability when $p_1 = p_2$. **1**

*Marks*

**A3.** Study of a queuing system revealed that the service time (minutes) for customers could be modelled by a highly skewed distribution with mean 5 and standard deviation 5.

(*a*) Find the approximate probability that the mean time to serve a random sample of 16 customers lies between 4 and 6 minutes. **3**

(*b*) Repeat the calculation for a sample of 100 customers. **2**

Statistical distribution theory may be used to calculate the exact answers to be 0·579 and 0·955 respectively.

(*c*) State the feature of the central limit theorem that comparison of the approximations with the true answers demonstrates. **1**

**[Turn over**

*Marks*

**A4.** The table below shows enzyme activity in a leaf sample from each one of a sample of 20 plants both before and after the application of a fungicide. The study was conducted by two biologists with the aim of determining whether or not there was evidence that the fungicide affected enzyme activity. The difference for each pair of measurements is also displayed.

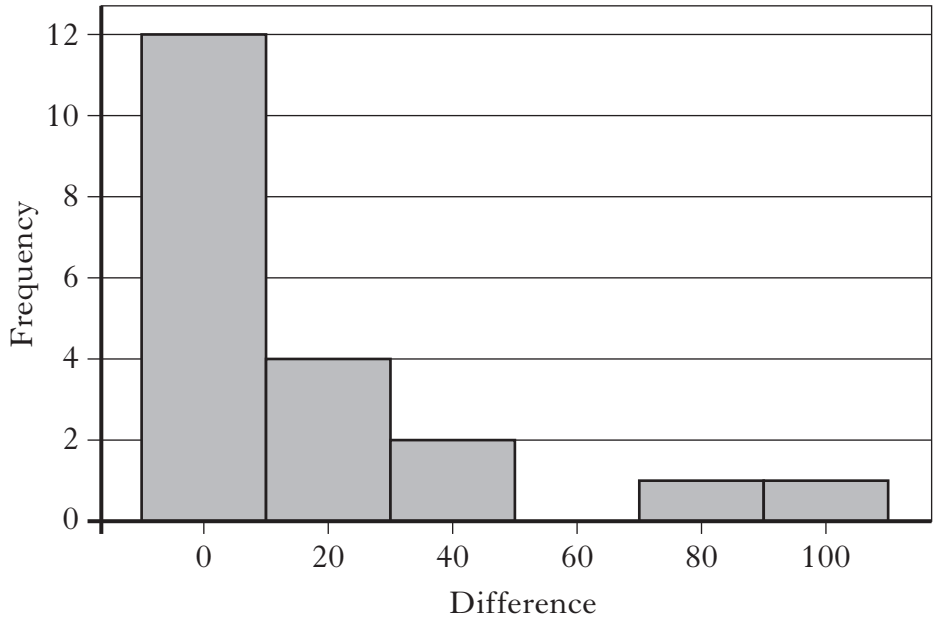
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pre | 22·0 | 47·3 | 62·7 | 234·0 | 249·1 | 51·7 | 165·8 | 43·9 | 16·1 | 45·8 |
| Post | 17·4 | 118·0 | 59·8 | 250·9 | 271·4 | 49·6 | 205·2 | 35·5 | 6·4 | 36·1 |
| Difference | −4·6 | 70·7 | −2·9 | 16·9 | 22·3 | −2·1 | 39·4 | −8·4 | −9·7 | −9·7 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Pre | 16·7 | 20·1 | 20·3 | 22·1 | 18·6 | 50·2 | 15·9 | 25·7 | 21·0 | 241·1 |
| Post | 53·4 | 112·6 | 11·4 | 23·9 | 13·3 | 59·5 | 28·1 | 32·9 | 26·4 | 268·5 |
| Difference | 36·7 | 92·5 | −8·9 | 1·8 | −5·3 | 9·3 | 12·2 | 7·2 | 5·4 | 27·4 |

The first biologist decided to perform a sign test.

(*a*) Obtain the *p*-value for a test of the null hypothesis that the median difference is zero with a two-sided alternative hypothesis and state what may be concluded. **4**

The second biologist performed a *t*-test. A histogram of the differences is shown below. A *t*-test of the null hypothesis that the mean difference is zero, with a two-sided alternative, yielded a *p*-value of 0·030.

