



- 1 Ruby records the mean daily screen time, s hours, and the mean nightly sleep, h hours, for a random sample of 12 Year 12 students. The product moment correlation coefficient between s and h is -0.614
- State what is measured by the product moment correlation coefficient.
 - Test, at the 5% level of significance, whether there is evidence of negative correlation between screen time and sleep. State your hypotheses clearly and the critical value used. (4 marks)

a)

Strength of linear correlation

b) $\rho = \text{PMCC of the population}$

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Table: $n = 12$, 5%, one tail

$$\text{critical value} = -0.4973$$

Compare r with the critical value

$$-0.614 < -0.4973$$

In critical region, so reject H_0

Evidence of negative correlation between screen time and sleep

2

A head of sixth form believes there is a relationship between the number of hours a week, w , that Year 13 students spend in part-time work and their mock exam score, m . She takes a random sample of 20 Year 13 students and finds that the product moment correlation coefficient between w and m is -0.402

Test, at the 5% level of significance, whether there is evidence of correlation between w and m . State your hypotheses clearly and the critical value used.

(4 marks)

"a relationship": no direction, so two tails

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

5% two tails: 2.5% in each tail

Table: $n = 20, 0.025$

$$\text{critical values} = \pm 0.4438$$

Compare

$$-0.402 > -0.4438$$

Not in critical region, so do not reject H_0

Insufficient evidence of correlation between hours of work and mock score

3

A microbiologist grows a culture of bacteria. She counts the number of bacteria, N , at time t hours, and models the data by $N = ab^t$, where a and b are constants.

She codes the data using $y = \log_{10} N$ and finds the regression line of y on t to be

$$y = 2.15 + 0.087t$$

- Find the value of a and the value of b , giving your answers to 3 significant figures.
- Interpret the value of b in this context.

(4 marks)

a) Take $\log_{10}$ of both sides

$$N = ab^t$$

$$\log N = \log a + \log b^t$$

$$\log N = \log a + t \log b$$

Compare with $y = 2.15 + 0.087t$

$$\log a = 2.15$$

$$a = 10^{2.15}$$

$$a = 141.25...$$

$$\log b = 0.087$$

$$b = 10^{0.087}$$

$$b = 1.2218...$$

$$a = 141, b = 1.22$$

b) b is the hourly growth factor

The number of bacteria increases by about 22% each hour

4

Laila uses records for a random sample of 40 days at a weather station in Wales. She finds that the product moment correlation coefficient between daily mean wind speed and daily mean temperature is -0.284

She carries out a test for negative correlation, and the p -value for her test is 0.0379

- Write down suitable hypotheses for Laila's test.
- Using a 5% level of significance, state the conclusion of her test in context.
- State, giving a reason, whether the conclusion would change if a 1% level of significance were used.
- Laila says "This proves that strong winds cause lower temperatures." Comment on her statement. (5 marks)

a) $\rho = \text{population PMCC}$

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

b) Compare the p -value with 5%

$$0.0379 < 0.05$$

Significant, so reject H_0

Evidence of negative correlation between wind speed and temperature

c) Compare with 1%

$$0.0379 > 0.01$$

Yes: not significant at 1%, so do not reject H_0

d) Correlation is not causation

$r = -0.284$ is only weak correlation

Correlation does not prove cause; another factor may affect both

5

A sprint coach measures the reaction time, x milliseconds, and the 60 m sprint time, y seconds, for a random sample of 9 athletes.

x	182	195	201	210	214	226	233	241	250
y	7.42	7.61	7.40	7.66	7.52	7.58	7.82	7.55	7.80

- Use your calculator to find the product moment correlation coefficient between x and y .
- Test, at the 5% level of significance, whether there is evidence of positive correlation between reaction time and sprint time. State your hypotheses and the critical value used.
- State, giving a reason, whether your conclusion would be different if the coach had tested for any correlation at the 5% level.

