GCE

FURTHER MATHEMATICS

UNIT 2: FURTHER STATISTICS A

SAMPLE ASSESSMENT MATERIALS

(1 hour 30 minutes)

ADDITIONAL MATERIALS

In addition to this examination paper, you will need:

• a 12 page answer book;
• a Formula Booklet;
• a calculator;
• statistical tables (RND/WJEC Publications).

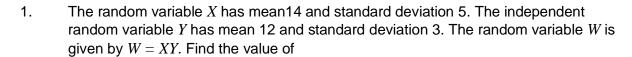**INSTRUCTIONS TO CANDIDATES**

Use black ink or black ball-point pen.
Answer **all** questions.
Sufficient working must be shown to demonstrate the **mathematical** method employed.
Unless the degree of accuracy is stated in the question, answers should be rounded appropriately.

INFORMATION FOR CANDIDATES

The number of marks is given in brackets at the end of each question or part-question.
You are reminded of the necessity for good English and orderly presentation in your answers.

1. The random variable X has mean 14 and standard deviation 5. The independent random variable Y has mean 12 and standard deviation 3. The random variable W is given by $W = XY$. Find the value of

    (a) E(W),   [1]

    (b) Var(W).   [6]

2. The queueing times, T minutes, of customers at a local Post Office are modelled by the probability density function

    $$f(t) = \frac{1}{2500}t(100 - t^2) \qquad \text{for } 0 \le t \le 10,$$
    $$f(t) = 0 \qquad \text{otherwise.}$$

    (a) Determine the mean queueing time.   [3]

    (b) (i)   Find the cumulative distribution function, $F(t)$, of T.

         (ii) Find the probability that a randomly chosen customer queues for more than 5 minutes.

         (iii) Find the median queueing time.   [10]

3. A class of 8 students sit examinations in History and Geography. The marks obtained by these students are given below.

Student	A	B	C	D	E	F	G	H
History mark	73	59	83	49	57	82	67	60
Geography mark	55	51	58	59	44	66	49	67

    (a) Calculate Spearman's rank correlation coefficient for this data set.   [6]

    (b) Hence determine whether or not, at the 5% significance level, there is evidence of a positive association between marks in History and marks in Geography.   [2]

    (c) Explain why it might not have been appropriate to use Pearson's product moment correlation coefficient to test association using this data set.   [1]

4. A year 12 student wishes to study at a Welsh university. For a randomly chosen year between 2000 and 2017 she collected data for seven universities in Wales from the Complete University Guide website. The data are for the variables:

    *   'Entry standards' – the average UCAS tariff score of new undergraduate students;

    *   'Student satisfaction' – a measure of student views of the teaching quality at the university taken from the National Student Survey (maximum 5);

    *   'Graduate prospects' – a measure of the employability of a university's first degree graduates (maximum 100);

    *   'Research quality' – a measure of the quality of the research undertaken in the university (maximum 4).

    (a) Pearson's product-moment correlation coefficients, for each pairing of the four variables, are shown in the table below.
    Discuss the correlation between graduate prospects and the other three variables.   [2]

Variable	Entry standards	Student satisfaction	Graduate prospects	Research quality
Entry standards	1			
Student satisfaction	-0.030	1		
Graduate prospects	0.772	0.236	1	
Research quality	0.866	0.066	0.827	1

    (b) Calculate the equation of the least squares regression line to predict 'Entry standards'(y) from 'Research quality'(x), given the summary statistics:

    $$\sum x = 22.24, \ \sum y = 2522, \ S_{xx} = 1.0542, \ S_{yy} = 20193.5, S_{xy} = 122.72.$$
                                                                                          [5]

    (c) The data for one of the Welsh universities are missing. This university has a research quality of 3.00. Use your equation to predict the entry standard for this university.   [2]

5.   The manager of a hockey team studies last season's results and puts forward the theory that the number of goals scored per match by her team can be modelled by a Poisson distribution with mean 2.0. The number of goals scored during the season are summarised below.

Goals scored	0	1	2	3	4 or more
Frequency	6	11	15	10	8

(a)   State suitable hypotheses to carry out a goodness of fit test.   [1]

(b)   Carry out a χ^2 goodness of fit test on this data set, using a 5% level of significance and draw a conclusion in context.   [9]

6.   Customers arrive at a shop such that the number of arrivals in a time interval of t minutes follows a Poisson distribution with mean $0.5t$.

(a)   Find the probability that exactly 5 customers arrive between 11 a.m. and 11.15 a.m.   [3]

(b)   A customer arrives at exactly 11 a.m.

   (i)   Let the next customer arrive at T minutes past 11 a.m. Show that

$$P(T > t) = e^{-0.5t}.$$

   (ii)   Hence find the probability density function, $f(t)$, of T.

   (iii)   Hence, giving a reason, write down the mean and the standard deviation of the time between the arrivals of successive customers. [7]

7.   The Pew Research Center's Internet Project offers scholars access to raw data sets from their research.

One of the Pew Research Center's projects was on teenagers and technology. A random sample of American families was selected to complete a questionnaire. For each of their children, between and including the ages of 13 and 15, parents of these families were asked:

Do you know your child's password for any of [his/her] social media accounts?

Responses to this question were received from 493 families. The table below provides a summary of their responses.

Parent know password	Age (years)			Total
	13	**14**	**15**	
Yes	76	75	67	218
No	66	103	106	275
Total	142	178	173	493

(a)   A test for significance is to be undertaken to see whether there is an association between whether a parent knows any of their child's social media passwords and the age of the child.

(i)   Clearly state the null and alternative hypotheses.

(ii)   Obtain the expected value that is missing from the table below, indicating clearly how it is calculated from the data values given in the table above.

Expected values:

Parent knows password	Age (years)		
	13	**14**	**15**
Yes	62.79	78.71	76.50
No		99.29	96.50

(iii)   Obtain the two chi-squared contributions that are missing from the table below.

Chi-squared contributions:

Parent knows password	Age (years)		
	13	14	15
Yes		0.175	1.180
No	2.203		0.935

The following output was obtained from the statistical package that was used to undertake the analysis:

Pearson chi-squared (2) = 7.409   *p*-value = 0.0305

(iv)   Indicate how the degrees of freedom have been calculated for the chi-squared statistic.

(v)   Interpret the output obtained from the statistical test in terms of the initial hypotheses.

[10]

(b)   Comment on the nature of the association observed, based on the contributions to the test statistic calculated in (a).   [2]