OCR
Oxford Cambridge and RSA

Friday 6 June 2025 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

You must have:
- the Printed Answer Booklet
- the Formulae Booklet for A Level Further Mathematics A
- a scientific or graphical calculator

QP

INSTRUCTIONS
- Use black ink. You can use an HB pencil, but only for graphs and diagrams.
- Write your answer to each question in the space provided in the **Printed Answer Booklet**. If you need extra space use the lined page at the end of the Printed Answer Booklet. The question numbers must be clearly shown.
- Fill in the boxes on the front of the Printed Answer Booklet.
- Answer **all** the questions.
- Where appropriate, your answer should be supported with working. Marks might be given for using a correct method, even if your answer is wrong.
- Give non-exact numerical answers correct to **3** significant figures unless a different degree of accuracy is specified in the question.
- The acceleration due to gravity is denoted by $g\,\mathrm{m\,s^{-2}}$. When a numerical value is needed use $g = 9.8$ unless a different value is specified in the question.
- Do **not** send this Question Paper for marking. Keep it in the centre or recycle it.

INFORMATION
- The total mark for this paper is **75**.
- The marks for each question are shown in brackets **[ ]**.
- This document has **8** pages.

ADVICE
- Read each question carefully before you start your answer.

1 A set of 16 observations of the bivariate data (X, Y) are summarised as follows.

$n = 16 \quad \Sigma x = 136 \quad \Sigma y = 352 \quad \Sigma x^2 = 1496 \quad \Sigma y^2 = 9104 \quad \Sigma xy = 3642$

Determine an estimate of the value of y corresponding to $x = 8.5$. **[4]**

2 The number of trees of a particular species found in 1km^2 of a forest can be modelled by the distribution $\text{Po}(\lambda)$.

The forest consists of two regions, A and B.

In the region A, $\lambda = 4$. The number of trees of this species in a randomly chosen area of 3km^2 in region A is denoted by X.

(a) Find $\text{P}(16 < X < 20)$. **[3]**

In region B, $\lambda = 8$. The number of trees of this species in a randomly chosen area of 3km^2 in region B is denoted by Y.

The random variable Z is defined by $Z = X + Y$. It may be assumed that X and Y are independent.

(b) Write down a formula for $\text{P}(Z = z)$, in terms of z. **[2]**

(c) The average numbers of trees of this species in one square kilometre are not the same in regions A and B.

Explain why the associated modelling assumption for the distribution of Z is nevertheless valid. **[1]**

(d) Show that $Y - X$ does **not** have a Poisson distribution. **[1]**

3 The length, in cm, of sand lizards when fully grown is a random variable which is known to have the distribution $N(13.2, 3.4^2)$.

Some scientists discover a colony of what they believe to be sand lizards in a remote location. The scientists measure the lengths of a random sample of 50 fully grown lizards from this remote location. The mean of this random sample is 14.02 cm. You may assume that the variance of the length of lizards on the remote island is 3.4^2 cm^2.

(a) Test, at the 5% significance level, whether the fully grown lizards in this location have a mean length which is different from 13.2 cm. **[7]**

(b) Now suppose it is **not** known that the lengths are normally distributed.

(i) Explain why the Central Limit Theorem can be used in this context. **[1]**

(ii) State where in your test the Central Limit Theorem would be used. **[1]**

4 In this question you must show the parameters of any distributions you use.

The continuous random variable V has the distribution $N(42, 3^2)$.

The continuous random variable W has the distribution $N(48, 4^2)$.

The random variable X is the sum of 5 independent observations of V.

The random variable Y is the sum of 7 independent observations of W.

(a) Determine $P(X + Y > 560)$. **[3]**

(b) Determine the probability that X is less than 60% of Y. **[4]**

5 For many years a coach company has used a make of tyre that has an average driving lifespan of 1.52×10^5 km. The coach company decides to fit a new make of tyre to 12 of its coaches, and records the lifespan of one of the tyres, randomly chosen, for each of these 12 coaches. The distances, in multiples of 10^5 km, are given below.

1.35	1.46	1.48	1.53	1.54	1.57
1.62	1.74	1.76	1.82	1.88	1.91

The coach company wishes to test whether the new make of tyres has improved the driving lifespan.

(a) Explain why a hypothesis test based on a normal distribution may not be valid. **[1]**

(b) Carry out a suitable Wilcoxon test at the 5% significance level. **[7]**

(c) Explain why a Wilcoxon test, if it is valid, is preferable to a sign test in this context. **[1]**

6 The editor of a mathematical journal investigated whether there was association between the types of articles published in the journal and the academic backgrounds of the authors.

The results for a random sample of 64 articles are shown in the table.

	Pure maths	Applied maths	Other
Schools and colleges	6	4	5
Universities	30	12	7

The editor carries out a test at the 10% significance level.

(a) Explain why it is not possible to combine rows of the table before carrying out the test. **[1]**

(b) Show that it is necessary to combine columns before carrying out the test. **[2]**

(c) Carry out the test. **[6]**

7 **(a)** A large number of competitors take part in two races. A random sample of 6 competitors who take part in both races is obtained. The value of Spearman's rank correlation coefficient for the finishing positions of these 6 competitors is 0.771, correct to 3 significant figures.

Test, at the 5% significance level, whether there is agreement between the finishing positions in the two races. **[4]**

(b) Show that, when only 3 competitors are selected at random, the value of Spearman's rank correlation coefficient can take only 4 different values. You should state what these values are. **[5]**

8 The proportion X of charge in a phone battery is measured on a scale from 0 (uncharged) to 1 (fully charged).

For a certain model of phone, X can be modelled as a continuous random variable with the following probability density function.

$$\mathrm{f}(x) = \begin{cases} (k+1)x^k & 0 \leqslant x \leqslant 1, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive constant.

(a) Show that $\mathrm{f}(x)$ is a valid probability density function for all positive values of k. **[3]**

(b) Suppose that $k = 3$.

Determine $\mathrm{Var}(X)$. **[5]**

(c) Suppose instead it is known that X is less than 0.8203 for 50% of this model of phone.

Determine the 70th percentile of X. **[5]**

9 A set of 50 independent observations of a discrete random variable X, with an unknown distribution, is found.

(a) The mean and variance of the 50 observations are 4.04 and 3.5184 respectively.

Determine whether these results are consistent with the following distributions.

- A geometric distribution
- A Poisson distribution **[4]**

(b) The values of the observations are given in the table.

x	0	1	2	3	4	5	6	7	8	> 8
Observed frequency	0	4	7	10	11	7	5	3	3	0

Without carrying out a goodness of fit test, determine whether these frequencies are consistent with the following distributions.

- A geometric distribution
- A Poisson distribution **[4]**

END OF QUESTION PAPER

BLANK PAGE