(6 marks)

a) Calculator: paired data, 9 values

$$r = 0.6747$$

$$r = 0.675$$

b) ρ = population PMCC

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

Table: $n = 9$, 5%, one tail

$$\text{critical value} = 0.5822$$

$$0.6747 > 0.5822$$

Significant, so reject H_0

Evidence of positive correlation between reaction time and sprint time

c) Two tails: 2.5% in each tail

Table: $n = 9$, 0.025

$$\text{critical value} = 0.6664$$

$$0.6747 > 0.6664$$

No change: still reject H_0

6

Ayo is studying the mass, m kg, and the resting heart rate, h beats per minute, of 14 species of mammal. He codes the data using

$$x = \log_{10} m \quad y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.962

- a. Test, at the 1% level of significance, whether there is evidence of negative correlation between x and y . State your hypotheses clearly and the critical value used.

The equation of the regression line of y on x is $y = 2.36 - 0.24x$

- b. Find a model for h in terms of m in the form $h = am^k$, where a and k are constants to be found.
c. Use your model to estimate the resting heart rate of a mammal of mass 70 kg.

(7 marks)

a) $\rho = \text{population PMCC}$

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Table: $n = 14$, 1%, one tail

$$\text{critical value} = -0.6120$$

$$-0.962 < -0.6120$$

Significant, so reject H_0

Evidence of negative correlation between x and y

b) Replace x and y by the logs

$$y = 2.36 - 0.24x$$

$$\log h = 2.36 - 0.24 \log m$$

$$h = 10^{2.36 - 0.24 \log m}$$

$$h = 10^{2.36} \times 10^{-0.24 \log m}$$

Use $10^{k \log m} = m^k$

$$h = 10^{2.36} \times m^{-0.24}$$

$$10^{2.36} = 229.09...$$

$$h = 229m^{-0.24}$$

c) Substitute $m = 70$

$$h = 229.09 \times 70^{-0.24}$$

$$h = 82.6$$

About 83 beats per minute

7

A start-up company launches a new revision app. The number of downloads, D , in week w was recorded for the first 10 weeks. The company models D by

$$D = kc^w$$

where k and c are constants. The regression line of $\log_{10} D$ on w is

$$\log_{10} D = 1.82 + 0.135w$$

- Find the values of k and c , giving your answers to 3 significant figures.
- Interpret the value of c in context.
- Use the regression line to find the first week in which the model predicts more than 2000 downloads.
- Comment on the reliability of your answer to part (c).

(7 marks)

a) Compare with $\log D = \log k + w \log c$

$$\log k = 1.82$$

$$k = 10^{1.82}$$

$$k = 66.069...$$

$$\log c = 0.135$$

$$c = 10^{0.135}$$

$$c = 1.3645...$$

$$k = 66.1, c = 1.36$$

b) c is the weekly growth factor

Downloads increase by about 36% each week

c) Need $D > 2000$

$$\log D > \log 2000$$

$$1.82 + 0.135w > 3.3010$$

$$0.135w > 1.4810$$

$$w > 10.97$$

Week 11

d) Data only covers weeks 1 to 10

Week 11 is outside the data: extrapolation

Only small extrapolation, but growth may not stay exponential, so treat with caution

Priya records the altitude, a metres, and the mean July temperature, t °C, at a random sample of 10 weather stations in Scotland. The altitudes range from 20 m to 950 m. The product moment correlation coefficient between a and t is -0.931

- Test, at the 1% level of significance, whether there is evidence that altitude and temperature are negatively correlated. State your hypotheses clearly and the critical value used.
- State, with a reason, which variable is the explanatory variable.

The regression line of t on a is $t = 15.8 - 0.0062a$

- Interpret the gradient of the regression line in context.
- Estimate the mean July temperature at a weather station at an altitude of 600 m.
- Explain why it would not be sensible to use the line to estimate the mean July temperature at the summit of Ben Nevis, 1345 m above sea level.

(7 marks)

a) ρ = population PMCC

$$H_0: \rho = 0$$

$$H_1: \rho < 0$$

Table: $n = 10$, 1%, one tail

$$\text{critical value} = -0.7155$$

$$-0.931 < -0.7155$$

Significant, so reject H_0

Evidence that altitude and temperature are negatively correlated

b) Which one affects the other?

Altitude a : it can affect temperature

c) Gradient = change in t per metre

Temperature falls by 0.0062 °C for each extra metre of altitude

d) Substitute $a = 600$

$$t = 15.8 - 0.0062 \times 600$$

$$t = 15.8 - 3.72$$

$$t = 12.08$$

$$12.1 \text{ °C}$$

e) Compare with the data range

$$1345 \text{ m} > 950 \text{ m}$$

Outside the range of the data: extrapolation

A researcher is planning a study of the link between the number of steps walked per day and resting blood pressure. A pilot study suggests that the sample correlation coefficient will be about $r = 0.42$

- a. Using the table of critical values, find the smallest sample size, n , for which $r = 0.42$ would give significant evidence of positive correlation at the 5% level.

