



GCE AS/A LEVEL

2305U20-1



Z22-2305U20-1

MONDAY, 23 MAY 2022 – AFTERNOON

FURTHER MATHEMATICS – AS unit 2 FURTHER STATISTICS A

1 hour 30 minutes

2305U201
01

ADDITIONAL MATERIALS

In addition to this examination paper, you will need:

- a WJEC pink 16-page answer booklet;
- 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.

Answers without working may not gain full credit.

Unless the degree of accuracy is stated in the question, answers should be rounded appropriately.

INFORMATION FOR CANDIDATES

The maximum mark for this paper is 70.

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.

Reminder: Sufficient working must be shown to demonstrate the **mathematical** method employed.

1. The probability distribution for the prize money, £ X per ticket, in a local fundraising lottery is shown below.

x	0	2	100	1000
$P(X = x)$	0.9	0.09	p	0.0001

(a) Calculate the value of p . [1]

(b) Find $E(X)$ and $\text{Var}(X)$. [4]

(c) (i) What is the minimum lottery ticket price that the organiser should set in order to make a profit in the long run?
(ii) Suggest why, in practice, people would be prepared to pay more than this minimum price. [2]

2. An economist suggested the rate of unemployment and the rate of wage inflation are independent. Amy sets about investigating this suggestion. She collects unemployment data and wage inflation data from a random sample of regions in the UK and decides that it is appropriate to carry out a significance test on Pearson's product moment correlation coefficient.

Amy's summary statistics for percentage unemployment, x , and percentage wage inflation, y , are shown below.

$$\begin{array}{lll} \sum x = 62.8 & \sum y = 19.4 & n = 10 \\ \sum x^2 = 413.44 & \sum y^2 = 46.16 & \sum xy = 113.16 \end{array}$$

(a) Calculate Pearson's product moment correlation coefficient for these data. [4]

(b) Carry out Amy's test at the 5% level of significance and state whether the economist's suggestion is reasonable. [4]

Amy also collects unemployment data and wage inflation data from a random sample of 10 regions in Spain and calculates Pearson's product moment correlation coefficient to be -0.2525 .

(c) Should this change Amy's opinion on the economist's suggestion above? What could she do to improve her investigation? [2]

(d) What assumption has Amy made in deciding that it is appropriate to carry out a significance test on Pearson's product moment correlation coefficient? [1]

3. Two basketball players, Steph and Klay, score baskets at random at a rate of 2.1 and 1.9 respectively per quarter of a game. Assume that baskets are scored independently, and that Steph and Klay each play all four quarters of the game.

(a) Stating the model that you are using, find the probability that they will score a combined total of exactly 20 baskets in a randomly selected game. [4]

(b) A quarter of a game lasts 12 minutes.

(i) State the distribution of the time between baskets for Steph. Give the mean and standard deviation of this distribution.

(ii) Given that Klay scores at the end of the third minute in a quarter of a game, find the probability that Klay doesn't score for the rest of the quarter. [5]

(c) When practising, Klay misses 4% of the free throws he takes. One week he takes 530 free throws. Calculate the probability that he misses more than 25 free throws. [2]

4. The continuous random variable R has probability density function $f(r)$ given by

$$f(r) = \begin{cases} kr(b-r) & \text{for } 1 \leq r \leq 4, \\ 0 & \text{otherwise,} \end{cases}$$

where k and b are positive constants.

(a) Explain why $b \geq 4$. [2]

(b) Given that $b = 4$,

(i) show that $k = \frac{1}{9}$,

(ii) find an expression for $F(r)$, valid for $1 \leq r \leq 4$, where F denotes the cumulative distribution function of R ,

(iii) find the probability that R lies between 2 and 3. [10]

5. John has a game that involves throwing a set of three identical, cubical dice with faces numbered 1 to 6. He wishes to investigate whether these dice are fair in terms of the number of sixes obtained when they are thrown. John throws the set of three dice 1100 times and records the number of sixes obtained for each throw. The results are shown in the table below.

Number of sixes	0	1	2	3
Frequency	625	384	81	10

Using these results, conduct a goodness of fit test and draw an appropriate conclusion. [11]

TURN OVER

6. An online survey on the use of social media asked the following question:

“Do you use any form of social media?”

The results for a total of 1953 respondents are shown in the table below.

Use social media	Age in years				Total
	18-29	30-49	50-64	65 or older	
Yes	310	412	348	196	1266
No	42	116	196	333	687
Total	352	528	544	529	1953

To test whether there is a relationship between social media use and age, a significance test is carried out at the 5 % level.

(a) State the null and alternative hypotheses. [1]

(b) Show how the expected frequency 228.18 is calculated in the table below. [1]

Expected values	Age in years			
	18-29	30-49	50-64	65 or older
Yes	228.18	342.27	352.64	342.92
No	123.82	185.73	191.36	186.08

(c) Determine the value of s in the table below. [2]

Chi-squared contributions	Age in years			
	18-29	30-49	50-64	65 or older
Yes	29.34	s	0.06	62.94
No	54.07	26.18	0.11	115.99

(d) Complete the significance test, showing all your working. [6]

(e) A student, analysing these data on a spreadsheet, obtains the following output.

