OCR
Oxford Cambridge and RSA

Thursday 13 June 2019 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

You must have:
- Printed Answer Booklet
- Formulae A Level Further Mathematics A

You may use:
- a scientific or graphical calculator

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.** If additional space is required, you should use the lined page(s) at the end of the Printed Answer Booklet. The question number(s) must be clearly shown.
- You are permitted to use a scientific or graphical calculator in this paper.
- 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 $g\,\mathrm{m\,s^{-2}}$. Unless otherwise instructed, when a numerical value is needed, use $g = 9.8$.

INFORMATION
- The total mark for this paper is **75**.
- The marks for each question are shown in brackets **[ ]**.
- **You are reminded of the need for clear presentation in your answers.**
- The Printed Answer Booklet consists of **16** pages. The Question Paper consists of **4** pages.

Answer **all** the questions.

1 A set of bivariate data (X, Y) is summarised as follows.

$n = 25, \Sigma x = 9.975, \Sigma y = 11.175, \Sigma x^2 = 5.725, \Sigma y^2 = 46.200, \Sigma xy = 11.575$

(a) Calculate the value of Pearson's product-moment correlation coefficient. **[1]**

(b) Calculate the equation of the regression line of y on x. **[2]**

It is desired to know whether the regression line of y on x will provide a reliable estimate of y when $x = 0.75$.

(c) State one reason for believing that the estimate will be reliable. **[1]**

(d) State what further information is needed in order to determine whether the estimate is reliable. **[1]**

2 The average numbers of cars, lorries and buses passing a point on a busy road in a period of 30 minutes are 400, 80 and 17 respectively.

(a) Assuming that the numbers of each type of vehicle passing the point in a period of 30 minutes have independent Poisson distributions, calculate the probability that the total number of vehicles passing the point in a randomly chosen period of 30 minutes is at least 520. **[3]**

(b) Buses are known to run in approximate accordance with a fixed timetable.

Explain why this casts doubt on the use of a Poisson distribution to model the number of buses passing the point in a fixed time interval. **[1]**

3 Six red counters and four blue counters are arranged in a straight line in a random order.

Find the probability that

(a) no blue counter has fewer than two red counters between it and the nearest other blue counter, **[3]**

(b) no two blue counters are next to one another. **[3]**

4 The greatest weight WN that can be supported by a shelving bracket of traditional design is a normally distributed random variable with mean 500 and standard deviation 80.

A sample of 40 shelving brackets of a new design are tested and it is found that the mean of the greatest weights that the brackets in the sample can support is 473.0 N.

(a) Test at the 1% significance level whether the mean of the greatest weight that a bracket of the new design can support is less than the mean of the greatest weight that a bracket of the traditional design can support. **[7]**

(b) State an assumption needed in carrying out the test in part **(a)**. **[1]**

(c) Explain whether it is necessary to use the central limit theorem in carrying out the test. **[1]**

5 Five runners, *A*, *B*, *C*, *D* and *E*, take part in two different races.

Spearman's rank correlation coefficient for the orders in which the runners finish is calculated and a test for positive agreement is carried out at the 5% significance level.

(a) State suitable hypotheses for the test. **[1]**

(b) Find the largest possible value of $\sum d^2$ for which the result of the test is to reject the null hypothesis. **[3]**

(c) In the first race, the order in which the five runners finished was: *A*, *B*, *C*, *D*, *E*. In the second race, three of the runners finished in the same positions as in the first race. The result of the test is to reject the null hypothesis.

Find a possible order for the runners to finish in the second race. **[3]**

6 Yusha investigates the proportion of left-handed people living in two cities, *A* and *B*. He obtains data from random samples from the two cities. His results are shown in the table, in which *L* denotes "left-handed".

	L	L'
A	14	9
B	26	51

(a) Test at the 10% significance level whether there is association between being left-handed and living in a particular city. **[7]**

A person is chosen at random from one of the cities *A* and *B*.
Let *A* denote "the person lives in city *A*".

(b) State the relationship between P(L) and $P(L|A)$ according to the model implied by the null hypothesis of your test. **[1]**

(c) Use the data in the table to suggest a value for $P(L|A)$ given by an improved model. **[2]**

7 The random variable D has the distribution Geo(p). It is given that $\mathrm{Var}(D) = \frac{40}{9}$.

Determine

(a) $\mathrm{Var}(3D+5)$, **[1]**

(b) $\mathrm{E}(3D+5)$, **[6]**

(c) $\mathrm{P}(D > \mathrm{E}(D))$. **[3]**

8 A university course was taught by two different professors. Students could choose whether to attend the lectures given by Professor Q or the lectures given by Professor R. At the end of the course all the students took the same examination.

The examination marks of a random sample of 30 students taught by Professor Q and a random sample of 24 students taught by Professor R were ranked. The sum of the ranks of the students taught by Professor Q was 726.

Test at the 5% significance level whether there is a difference in the ranks of the students taught by the two professors. **[10]**

9 The continuous random variable T has cumulative distribution function

$$F(t) = \begin{cases} 0 & t < 0, \\ 1 - \mathrm{e}^{-0.25t} & t \geqslant 0. \end{cases}$$

(a) Find the cumulative distribution function of $2T$. **[3]**

(b) Show that, for constant k, $\mathrm{E}(\mathrm{e}^{kt}) = \dfrac{1}{1-4k}$.

You should state with a reason the range of values of k for which this result is valid. **[7]**

(c) T is the time before a certain event occurs.

Show that the probability that no event occurs between time $T = 0$ and time $T = \theta$ is the same as the probability that the value of a random variable with the distribution $\mathrm{Po}(\lambda)$ is 0, for a certain value of λ. You should state this value of λ in terms of θ. **[4]**

END OF QUESTION PAPER