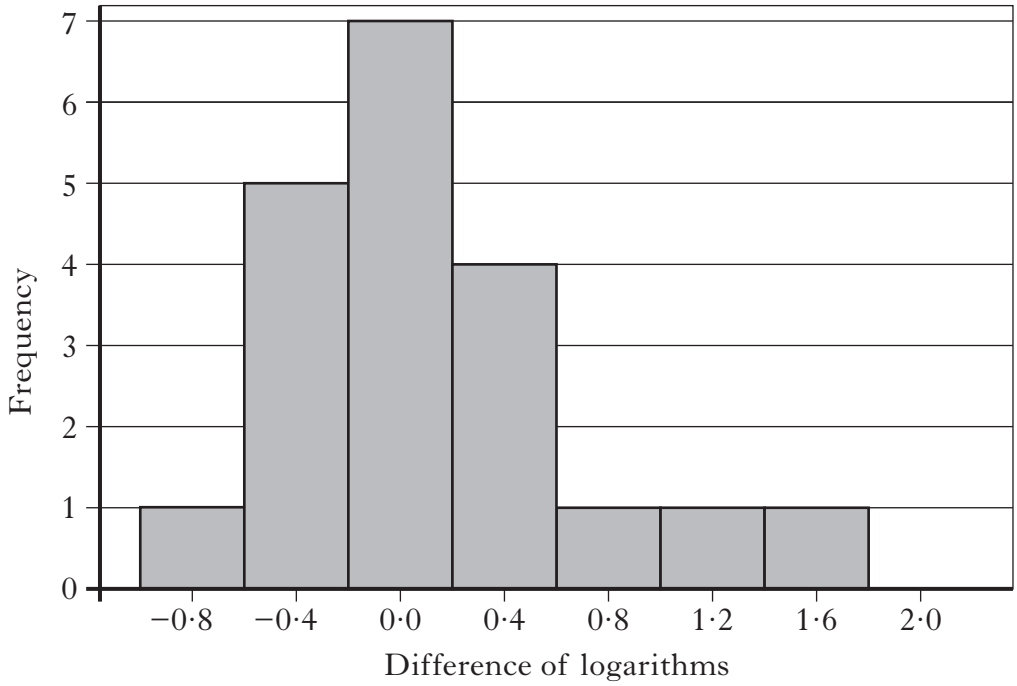



(*b*) State the conclusion that would be reached purely on the basis of the *p*-value and give a reason why that conclusion is unsafe. **2**

Concerned at the apparent contradictory analyses, the biologists consulted a statistician who suggested conducting the *t*-test after taking logarithms of the activities.

**A4. (continued)**

A histogram of the differences between the logarithms is shown below. A *t*-test of the null hypothesis that the mean difference between the logarithms is zero, with a two-sided alternative, yielded a *p*-value of 0·279.




(*c*) Explain how the apparent contradiction may now be resolved. **2**

**A5.** In teaching confidence intervals to her Advanced Higher Statistics class of 20 students, a teacher gave each student a different, independent, random sample of size 16 from a known, normal distribution. The students were given their data together with the information that the sample was from a normal distribution with standard deviation 2 and each was asked to obtain a 95% confidence interval for the population mean.

(*a*) Show that the interval is given approximately by $\bar{x} \pm 1$. **2**

(*b*) Calculate the probability that all of the students would obtain a confidence interval that captured the population mean. **2**

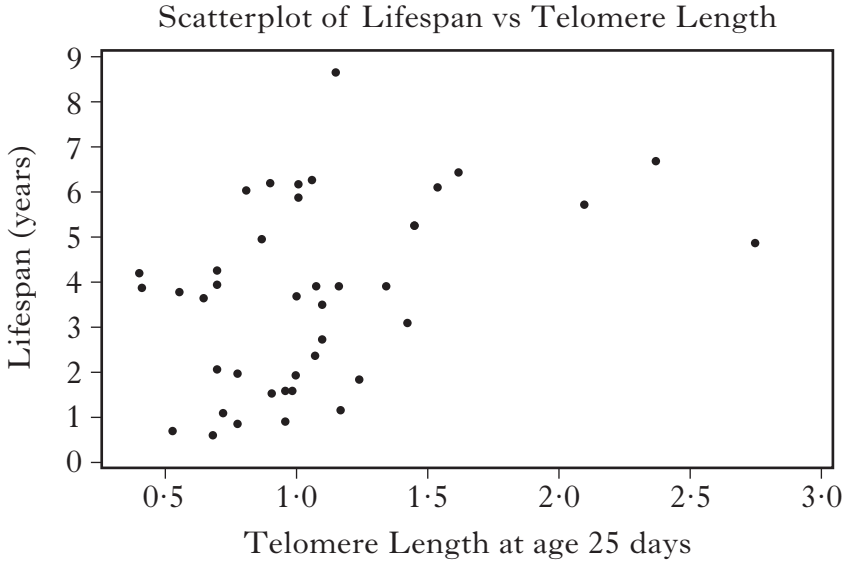
In order to enhance the lesson the teacher also asked the students to compute a 50% confidence interval.