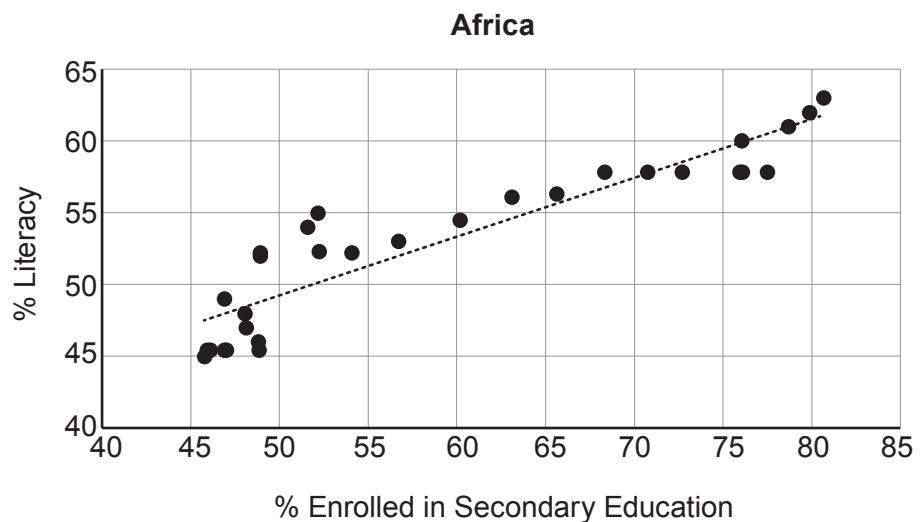
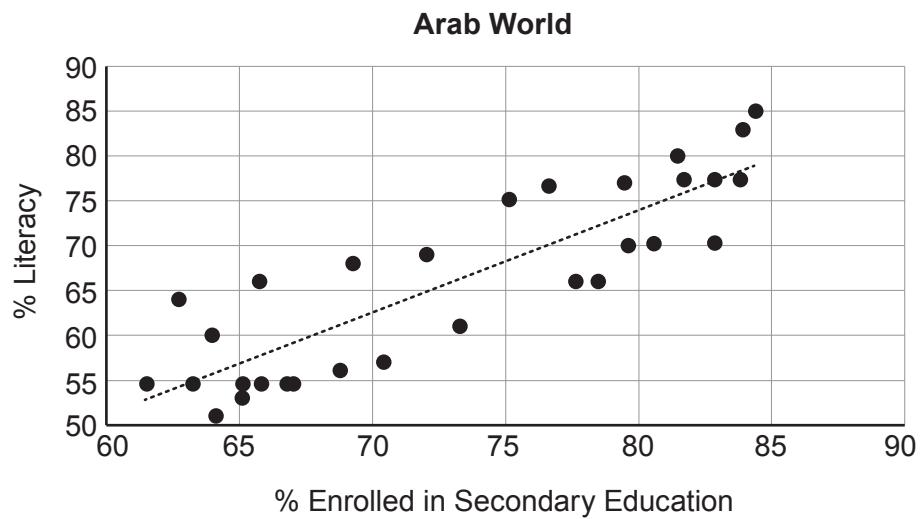
<code>CHISQ.TEST(B13:E14,I13:L14)</code>	<code>p value=</code>	<code>0.07894</code>	
--	-----------------------	----------------------	--

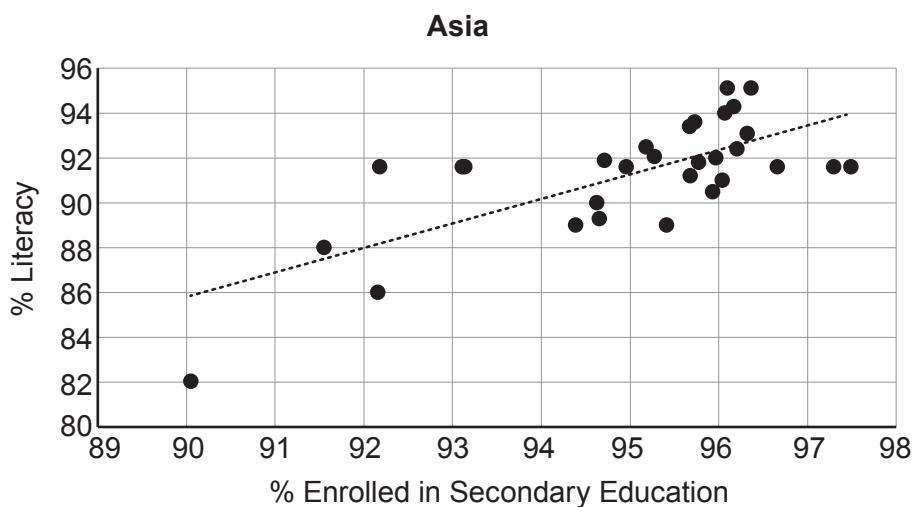
Explain why the student must have made an error in calculating the *p*-value.

[1]

TURN OVER

7. Data from a large dataset shows the percentage of children enrolled in secondary education and the percentage of the adult population who are literate. The following graphs show data from 30 randomly selected regions from each of the Arab World, Africa and Asia. In each case, the least squares regression line of '% Literacy' on '% Enrolled in Secondary Education' is shown.





(a) Calculate the equation of the least squares regression line of '% Literacy' (y) on '% Enrolled in Secondary Education' (x) for Asia, given the following summary statistics. [5]

$$\sum x = 2850.836 \quad \sum y = 2738.656 \quad S_{xx} = 88.42142$$

$$S_{yy} = 204.733 \quad S_{xy} = 96.60984 \quad n = 30$$

(b) The Arab World, Africa and Asia each contain a region where 70% are enrolled in secondary education. The three regression lines are used to estimate the corresponding % Literacy. Which of these estimates is likely to be the most reliable? Clearly explain your reasoning. [2]

END OF PAPER