

QQQ – Statistics Yr2 - Chapter 1 – Correlation & Regression

Total Marks: 21

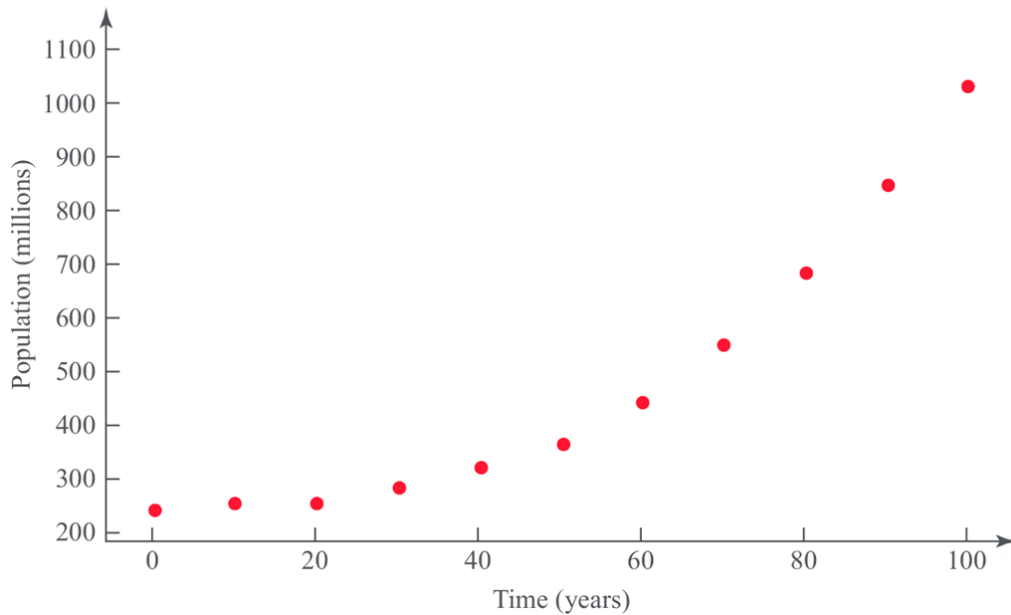
(21 = Platinum, 19 = Gold, 17 = Silver, 15 = Bronze)

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.

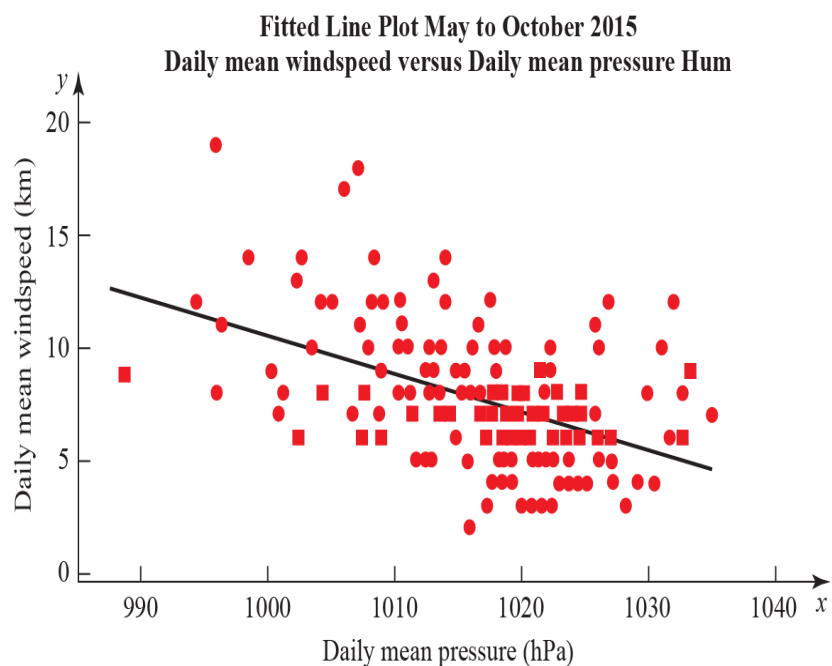


Figure 3

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)

Solutions (

1.

4a	The data seems to follow an exponential distribution.	B1	2.4	6th Understand exponential models in bivariate data.
		(1)		
4b	$r = 0.9735$ is close to 1	B1	2.2a	2nd
	which gives a strong positive correlation.	B1	2.4	Know and understand the language of correlation and regression.
		(2)		
4c	Model is a good fit with a reason. For example, Very strong positive linear correlation between t and $\log_{10} p$. The transformed data points lie close (enough) to a straight line.	B2	3.2a	6th Understand exponential models in bivariate data.
		(2)		

(5 marks)

2.

G4	Scheme	Marks	AOs	Pearson Progression Step and Progress descriptor
a	Linear association between two variables.	B1	1.2	2nd Know and understand the language of correlation and regression.
		(1)		
b	Negative correlation.	B1	1.2	2nd Know and understand the language of correlation and regression.
		(1)		

c	As daily mean pressure increases (rises) daily mean wind speed decreases (falls) in Hurn May to October in 2015. or As daily mean pressure decreases (falls) daily mean wind speed increases (rises) in Hurn May to October in 2015.	B1	3.2	5th Interpret the PPMC as a measure of correlation.
		(1)		
d	$H_0 : \rho = 0, H_1 : \rho < 0$ p -value < 0.05 There is evidence to reject H_0 . There is (strong) evidence of negative correlation between the daily mean wind speed and daily mean pressure.	B1 M1 A1	2.5 1.1b 2.2b	6th Carry out a hypothesis test for zero correlation.
		(3)		
e	Daily mean wind speed = $180 - 0.170 \times$ daily mean pressure.	B2	1.1b	4th Use the principles of bivariate data analysis in the context of the large data set.
		(2)		
f	The regression model suggests for every hPa increase in daily mean pressure the daily mean wind speed decreases by 0.1694 knots. or The regression model suggests for every hPa decrease in daily mean pressure the daily mean wind speed increases by 0.1694 knots.	B1	3.2	4th Use the principles of bivariate data analysis in the context of the large data set.
		(1)		
g	Sensible comment. For example, Not very accurate as very few or no points Not very accurate as near the bottom range for the data.	B1	3.5b	4th Make predictions using the regression line within the range of the data.
		(1)		
(10 marks)				
Notes				
e	B1 $y = 180.0 - 0.1694x$ unless x and y are defined.			

3.

2(a)	e.g. It requires extrapolation so will be unreliable (o.e.)	B1	1.2
		(1)	
(b)	e.g. Linear association between w and t	B1	1.2
		(1)	
(c)	$H_0: \rho = 0$ $H_1: \rho > 0$	B1	2.5
	Critical value 0.5822	M1	1.1a
	Reject H_0		
	There is evidence that the product moment correlation coefficient is greater than 0	A1	2.2b
		(3)	
(d)	Higher \bar{t} suggests overseas and not Perth...lower wind speed so perhaps not close to the sea so suggest Beijing	B1	2.4
		(1)	

(6 marks)

(a)

B1: for a correct statement (unreliable) with a suitable reason

(b)

B1: for a correct statement

(c)

B1: for both hypotheses in terms of ρ

M1: for selecting a suitable 5% critical value compatible with their H_1

A1: for a correct conclusion stated

(d)

B1: for suggesting Beijing with some supporting reason based on t or w

Allow Jacksonville with a reason based just on higher \bar{t}