2.5 Further Working with Data (A Level only)

Question Paper

Course	OCR A Level Maths: Statistics
Section	2. Data Presentation & Interpretation
Topic	2.5 Further Working with Data (A Level only)
Difficulty	Medium

Time allowed: **50**

Score: **/39**

Percentage: **/100**

Question 1

A teacher, Ms Pearman, claims that there is a positive correlation between the number of hours spent studying for a test and the percentage scored on it.

(a) Write down suitable null and alternative hypotheses to test Ms Pearman's claim.

[1 mark]

Question 1

Ms Pearman takes a random sample of 25 students and gives them a week to prepare for a test. She records the percentage they score in the test, s %, and the amount of revision they did, h hours.

Ms Pearson calculates the product moment correlation coefficient for these data as $r = 0.874$.

(b) Given that the p-value for the test statistic $r = 0.874$ is 0.0217, test at the 5% level of significance whether Ms Pearman's claim is justified.

[2 marks]

Question 1

Ms Pearman decides to use a linear regression model for these data. She calculates the equation for the regression line of s on h to be $s = 21.3 + 5.29h$.

(c) (i) Give an interpretation of the value 21.3 in context.

(ii) Give an interpretation of the value 5.29 in context.

[2 marks]

Question 2

The following table shows the number of hours spent learning to drive, d, and the number of mistakes made in the driving test, m, of ten college students.

d	48	51	51	57	61	68	70	72	73	75
m	19	21	17	12	8	16	7	4	0	1

The product moment correlation coefficient for these data is $r = -0.869$. A driving instructor, Dave, believes there is a negative correlation between the number of hours spent learning to drive and the number of mistakes made in the driving test.

(a) (i) Write down suitable null and alternative hypotheses to test Dave's claim.

(ii) Test, at the 1% level of significance, whether Dave's claim is justified, given that the relevant critical value is -0.7155.

[3 marks]

Question 2

Dave calculates the equation of the regression line of m on d to be $m = 50.7 - 0.642d$.

(b) State, giving a reason, whether or not the correlation coefficient is consistent with the use of a linear regression model.

[1 mark]

Question 2

(c) (i) Explain why the linear regression model could be unreliable for predicting the number of mistakes a student would make on their driving test after learning for 30 hours.

(ii) By considering a student who has spent 80 hours learning to drive, give a limitation to the linear regression model.

[2 marks]

Question 3

The table below shows data from the United States regarding annual per capita chicken consumption (in pounds) and the unemployment rate (% of population) between the years 2005 and 2014:

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Chicken consumption (pounds)	86.4	86.9	85.5	83.8	80.0	82.8	83.3	80.8	82.3	83.8
Unemployment rate (%)	5.08	4.62	4.62	5.78	9.25	9.63	8.95	8.07	7.38	6.17

The product moment correlation coefficient for these data is $r = -0.821$.
The critical values for a 10% two-tailed test are ± 0.5495.

(a) State what is measured by the product moment correlation coefficient.

[1 mark]

Question 3

(b) (i) Write down suitable null and alternative hypotheses for a two-tailed test of the correlation coefficient.

(ii) Show that, at the 10% level of significance, there is evidence that the correlation coefficient is different from zero.

[3 marks]

Question 3

A newspaper's headline states:

"Eating chicken is the secret to reducing the unemployment rate in the US!"

(c) Explain whether this headline is fully justified.

[1 mark]

Question 4

Jessica is researching whether there is a correlation between the productivity of university students and the number of hours sleep they get per night.

(a) Write suitable null and alternative hypotheses to test for linear correlation.

[1 mark]

Question 4

Jessica takes a random sample of 25 students, measures their productivity during the day, and records how many hours sleep they had during the previous night. She calculates the product moment correlation coefficient and finds that $r = -0.107$.

The table below gives the critical values, for different significance levels, of the product moment correlation coefficient, r, for a sample of size 25.

One tail	10%	5%	2.5%	1%	0.5%	One tail
Two tail	20%	10%	5%	2%	1%	Two tail
	0.2653	0.3365	0.3961	0.4622	0.5052	

Jessica wishes to test, at the 10% level of significance, whether there is evidence that the correlation coefficient for the population is different from zero.

(b) (i) Find the critical regions for Jessica's test.

(ii) Show that, at the 10% level of significance, there is no evidence of a linear correlation.

[3 marks]

Question 4

(c) State, with a reason, whether there could be a relationship between students' hours of sleep and their productivity.

[1 mark]

Question 5

During a zombie attack, Richard suspects that the number of flies in the area, f, is dependent on the number of zombies, z.
Richard is trying to decide whether the correlation is linear or non-linear, so he uses a graphical software package to plot two scatter graphs. **Figure 1** shows the graph of f plotted against z, and **Figure 2** show the graph of $\log f$ plotted against $\log z$.