In the full study, she takes a random sample of 25 adults and obtains $r = 0.42$

- b. Test, at the 5% level of significance, whether there is evidence of correlation between the two variables. State your hypotheses clearly and the critical value used.
c. Explain why a significant result does not mean that the correlation is strong.

(6 marks)

a) 5% one tail: need $r >$ critical value

Look down the 0.05 column

$n = 16$: critical value = 0.4259

$0.42 < 0.4259$: not significant

$n = 17$: critical value = 0.4124

$0.42 > 0.4124$: significant

$n = 17$

b) "correlation": two tails

$H_0: \rho = 0$

$H_1: \rho \neq 0$

Table: $n = 25$, 0.025

critical values = ± 0.3961

$0.42 > 0.3961$

Significant, so reject H_0

Evidence of correlation between steps and blood pressure

c) Significant does not mean strong

The test only shows ρ is unlikely to be 0

Larger samples make weak r significant; $r = 0.42$ is fairly weak

A chemist measures the concentration, C mg per litre, of a substance t minutes after a reaction starts.

t	0	2	4	6	8	10	12
C	81.2	57.6	43.1	30.4	22.6	15.9	11.7

a. Use your calculator to find the product moment correlation coefficient between

(i) t and C

(ii) t and y , where $y = \log_{10} C$

b. Explain why your answers to part (a) suggest that $C = ab^t$ is a suitable model.

c. Find the equation of the regression line of y on t , giving the constants to 4 decimal places.

d. Hence find the values of a and b to 3 significant figures.

e. Using your model, find the time taken for the concentration to halve.

(8 marks)

a) Calculator, paired lists

$$(i) r = -0.9641$$

(ii) Make a list of $y = \log_{10} C$ first

$$y: 1.9096, 1.7604, 1.6345, 1.4829, 1.3541, 1.2014, 1.0682$$

$$(ii) r = -0.9998$$

$$(i) -0.964 \quad (ii) -1.00 \quad (-0.9998)$$

b) Which is closer to a straight line?

