
Stats1 Chapter 7: Correlation

Chapter Practice

Key Points

Summary of key points

- 1 If $y = ax^n$ for constants a and n then $\log y = \log a + n \log x$
- 2 If $y = kb^x$ for constants k and b then $\log y = \log k + x \log b$
- 3 The **product moment correlation coefficient** describes the linear correlation between two variables. It can take values between -1 and 1 .
- 4 For a one-tailed test use either:
 - $H_0: \rho = 0, H_1: \rho > 0$ or
 - $H_0: \rho = 0, H_1: \rho < 0$For a two-tailed test use:
 - $H_0: \rho = 0, H_1: \rho \neq 0$

Chapter Exercises

- 1 Conor uses a 3D printer to produce various pieces for a model. He records the time taken, t hours, to produce each piece, and its base area, $x \text{ cm}^2$.

Base area, $x \text{ (cm}^2\text{)}$	1.1	1.3	1.9	2.2	2.5	3.7
Time, $t \text{ (hours)}$	0.7	0.9	1.5	1.8	2.2	3.8

- a Calculate the product moment correlation coefficient between $\log x$ and $\log t$. (2 marks)
- b Use your answer to part a to explain why an equation of the form $t = ax^n$, where a and n are constants, is likely to be a good model for the relationship between x and t . (1 mark)
- c The regression line of $\log t$ on $\log x$ is given as $\log t = -0.210 + 1.38 \log x$. Determine the values of the constants a and n in the equation given in part b. (2 marks)
- 2 The table shows some data collected on the temperature in $^{\circ}\text{C}$ of a chemical reaction (t) and the amount of dry residue produced (d grams).

Temperature, $t \text{ (}^{\circ}\text{C)}$	38	51	72	83	89	94
Dry residue, $d \text{ (grams)}$	4.3	11.7	58.6	136.7	217.0	318.8

The data are coded using the changes of variable $x = t$ and $y = \log d$. The regression line of y on x is found to be $y = -0.635 + 0.0334x$.

- a Given that the data can be modelled by an equation of the form $d = ab^t$ where a and b are constants, find the values of a and b . (3 marks)
- b Explain why this model is not reliable for estimating the amount of dry residue produced when the temperature is 151°C . (1 mark)
- 3 The product moment correlation coefficient for a person's age and his score on a memory test is -0.86 . Interpret this value.

Chapter Exercises

- 4 Each of 10 cows was given an additive (x) every day for four weeks to see if it would improve the milk yield (y). At the beginning, the average milk yield per day was 4 gallons. The milk yield of each cow was measured on the last day of the four weeks. The data collected is shown in the table.

Cow	A	B	C	D	E	F	G	H	I	J
Additive, x (25 gm units)	1	2	3	4	5	6	7	8	9	10
Yield, y (gallons)	4.0	4.2	4.3	4.5	4.5	4.7	5.2	5.2	5.1	5.1

- a By drawing a scatter diagram or otherwise, suggest the maximum amount of additive that should be given to the cows to maximise yield.
- b Use your calculator to find the value of the product moment correlation coefficient for the first seven cows.
- c Without further calculation, write down, with a reason, how the product moment correlation coefficient for all 10 cows would differ from your answer to b.
- 5 The following table shows the engine size (c), in cubic centimetres, and the fuel consumption (f), in miles per gallon to the nearest mile, for 10 car models.

c (cm ³)	1000	1200	1400	1500	1600	1800	2000	2200	2500	3000
f (mpg)	46	42	43	39	41	37	35	29	28	25

- a Use your calculator to find the value of the product moment correlation coefficient between c and f . (1 mark)
- b Interpret your answer to part a. (2 marks)

Chapter Exercises

- 6 As part of a survey in a particular profession, age, x years, and yearly salary, $\pounds y$ thousands, were recorded.

The values of x and y for a randomly selected sample of ten members of the profession are as follows:

x	30	52	38	48	56	44	41	25	32	27
y	22	38	40	34	35	32	28	27	29	41

- a Calculate, to 3 decimal places, the product moment correlation coefficient between age and salary. **(1 mark)**

It is suggested that there is no correlation between age and salary.

- b Test this suggestion at the 5% significance level, stating your null and alternative hypotheses clearly. **(3 marks)**

- 7 A machine hire company kept records of the age, X months, and the maintenance costs, $\pounds Y$, of one type of machine. The following table summarises the data for a random sample of 10 machines.

Machine	A	B	C	D	E	F	G	H	I	J
Age, X	63	12	34	81	51	14	45	74	24	89
Maintenance costs, Y	111	25	41	181	64	21	51	145	43	241

- a Calculate, to 3 decimal places, the product moment correlation coefficient. **(1 mark)**

It is believed that there is a relationship between the age and maintenance cost of these machines.

- b Using a 5% level of significance and quoting from the table of critical values, interpret your correlation coefficient. Use a two-tailed test and state clearly your null and alternative hypotheses. **(3 marks)**

Chapter Exercises

- 8 The data below show the height above sea level, x metres, and the temperature, y °C, at 7.00 a.m., on the same day in summer at nine places in Europe.

Height, x (m)	1400	400	280	790	390	590	540	1250	680
Temperature, y (°C)	6	15	18	10	16	14	13	7	13

The product moment correlation coefficient is -0.975 . Use this value to test for negative correlation at the 5% significance level. Interpret your result in context. **(3 marks)**

- 9 The ages, in months, and the weights, in kg, of a random sample of nine babies are shown in the table below.