(*c*) Obtain the formula required in the form $\bar{x} \pm a$ and suggest a reason why this might illustrate the use of confidence intervals more clearly. **3**

**[Turn over**

*Marks*

**A6.** Researchers at the University of Glasgow obtained data on telomere length at age 25 days and lifespan (years) of zebra finches. Telomeres cap the ends of eukaryotic chromosomes. A scatter plot of data for a sample of 40 birds is shown below. The data was provided by Professor Pat Monaghan.




The product moment correlation is $r = 0{\cdot}395$.

Test, at the 5% and 1% levels of significance, the null hypothesis that the population product moment correlation coefficient, $\rho$, is zero against the alternative that it is non-zero. **6**

*Marks*

**A7.** A researcher was interested in whether the occurrences of major hurricanes in a particular area could be modelled by a Poisson distribution. The numbers of major hurricanes per annum affecting an area during the 25-year period 1970–1994 had mean 1·52. The following table was obtained during analysis of the data.

| Major hurricanes per annum | Observed frequency | Expected Poisson frequency | Contribution to chi-squared |
|---|---|---|---|
| 0 | 3 | | |
| 1 | 10 | 8·311 | 0·343 |
| 2 | 8 | 6·316 | 0·449 |
| 3 or more | 4 | | |

(*a*) Calculate the missing entries in the table and complete a chi-squared goodness-of-fit test to demonstrate that the Poisson distribution provides an adequate model. State what may be concluded with regard to the occurrence of major hurricanes in the area. **6**

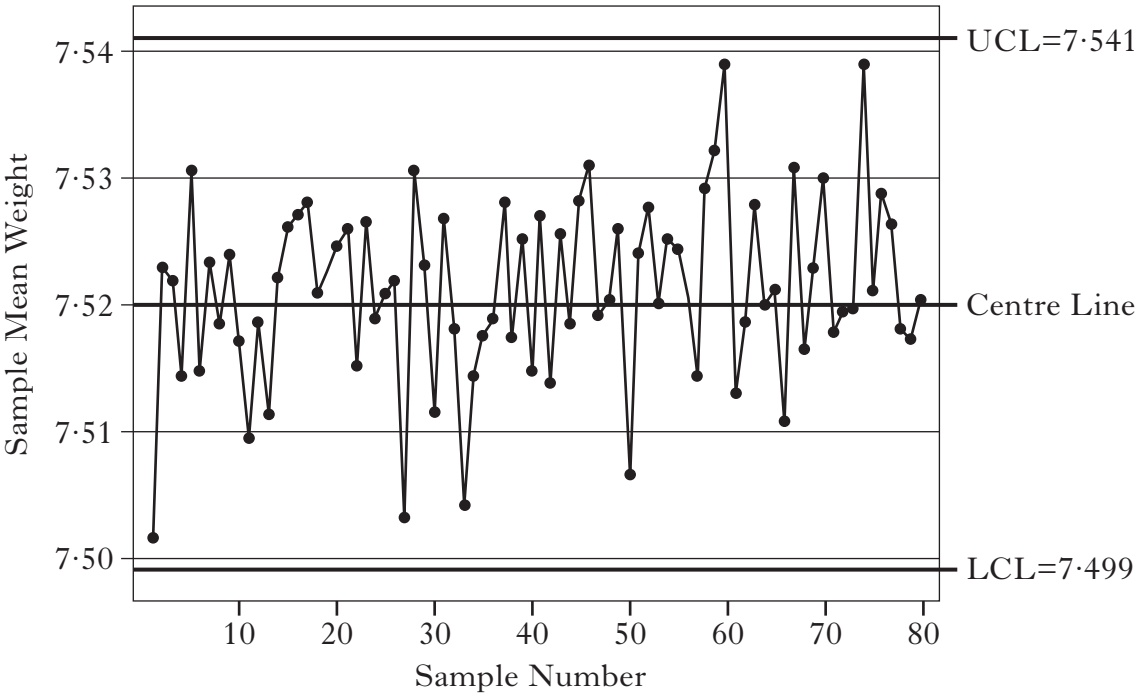
Over the next five years 17 major hurricanes occurred.

