QQQ – Statistics Yr2 - Chapter 1 – Correlation & Regression

Total Marks: 21
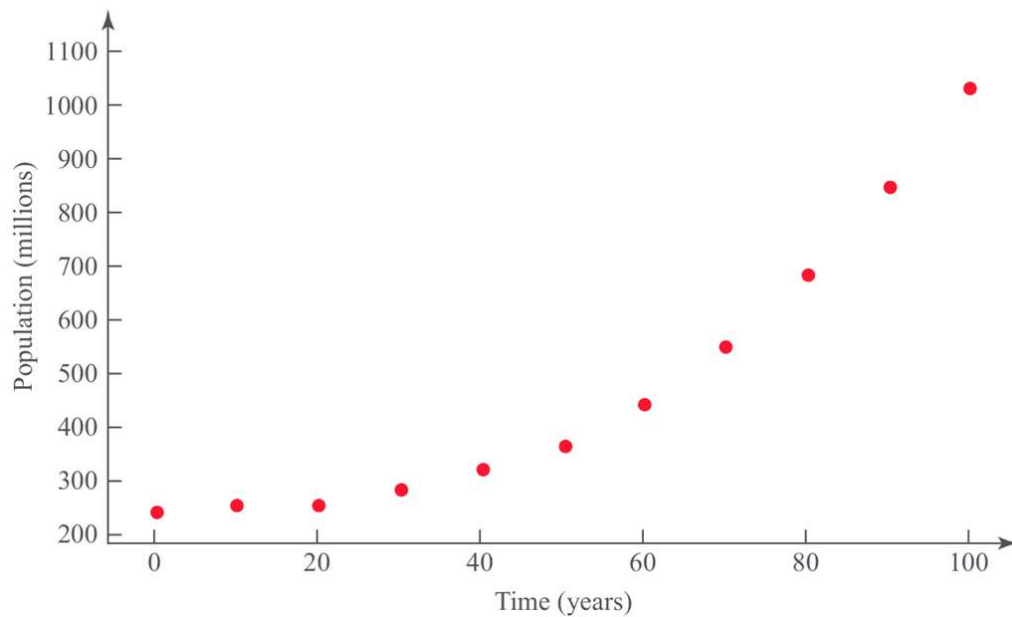
(21 = Platinum, 19 = Gold, 17 = Silver, 15 = Bronze)

1.

The data and scatter diagram in Figure 1 show the population, p, in millions, of a country taken t years since their first census.

t	0	10	20	30	40	50	60	70	80	90	100
p	238.4	252.1	251.3	279	318.7	361.1	439.2	548.2	683.3	846.4	1028.7

Population versus number of years since first census for a country

(a) Give a reason why the data is coded using the changes of variable $x = t$ and $y = \log_{10} p$.

(1)

(b) The product moment correlation coefficient for the coded data is $r = 0.9735$. Comment on r for this model.

(2)

(c) With reference to your answer to part (b), state whether a model in the form $p = ab^t$, where a and b are constants, is a good fit for this data.

(2)

(Total 5 marks)

2.

To investigate if there is a correlation between daily mean pressure (hPa) and daily mean wind speed (kn) 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.)

The statistical software output for these data is shown in Figure 3 below.
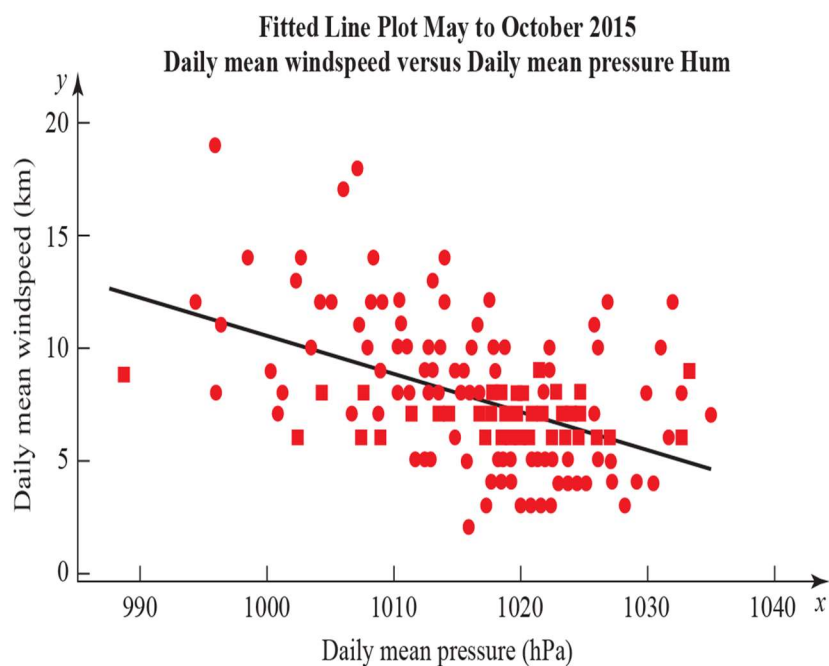
Figure 3

Correlation coefficient.
Daily mean winds and Daily mean pressure = −0.477 p-value < 0.001.
Regression summary output for daily mean wind speed versus daily mean pressure.

	Coefficients	Lower 95%	Upper 95%
Intercept	180.00	133.5424	226.4128
Daily Mean Pressure (hPa) Gradient	−0.1694	−0.21512	−0.12377

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

(1)

(b) Comment on the correlation between the two variables.

(1)

(c) Give an interpretation of the correlation between the two variables.

(1)

(d) Test at 5% significance level whether or not the product moment correlation coefficient for the population is less than zero. State your hypotheses clearly.

(3)

(e) Write down the regression model for daily mean wind speed versus daily mean pressure.

(2)

(f) Interpret the gradient of the line of regression stated in part **e**.

(1)

(g) The regression model (equation of regression) was used to predict the daily mean wind speed of 11.15 knots for a daily mean pressure of 995 hPa. Comment on the accuracy of this prediction.

(1)

3.

A meteorologist believes that there is a relationship between the daily mean windspeed, w kn, and the daily mean temperature, t °C. A random sample of 9 consecutive days is taken from past records from a town in the UK in July and the relevant data is given in the table below.

t	13.3	16.2	15.7	16.6	16.3	16.4	19.3	17.1	13.2
w	7	11	8	11	13	8	15	10	11

The meteorologist calculated the product moment correlation coefficient for the 9 days and obtained $r = 0.609$

(a) Explain why a linear regression model based on these data is unreliable on a day when the mean temperature is 24 °C

(1)

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

(1)

(c) Stating your hypotheses clearly test, at the 5% significance level, whether or not the product moment correlation coefficient for the population is greater than zero.

(3)

Using the same 9 days a location from the large data set gave $\bar{t} = 27.2$ and $\bar{w} = 3.5$

(d) Using your knowledge of the large data set, suggest, giving your reason, the location that gave rise to these statistics.

(1)