Baby	A	B	C	D	E	F	G	H	I
Age, x	1	2	2	3	3	3	4	4	5
Weight, y	4.4	5.2	5.8	6.4	6.7	7.2	7.6	7.9	8.4

The product moment correlation coefficient between weight and age for these babies was found to be 0.972 . By testing for positive correlation at the 5% significance level interpret this value. **(3 marks)**

- 10 An ice-cream seller believes that there is a positive correlation between the amount of sunshine and sales of ice cream. He collects data on six days during June 2015 at his 'pitch' in Camborne:

Sunshine (hours)	4.2	7.9	13.8	8.7	6.2	0.7
Ice-cream sales (£100s)	7.0	8.3	12.4	8.1	7.9	6.2

- a Calculate the product moment correlation coefficient for these data. **(1 mark)**
- b Carry out a hypothesis test to determine, at the 5% level, if there is significant evidence in support of the ice-cream seller's belief. State your hypotheses clearly. **(3 marks)**

Chapter Exercises

- 11 A meteorologist believes that there is a positive correlation between daily mean windspeed and daily maximum gust. She collects data from the large data set for 5 days during August 2015 in the town of Hurn.

Mean windspeed (knots)	4	7	7	8	5
Daily maximum gust (knots)	14	22	18	20	17

© Crown Copyright Met Office

By calculating the product moment correlation coefficient for these data, test at the 5% level of significance whether there is evidence to support the meteorologist's claim. State your hypotheses clearly. **(4 marks)**

- 12 The table shows data from the large data set on the daily mean air temperature and the daily mean pressure during May and June 2015 in Beijing.

Temperature (°C)	17.5	18.5	18.0	24.6	22.2	23.1	27.3
Pressure (hPa)	1010	1011	1012	997	1009	998	1002

© Crown Copyright Met Office

Test at the 2.5% level of significance the claim that there is negative correlation between the daily mean air temperature and the daily mean pressure. State your hypotheses clearly. **(4 marks)**

Chapter Exercises

Large data set

You will need access to the large data set and spreadsheet software to answer these questions.

- 1 a** Take a random sample of size 20 from the data for Heathrow in 2015, and record the daily mean air temperature and daily total rainfall.
 - b** Calculate the product moment correlation coefficient between these variables for your sample.
 - c** Test, at the 5% level of significance, the claim that there is a correlation between the daily mean air temperature and the daily total rainfall.
-
- 2 a** State with a reason whether you would expect to find a relationship between daily mean total cloud cover and daily mean visibility.
 - b** Use a random sample from the large data set to test for this relationship. You should state clearly:
 - Your sample size and location
 - Your sampling method
 - The hypotheses and significance level for your test
 - A conclusion in the context of the question

Hint

You might be able to use the **Correl** or **CorrelationCoefficient** commands in your spreadsheet software to calculate the PMCC.

Chapter Answers

- 1 **a** 0.9998
b r is close to 1, so a graph of $\log t$ against $\log x$ shows a straight line, suggesting that the relationship is in the form $t = ax^n$.
c $n = 1.38$, $a = 0.617$ (3 s.f.)
- 2 **a** $a = 0.232$ (3 s.f.), $b = 1.08$ (3 s.f.)
b 151°C is outside the range of the data (extrapolation).
- 3 As a person's age increases, their score on the memory test decreases.
- 4 **a** Each cow should be given 7 units. The yield levels off at this point.
b 0.952 (3 s.f.)
c It would be less than 0.952. The yield of the last 3 cows is no greater than that of the 7th cow.
- 5 **a** -0.972
b There is strong negative correlation. As c increases, f decreases.
- 6 **a** 0.340 (3 d.p.)
b $H_0: \rho = 0$, $H_1: \rho \neq 0$, critical value $= \pm 0.6319$. Accept H_0 . There is not enough evidence that there is a correlation between age and salary.
- 7 **a** 0.937 (3 s.f.)
b $H_0: \rho = 0$, $H_1: \rho \neq 0$, critical value $= \pm 0.6319$. Reject H_0 . There is evidence that there is a correlation between the age of a machine and its maintenance costs.
- 8 $H_0: \rho = 0$, $H_1: \rho < 0$, critical value $= -0.5822$. Reject H_0 . There is evidence that the greater the altitude, the lower the temperature.
- 9 $H_0: \rho = 0$, $H_1: \rho > 0$, critical value 0.5822, $0.972 > 0.5822$. Reject H_0 . There is evidence that age and weight are positively correlated.
- 10 **a** 0.940
b $H_0: \rho = 0$, $H_1: \rho > 0$, critical value 0.7293. Reject H_0 . There is evidence that sunshine hours and ice cream sales are positively correlated.
- 11 $r = 0.843$ (3 s.f.), $H_0: \rho = 0$, $H_1: \rho > 0$, critical value 0.8054. Reject H_0 . There is evidence that mean windspeed and daily maximum gust are positively correlated.
- 12 $r = -0.793$ (3 s.f.), $H_0: \rho = 0$, $H_1: \rho < 0$, critical value -0.7545 . Reject H_0 . There is evidence that temperature and pressure are negatively correlated.

Large data set

Student's own answers