# Hypothesis testing (correlation)

1

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)**

2

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)**

3

A machine hire company kept records of the age, $X$ months, and the maintenance costs, £$Y$, of one type of machine. The 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)**

4 A small company decided to import fine Chinese porcelain. They believed that in the long term this would prove to be an increasingly profitable arrangement with profits increasing proportionally to sales. Over the next 6 years their sales and profits were as shown in the table below.

| **Year** | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 |
|---|---|---|---|---|---|---|
| **Sale in thousands** | 165 | 165 | 170 | 178 | 178 | 175 |
| **Profits in £1000** | 65 | 72 | 75 | 76 | 80 | 83 |

Using a 1% significance level, test to see if there is any evidence that the company's beliefs were correct, and that profit and sales were positively correlated.

5

A safari ranger believes that there is a positive correlation between the amount of grass per square kilometre and the number of meerkats that graze there. He decides to carry out a hypothesis test to see if there is evidence for his claim. He takes a random sample of 10 equal-sized areas of grassland, records the amount of grass and the number of meerkats grazing in each, and finds that the correlation coefficient is 0.66.

Given that this result provided the ranger with sufficient evidence to reject his null hypothesis, suggest the least possible significance level for the ranger's test.

6

Data on the daily mean temperature and the daily total sunshine is taken from the large data set for Leuchars in May and June 1987. A meteorologist finds that the product moment correlation coefficient for these data is 0.715. Given that the researcher tests for positive correlation at the 2.5% level of significance, and concludes that the value is significant, find the smallest possible sample size.

7

A scientist wishes to test, at the 5% level, whether there is any correlation between the masses of two reactants in an experiment. She conducts the experiment 20 times and observes a product moment correlation coefficient of $r = 0.4$.

The probability of obtaining this value of $r$ or higher, given $\rho = 0$, is 0.0403. The scientist claims that this means her result is significant at the 5% level.

**a** Explain why the scientist is incorrect.

**b** Find the critical values of $r$ for this test at the 5% level.

# Answers to hypothesis testing

1

$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho < 0 \end{array}\right\}$ 1-tail $\alpha = 0.05$

Test statistic $= -0.975$

$n = 9$, critical value $= -0.5822$

Lower tail test, t.s. < c.v. since $-0.975 < -0.5822$ reject $H_0$.

Conclude there is evidence of negative correlation. There is evidence that the greater the height above sea level, the lower the temperature at 7.00 a.m. is likely to be.

2

$r = 0.843$ (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.

3

**a** $r = 0.937$ (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.

4

$r = 0.793$

(NB. In the exam get this directly from your calculator. If you set up a table of results you are likely to run out of time.)

$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\}$ 1-tail $\alpha = 0.01$

test statistic $= 0.793$

critical values $= 0.8822$

t.s. < c.v. so accept $H_0$.

Conclude there is insufficient evidence at the 1% significance level to support the company's belief.

5

The safari ranger's test.
Type: 1-tailed

$H_0 : \rho = 0$

$H_1 : \rho > 0$

Sample size: 10
$r = 0.66$
He has sufficient evidence to reject $H_0$. The corresponding part of the table reads:

| 0.10 | 0.05 | 0.025 | 0.01 | 0.005 | Sample size |
|---|---|---|---|---|---|
| 0.4428 | 0.5494 | 0.6319 | 0.7155 | 0.7646 | 10 |

Therefore the least possible significance level for the ranger's test is 2.5%.

6

The information from the question is as follows:
Type: 1-tailed

$H_0 : \rho = 0$

$H_1 : \rho > 0$

Sample size: unknown
$r = 0.715$.
He has sufficient evidence to reject $H_0$. Part of the corresponding column of the table reads:

| PMCC at 0.025 level of significance | Sample size |
|---|---|
| 0.9500 | 4 |
| 0.8783 | 5 |
| 0.8114 | 6 |
| 0.7545 | 7 |
| 0.7067 | 8 |

Therefore the smallest possible sample size is 8.

7

**a** This is a two-tailed test, so the scientist would need to halve the significance level or double her *p*-value.
**b** ±0.4438