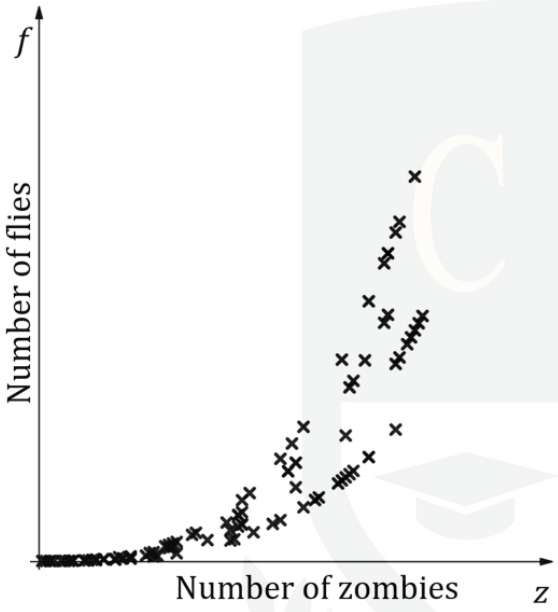


Figure 1

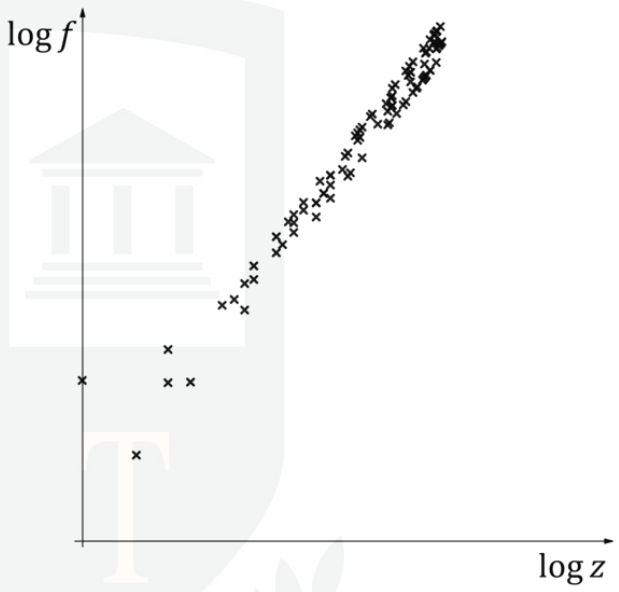


Figure 2

Richard calculates the product moment correlation coefficient for each graph. One value is found to be 0.847 while the other is 0.985.

(a) State, with a reason, which PMCC value corresponds with **Figure 2**.

[2 marks]

Question 5

(b) Test, using a 5% level of significance, whether there is positive linear correlation in the graph shown in **Figure 2**. State your hypotheses clearly.
You are given that the critical value for this test is 0.1654.

[3 marks]

Question 5

(c) State, with a reason, whether the relationship between number of zombies and number of flies is better represented as linear or non-linear.

[1 mark]

Question 6

Nicole is a Biologist studying the growth of bacteria. She records the number of bacteria on an organism every hour. The table below shows her results for the first eight hours.

Hours (t)	1	2	3	4	5	6	7	8
Number of bacteria (B)	10	50	170	520	1730	5200	17020	58140

Nicole calculates the product moment correlation coefficient for this data as $r = 0.735$. The table below gives the critical values, for different significance levels, of the product moment correlation coefficient, r, for a sample of size 8.

One tail	10%	5%	2.5%	1%	0.5%	One tail
Two tail	20%	10%	5%	2%	1%	Two tail
	0.5067	0.6215	0.7067	0.7887	0.8343	

Nicole claims that there is a positive linear correlation between the number of hours and the number of bacteria.

(a) Test, at the 1% level of significance, whether Nicole's claim is justified. State your hypotheses clearly.

[3 marks]

Question 6

Mariam, Nicole's lab assistant, claims that there is an exponential relationship between the two variables. To test this Mariam calculates the values of $\ln(B)$ for the different values of t.

(b) Complete the table, giving your answers to three decimal places.

t	1	2	3	4	5	6	7	8
$\ln(B)$	2.303	3.912	5.136	6.254	7.456	8.556		

[2 marks]

Question 6

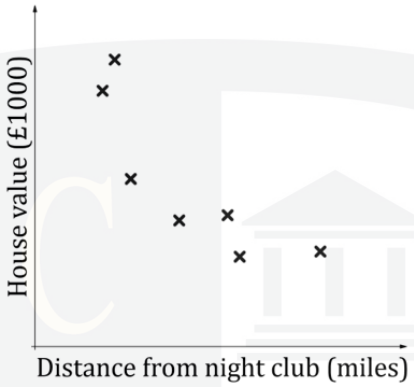
(c) Given that the product moment correlation coefficient for these eight pairs of data is $r = 0.999$, comment on Mariam's claim that there is an exponential relationship between B and t.

[1 mark]

Question 7

An estate agent, Terry, claims that there is a correlation between the value of a house (£1000) and the distance between that house and the nearest nightclub (miles).

Terry has a database containing over 100 houses and he takes a random sample of seven houses to investigate his claim. The scatter graph below shows the results:



(a) Terry calculates the product moment correlation coefficient as $r = 0.837$. Using the scatter graph, explain how you know Terry's PMCC value is incorrect.

[1 mark]

Question 7

Terry corrects his mistake and calculates the correct PMCC as $r = -0.837$.
The table below gives the critical values, for different significance levels, of the product moment correlation coefficient, r, for a sample of size 7.

One tail	10%	5%	2.5%	1%	0.5%	One tail
Two tail	20%	10%	5%	2%	1%	Two tail
	0.5509	0.6694	0.7545	0.8329	0.8745	

(b) (i) Write down suitable null and alternative hypotheses for a two-tailed test to investigate Terry's claim.

(ii) Test Terry's claim using a 5% level of significance.

[3 marks]

Question 7

(c) State, giving a reason, whether the conclusion to the test would be different if a 1% level of significance had been used.

[1 mark]

Question 7

(d) Suggest one way in which Terry could improve his investigation.

[1 mark]