$$|-0.9998| > |-0.9641|: \log C \text{ is almost linear in } t$$

$$C = ab^t$$

$$\log C = \log a + t \log b$$

$$\log C \text{ linear in } t \text{ fits } C = ab^t$$

c) Calculator: regression of y on t

$$y = 1.9076 - 0.0700t$$

d) Compare with $\log C = \log a + t \log b$

$$\log a = 1.9076$$

$$a = 10^{1.9076} = 80.83...$$

$$\log b = -0.07005$$

$$b = 10^{-0.07005} = 0.8510...$$

$$a = 80.8, b = 0.851$$

e) Halved: $C = 0.5a$

$$ab^t = 0.5a$$

$$b^t = 0.5$$

$$t \log b = \log 0.5$$

$$t = \frac{\log 0.5}{\log b}$$

$$t = \frac{-0.30103}{-0.07005}$$

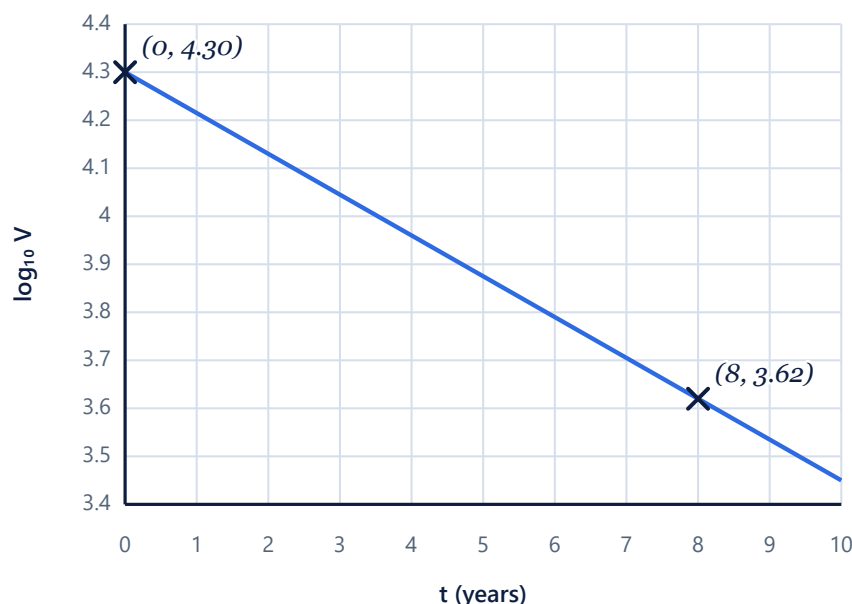
$$t = 4.297...$$

4.30 minutes

11

The value, £ V , of a model of car is modelled by $V = ab^t$, where t is the age of the car in years and a and b are constants. The graph shows the straight line of $\log_{10} V$ against t . The line passes through the points $(0, 4.30)$ and $(8, 3.62)$.

- Find the value of a and the value of b , giving your answers to 3 significant figures.
- Interpret the value of b in context.
- Find, according to the model, the age at which the car's value first falls below £5000
- Give one reason why the model may not be suitable for very old cars.



(7 marks)

a) Straight Line: $\log V = \log a + t \log b$

Intercept = $\log a$

$$\log a = 4.30$$

$$a = 10^{4.30} = 19\,952.6\dots$$

Gradient = $\log b$

$$\text{gradient} = \frac{3.62 - 4.30}{8 - 0}$$

$$\text{gradient} = -0.085$$

$$b = 10^{-0.085} = 0.8222\dots$$

$$a = 20\,000 \text{ (19\,953)}, b = 0.822$$

$$b) 1 - 0.822 = 0.178$$

The car loses about 17.8% of its value each year

c) Need $V < 5000$

$$\log V = 4.30 - 0.085t$$

$$4.30 - 0.085t < \log 5000$$

$$4.30 - 0.085t < 3.69897$$

$$0.085t > 0.60103$$

$$t > 7.0709\dots$$

7.07 years

d) What happens as t gets large?

Model: V gets closer and closer to 0

An old car keeps a scrap value, or classic cars may gain value

Ellie records the number of pupils, P , at a school of 600 pupils who have caught a winter virus by the end of day d of an outbreak.

d	1	2	3	4	5	6	7	8
P	6	9	13	21	30	47	68	104

Ellie suggests the model $P = ab^d$, where a and b are constants.

- Show that this model gives a straight line when $\log_{10} P$ is plotted against d .
- Find the product moment correlation coefficient between d and $\log_{10} P$, and test, at the 5% level of significance, whether there is evidence of positive correlation between them. State your hypotheses and the critical value used.
- Find the regression line of $\log_{10} P$ on d and hence find the values of a and b , to 3 significant figures.
- Use the model to find the first day on which more than 500 pupils would have caught the virus, and comment on this prediction.

(10 marks)

a) Take $\log_{10}$ of both sides

$$P = ab^d$$

$$\log P = \log a + \log b^d$$

$$\log P = \log a + d \log b$$

Form $y = c + mx$ with $y = \log P$, $x = d$

Straight line: gradient $\log b$, intercept $\log a$

b) List of $\log_{10} P$ first

0.7782, 0.9542, 1.1139, 1.3222, 1.4771, 1.6721, 1.8325, 2.0170

$$r = 0.9997$$

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

Table: $n = 8$, 5%, one tail

$$\text{critical value} = 0.6215$$

$$0.9997 > 0.6215$$

Significant, so reject H_0

Evidence of positive correlation between d and $\log P$

c) Calculator: regression of $\log P$ on d

$$\log P = 0.5981 + 0.1773d$$

$$\log a = 0.5981$$

$$a = 10^{0.5981} = 3.963...$$

$$\log b = 0.1773$$

$$b = 10^{0.1773} = 1.504\dots$$

$$a = 3.96, b = 1.50$$

d) Need $P > 500$

$$0.5981 + 0.1773d > \log 500$$

$$0.5981 + 0.1773d > 2.69897$$

$$0.1773d > 2.10087$$

$$d > 11.85$$

Day 12

Comment

Day 12 is outside the data (days 1 to 8)

Only 600 pupils, so growth must slow down

Unreliable: extrapolation, and the model grows too fast