(*b*) On the basis of the model fitted to the data in part (*a*), state a model for the number of occurrences of major hurricanes per 5-year period. Find the approximate probability of 17 or more major hurricanes occurring in a 5-year period and comment. **3**

**[Turn over**

*Marks*

**A8.** The manufacture of metal tokens for use in a car parking system is monitored using control charts for the mean weight (g) of samples of 25 tokens taken at regular intervals. A chart, with 3-sigma limits, for a period of production is shown below.




(*a*) State the features of the plot that indicate that the manufacturing process is behaving in a stable, predictable manner. **1**

For the machines that receive tokens from motorists about to park their cars to operate successfully, it is desirable that the weights of individual tokens lie in the specification range 7·400 g to 7·600 g.

(*b*) On the assumption that weight is normally distributed, use information from the chart to determine the proportion of tokens that would meet specifications with current process performance. **5**

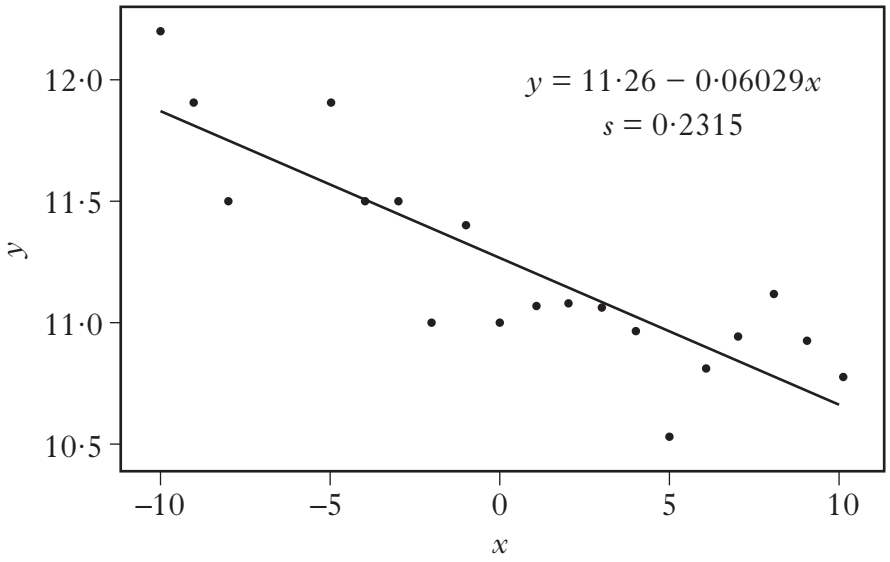
(*c*) State an adjustment to the process mean that would yield the greatest proportion of tokens meeting specifications and estimate this proportion. **3**

(*d*) Determine what would have to be done to increase the proportion still further. **1**

*Marks*

**A9.** The table below gives the winning times for the Olympic 100 m women's sprint final, $y$ seconds, from 1928 until 2008, together with a coded variable $x$. Also shown is a scatter plot created by a computer software package, together with the least squares regression line of $y$ on $x$ and associated information in the usual notation.

| Year | $x$ | Time $y$ |
|---|---|---|
| 1928 | –10 | 12·20 |
| 1932 | –9 | 11·90 |
| 1936 | –8 | 11·50 |
| 1948 | –5 | 11·90 |
| 1952 | –4 | 11·50 |
| 1956 | –3 | 11·50 |
| 1960 | –2 | 11·00 |
| 1964 | –1 | 11·40 |
| 1968 | 0 | 11·00 |
| 1972 | 1 | 11·07 |
| 1976 | 2 | 11·08 |
| 1980 | 3 | 11·06 |
| 1984 | 4 | 10·97 |
| 1988 | 5 | 10·54 |
| 1992 | 6 | 10·82 |
| 1996 | 7 | 10·94 |
| 2000 | 8 | 11·12 |
| 2004 | 9 | 10·93 |
| 2008 | 10 | 10·78 |



