OCR
Oxford Cambridge and RSA

Wednesday 15 June 2022 – Afternoon

A Level Further Mathematics A

Y542/01 Statistics

Time allowed: 1 hour 30 minutes

9459103657

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

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 pages 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 $g\,\text{m}\,\text{s}^{-2}$. 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.

Answer **all** the questions.

1 A researcher wishes to find people who say that they support a specific plan. Each day the researcher interviews people at random, one after the other, until they find one person who says that they support this plan. The researcher does not then interview any more people that day. The total number of people interviewed on any one day is denoted by R.

(a) Assume that in fact 1% of the population would say that they support the plan.

(i) State an appropriate distribution with which to model R, giving the value(s) of any parameter(s). **[1]**

(ii) Find $\mathrm{P}(50 < R \leqslant 150)$. **[2]**

The researcher incorrectly believes that the variance of a random variable X with **any** discrete probability distribution is given by the formula $[\mathrm{E}(X)]^2 - \mathrm{E}(X)$.

(b) Show that, for the type of distribution stated in part **(a)**, they will obtain the correct value of the variance, regardless of the value(s) of the parameter(s). **[2]**

2 The directors of a large company believe that there are more computer failures in the Head Office when temperatures are higher. They obtain data for the Head Office for the maximum temperature, T°C, and the number of computer failures, X, on each of 12 randomly chosen days.

(a) State which of the following words can be applied to T.

Dependent Independent Controlled Response **[1]**

The data is summarised as follows.

$n = 12 \quad \Sigma t = 261 \quad \Sigma x = 41 \quad \Sigma t^2 = 5869 \quad \Sigma x^2 = 311 \quad \Sigma tx = 1021$

(b) Calculate the value of the product moment correlation coefficient r. **[2]**

(c) The directors wish to investigate their belief using a significance test at the 1% level.

(i) Explain why a 1-tail test is appropriate in this situation. **[1]**

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

(d) One of the directors prefers the temperatures to be given in Fahrenheit (°F), rather than Centigrade (°C). The relationship between F and C is $\mathrm{F} = \frac{9}{5}\mathrm{C} + 32$.
State the value of r that would result from using temperatures in Fahrenheit in the calculation. **[1]**

3 **In this question you must show detailed reasoning.**

A discrete random variable V has the following probability distribution, where p and q are constants.

v	0	1	2	3
$\mathrm{P}(V = v)$	p	q	0.12	0.2

It is given that $\mathrm{E}(V) = \mathrm{Var}(V)$.

Determine the value of p and the value of q. **[8]**

4 The manager of a car breakdown service uses the distribution Po(2.7) to model the number of punctures, R, in a 24-hour period in a given rural area. The manager knows that, for this model to be valid, punctures must occur randomly and independently of one another.

(a) State a further assumption needed for the Poisson model to be valid. **[1]**

(b) State the value of the standard deviation of R. **[1]**

(c) Use the model to calculate the probability that, in a randomly chosen period of 168 hours, at least 22 punctures occur. **[3]**

The manager uses the distribution Po(0.8) to model the number of flat batteries in a 24-hour period in the same rural area, and he assumes that instances of flat batteries are independent of punctures. A day begins and ends at midnight, and a "bad" day is a day on which there are more than 6 instances, in total, of punctures and flat batteries.

(d) Assume first that both the manager's models are correct.

Calculate the probability that a randomly chosen day is a "bad" day. **[2]**

(e) It is found that 12 of the next 100 days are "bad" days.

Comment on whether this casts doubt on the validity of the manager's models. **[2]**

5 A company uses two drivers for deliveries.

Driver *A* charges a fixed rate of £80 per day plus £2 per mile travelled on that day.

Driver *B* charges a fixed rate of £120 per day plus £1.50 per mile travelled on that day.

On each working day the total distance, in miles, travelled by each driver is a random variable with the distribution N(83, 360).

(a) Find the probability that driver *A* charges the company less than £235.00 for a randomly chosen day's deliveries. **[4]**

(b) Find the probability that the total charge to the company of three randomly chosen days' deliveries by driver *A* is at least £300 more than the total charge of two randomly chosen days' deliveries by driver *B*. **[6]**

6 The random variable X was assumed to have a normal distribution with mean μ. Using a random sample of size 128, a significance test was carried out using the following hypotheses.

H_0: $\mu = 30$

H_1: $\mu > 30$

It was found that $\sum x = 3929.6$ and $\sum x^2 = 123\,483.52$. The conclusion of the test was to reject the null hypothesis.

(a) Determine the range of possible values of the significance level of the test. **[5]**

(b) It was subsequently found that X was not normally distributed.

Explain whether this invalidates the conclusion of the test. **[2]**

7 The continuous random variable X has probability density function

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

where k is a constant and n is a parameter whose value is positive.

It is given that the median of X is 0.8816 correct to 4 decimal places.

Ten independent observations of X are obtained.

Find the expected number of observations that are less than 0.8. **[8]**

8 The critical region for an $r\%$ two-tailed Wilcoxon signed-rank test, based on a large sample of size n, is $\{W_+ \leqslant 113\} \cup \{W_+ \geqslant 415\}$.

(a) Show that $n = 32$. **[3]**

(b) Using a suitable approximation, determine the value of r. **[4]**

9 The head teacher of a school believes that, on average, pupil absences on the days Monday, Tuesday, Wednesday, Thursday and Friday are in the ratio 3 : 2 : 2 : 2 : 3. The head teacher takes a random sample of 120 pupil absences. The results are as follows.

Day of week	Monday	Tuesday	Wednesday	Thursday	Friday
Number of absences	28	16	24	16	36

(a) Test at the 5% significance level whether these results are consistent with the head teacher's belief. **[7]**

A significance test at the 5% level is also carried out on a second, independent, random sample of n pupil absences. All the numbers of absences are integers. The ratio of the numbers of absences for each day in this sample is identical to the ratio of the numbers of absences for each day in the original sample of size 120.

(b) Determine the smallest value of n for which the conclusion of this significance test is that the data are not consistent with the head teacher's belief. **[3]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE