QQQ – Statistics Yr2 - Chapter 1 – Correlation & Regression

Total Marks: 25

(25 = Platinum, 23 = Gold, 20 = Silver, 18 = Bronze)

1.
An engineer believes that there is a relationship between the CO_2 emissions and fuel consumption for cars.

A random sample of 40 different car models (old and new) was taken and the CO_2 emission figures, e grams per kilometre, and fuel consumption, f miles per gallon, were recorded, as shown in Figure 1. The engineer calculates the product moment correlation coefficient for the 40 cars and obtains $r = -0.803$.

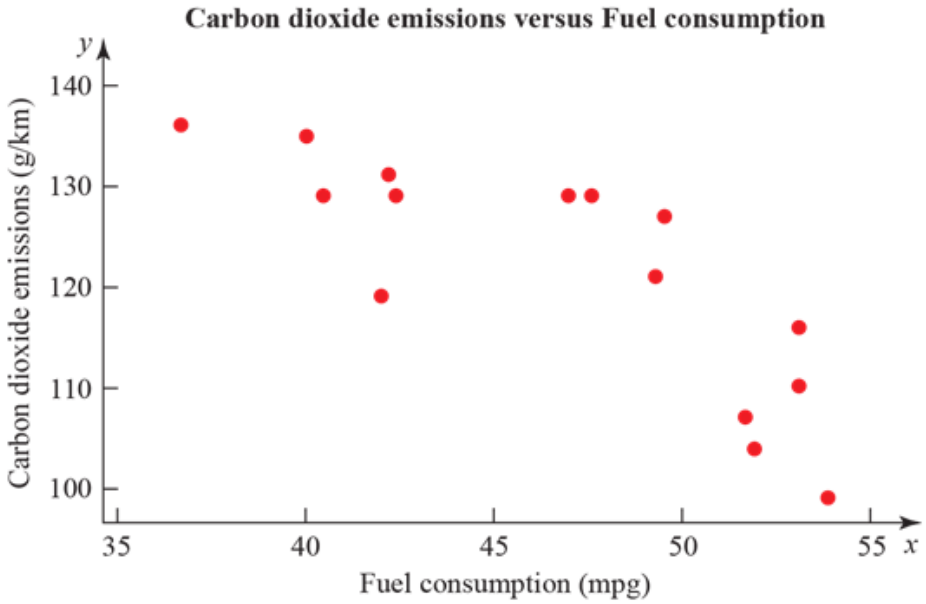

Figure 1

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

(1)

(b) State, with a reason, whether a linear regression model based on these data is reliable or not for a car when the fuel consumption is 60 mpg.

(1)

(c) For the linear regression model $e = 198 - 1.71 \times f$ write down the explanatory variable.

(1)

(d) State the definition of a hypothesis test.

(1)

(e) Test at 1% significance level whether or not the product moment correlation coefficient for CO_2 emissions and fuel consumption is less than zero. State your hypotheses clearly.

(3)

(Total 7 marks)

2.

Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £w, and the average weekly temperature, t°C, for 8 weeks during the summer.
The product moment correlation coefficient for these data is −0.915

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative.

(3)

(b) Suggest a possible reason for this correlation.

(1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.

(1)

(d) State, giving a reason, which variable would be the explanatory variable.

(1)

Tessa calculated the linear regression equation as $w = 10\ 755 - 171t$

(e) Give an interpretation of the gradient of this regression equation.

(1)

3.

The table shows some data collected on the temperature, in °C, of a cup of coffee, c, and the time, t in minutes, after which it was made.

t	0	2	4	5	7	11	13	17	25
c	81.9	75.9	70.1	65.1	60.9	51.9	50.8	45.1	39.2

The data is coded using the changes of variable $x = t$ and $y = \log_{10} c$.

The regression line of y on x is found to be $y = 1.89 - 0.0131x$.

(a) Given that the data can be modelled by an equation of the form $c = ab^t$ where a and b are constants, find the values of a and b.

(3)

(b) Give an interpretation of the constant b in this equation.

(1)

(c) Explain why this model is not reliable for estimating the temperature of the coffee after an hour.

(1)

(Total 5 marks)

4.

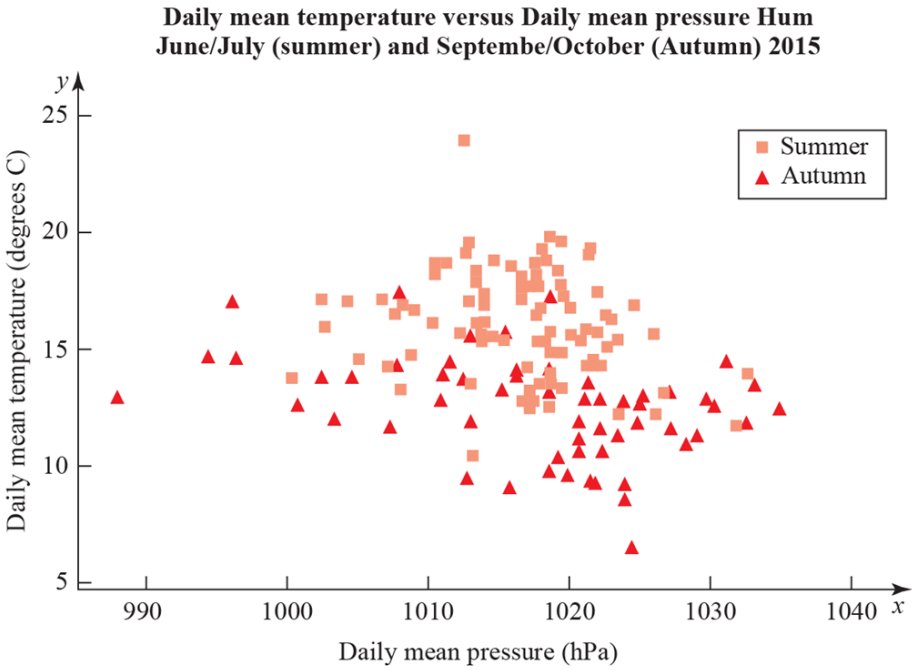

Figure 2

To investigate if there is a correlation between daily mean temperature (°C) and daily mean pressure (hPa) the location Hurn 2015 was randomly selected from:

Camborne 2015	Camborne 1987
Hurn 2015	Hurn 1987
Leuchars 2015	Leuchars 1987
Leeming 2015	Leeming 1987
Heathrow 2015	Heathrow 1987

(Source: Pearson Edexcel GCE AS and A Level Mathematics data set.)

(a) State the definition of a test statistic.

(1)

(b) The product moment correlation coefficient between daily mean temperature and daily mean pressure for these data is −0.258 with a p-value of 0.001. Use a 5% significance level to test whether or not there is evidence of a correlation between the daily mean temperature and daily mean pressure.

(3)

(c) The scatter diagram in Figure 2 shows daily mean temperature versus daily mean pressure, by season, for Hurn 2015. Give two interpretations on the split of the data between summer and autumn.

(2)

Solutions (

1.

1a	Linear association between e and f.	**B1**	1.2	2nd Know and understand the language of correlation and regression.
		(1)		
1b	It requires extropolation and hence it may be unreliable.	**B1**	1.2	4th Understand the concepts of interpolation and extrapolation.
		(1)		
1c	Fuel consumption (f)	**B1**	1.2	2nd Know and understand the language of correlation and regression.
1d	A hypothesis test is a statistical test that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.	**B1**	1.2	5th Understand the language of hypothesis testing.
		(1)		
1e	$H_0 : \rho = 0$, $H_1 : \rho < 0$ Critical value = -0.3665 $-0.803 < -0.3665$ (test statistic in critical region) Reject H_0 There is evidence that the product moment correlation coeficient for CO_2 emissions and fuel consumption is less than zero.	**B1** **M1** **A1**	2.5 1.1b 2.2b	6th Carry out a hypothesis test for zero correlation.
		(3)		

(7 marks)

2.

(a)	$H_0 : \rho = 0$ $\quad H_1 : \rho < 0$ Critical value: -0.6215 (Allow any cv in range $0.5 < \lvert cv \rvert < 0.75$) $r < -0.6215$ so significant result and there is evidence of a negative correlation between w and t	B1 M1 A1 (3)	2.5 1.1a 2.2b
(b)	e.g. As temperature increases people spend more time on the beach and less time shopping (o.e.)	B1 (1)	2.4
(c)	Since r is close to -1 , it **is** consistent with the suggestion	B1 (1)	2.4
(d)	t will be the explanatory variable since sales are likely to depend on the temperature	B1 (1)	2.4
(e)	Every degree rise in temperature leads to a drop in weekly earnings of £171	B1 (1)	3.4
		(7 marks)	

3.

2a	$\log_{10} c = 1.89 - 0.0131t$ $c = 10^{1.89-0.0131t}$ $c = 77.6 \times 0.970^t$ (3 s.f.)	**M1** **M1** **A1**	1.1a 1.1b 1.1b	6th Understand exponential models in bivariate data.
		(3)		
2b	b is the proportional rate at which the temperature changes per minute.	**A1**	3.2a	6th Understand exponential models in bivariate data.
		(1)		
2c	Extrapolation/out of the range of the data.	**A1**	2.4	4th Understand the concepts of interpolation and extrapolation.
		(1)		

(5 marks)

4.

8a	A statistic that is calculated from sample data in order to test a hypothesis about a population.	**B1**	1.2	5th Understand the language of hypothesis testing.
		(1)		
8b	$H_0 : \rho = 0$, $H_1 : \rho \neq 0$ p-value < 0.05 There is evidence to reject H_0 There is evidence (at 5% level) of a correlation between the daily mean temperature and daily mean pressure.	**B1** **M1** **A1**	2.5 1.1b 2.2b	6th Carry out a hypothesis test for zero correlation.
		(3)		
8c	Two sensible interpretations or observations. For example, Two distinct distributions Similar gradients of regression line. Similar correlations for each season. Lower temperaure in autumn. More spread for the daily mean pressure in autumn.	**B2**	3.2a	4th Use the principles of bivariate data analysis in the context of the large data set.
		(2)		

(6 marks)