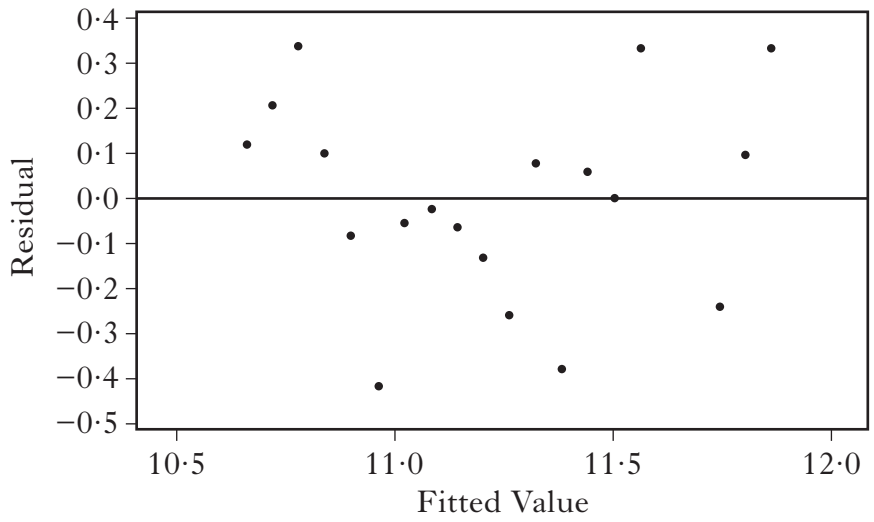

(*a*) Calculate $S_{xx}$ and assess the evidence that the slope parameter differs significantly from zero at the 0·1% level of significance. **4**

(*b*) Use the model to estimate the winning time in the 2012 Olympics and explain why the model would be unrealistic for making predictions into the distant future. **2**

**[Turn over**

**A9. (continued)**

A plot of residuals versus fitted values is shown below.




(*c*) Calculate the residual for the data point corresponding to 2008 and state the feature of the residual plot that indicates that a non-linear model might be appropriate. **3**

The software package yielded the non-linear model

$$y = 10{\cdot}604 + 0{\cdot}5554e^{-0{\cdot}0983x}.$$

(*d*) Explain why this model is more realistic than the original, linear model. State a prediction that may be made for the winning time, in the women's 100 m sprint final, in the distant future. **3**

[*END OF SECTION A*]

## Section B (Mathematics for Applied Mathematics)

*Marks*

**Answer all the questions**

**B1.** Given that $y = e^{5x} \tan 2x$, find $\frac{dy}{dx}$. **3**

**B2.** (*a*) Given matrix $A = \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}$, find $A^2$ and show that the inverse of $A^2$ exists. **2**

(*b*) Hence, or otherwise, find matrix $B$ such that $A^2B = \begin{pmatrix} 4 & 6 \\ 2 & -2 \end{pmatrix}$. **3**

**B3.** A curve is defined by

$$y = \frac{\sin x}{2 - \cos x} \text{ for } 0 \leq x \leq \pi.$$

Find the exact values of the coordinates of the stationary point of this curve. **5**

**B4.** Express $\log_a 2 + \log_a 4 + \log_a 8$ in the form $p\log_a 2$, where $p$ is a constant. **1**

Hence evaluate $\sum_{r=1}^{100} \log_a 2^r$, giving your answer in the form $q\log_a 2$, where $q$ is a constant. **3**

**B5.** Find the general solution, in the form $y = f(x)$, of the differential equation

$$\frac{1}{\cos x}\frac{dy}{dx} + y \tan x = \tan x, \ 0 < x < \frac{\pi}{2}$$ **6**

**B6.** (*a*) Express $\frac{1}{1 - y^2}$ in partial fractions. **3**

(*b*) Use the substitution $u = \sqrt{1 - x}$ to obtain $\int \frac{dx}{x\sqrt{1 - x}}$, $0 < x < 1$. **6**

[*END OF SECTION B*]

[*END OF QUESTION PAPER*]

ACKNOWLEDGEMENT

Statistics – Question A6 – Data on telomere length and lifespan of zebra finches is reproduced by kind permission of Professor Pat Monaghan, University of Glasgow.