Summary of key points

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:

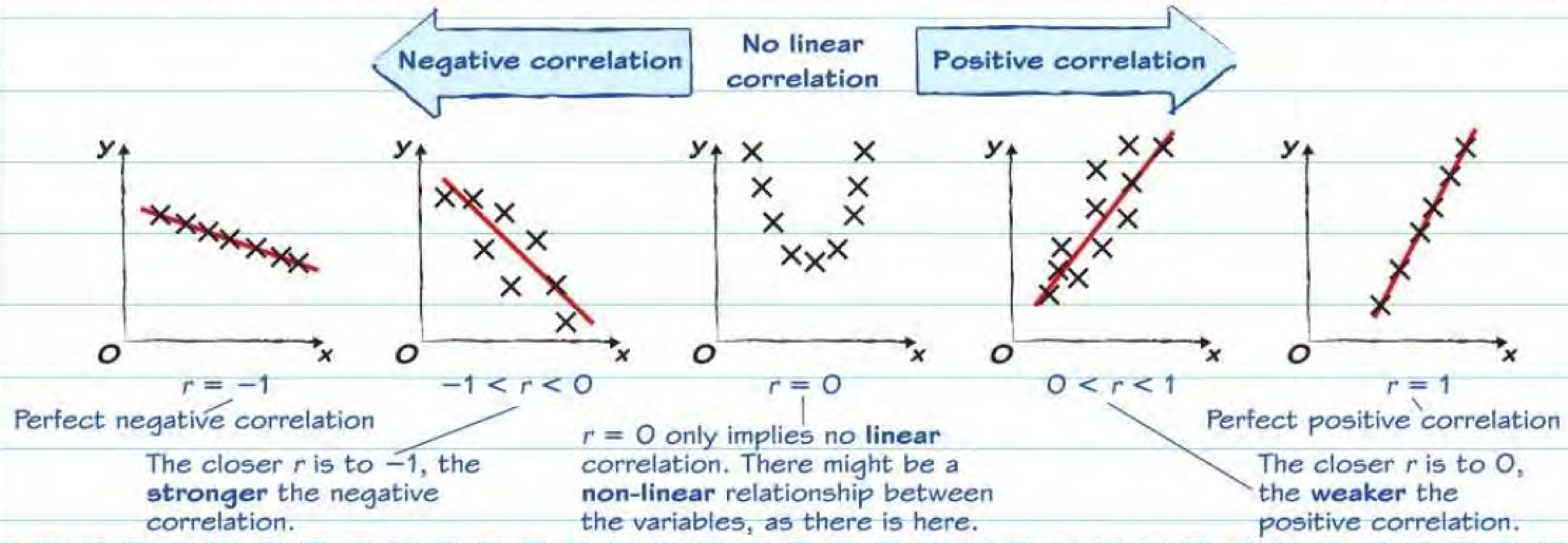
- $H_0: \rho = 0$, $H_1: \rho > 0$ or
- $H_0: \rho = 0$, $H_1: \rho < 0$

For a two-tailed test use:

- $H_0: \rho = 0$, $H_1: \rho \neq 0$

Measuring correlation

You can use the **product moment correlation coefficient (PMCC)** to measure the strength of the linear correlation between two variables. You need to be able to use your calculator to find the PMCC.



Worked example

An engineer is using a wind tunnel to investigate the effects of air speed on drag over an aircraft wing. She records her results in a table.

Air speed, a (mph)	20	40	80	120	160
Drag, d (N)	80	260	1670	5400	12700

The engineer believes the data can be modelled by a relationship of the form $d = Pa^k$ for constants P and k. She codes the data using $x = \log a$ and $y = \log d$.

(a) Calculate the product moment correlation coefficient of the coded data. **(3 marks)**

x	1.301	1.602	1.903	2.079	2.204
y	1.903	2.415	3.223	3.732	4.104

$r = 0.994$

Use your calculator to find the product moment correlation coefficient. You might have to select the type of model you are using. Because the PMCC measures **linear** correlation, you should choose a model of the form $y = a + bx$. The letter r is usually used to denote the PMCC.

The equation of the regression line of y on x is found to be $y = 2.470x - 1.415$

(b) Estimate the values of P and k in the engineer's model, and comment on the validity of this model. **(4 marks)**

$\log d = 2.470 \log a - 1.415$

$\log d = \log a^{2.470} - \log 10^{1.415}$

$\log d = \log \dfrac{a^{2.470}}{10^{1.145}}$

$d = 10^{-1.415} a^{2.470}$

So $P = 10^{-1.415} = 0.0385$ and $k = 2.470$

The PMCC is close to 1 ($r = 0.994$), which suggests this model is valid.

Non-linear models

You can use **logarithms** to test for non-linear relationships between two variables x and y. You need to know about two different forms:

1 If $y = ax^n$ for constants a and n then $\log y = \log a + n \log x$

You need to use the coding $Y = \log y$ and $X = \log x$ to obtain a linear relationship.

2 If $y = kb^x$ for constants k and b then $\log y = \log k + x \log b$

You need to use the coding $Y = \log y$ and $X = x$ to obtain a linear relationship.

You can revise the techniques needed to determine these relationships on page 53.

Now try this

The daily maximum relative humidity (%) and the daily mean visibility (km) recorded in Leeming for the first 8 days in May 2015 were recorded in the large data set.

Humidity (%)	99	94	95	94	99	87	88
Visibility (km)	25	16	14	20	15	24	32

Calculate the product moment correlation coefficient of the data, and describe the nature of the linear relationship (if any).

Answer **(2 marks)**

Had a look ☐ Nearly there ☐ Nailed it! ☐

Hypothesis testing for 0 correlation

You might need to determine whether a calculated value of the product moment correlation coefficient is statistically significant. You can do this by carrying out a **hypothesis test** for zero correlation.

Look at page 132 for a reminder about hypothesis testing.

Using tables

You need to use the table of **Critical Values for Correlation Coefficients** given in the formulae booklet. Choose the correct significance level and sample size to determine the critical value.

Product Moment Coefficient					Sample size, n
Level					
0.10	0.05	0.025	0.01	0.005	
0.8000	0.9000	0.9500	0.9800	0.9900	4
0.6870	0.8054	0.8783	0.9343	0.9587	5
0.6084	0.7293	0.8114	0.8822	0.9172	6
0.5509	0.6694	0.7545	0.8329	0.8745	7
0.5067	0.6215	0.7067	0.7887	0.8343	8

Worked example

A doctor recorded the head circumference, y cm, and gestation period, x weeks, for a random sample of six new-born babies at a hospital.

x	36	40	33	37	40	39
y	30.0	35.0	29.8	32.5	33.2	32.1

(a) Calculate the product moment correlation coefficient for these data. **(2 marks)**

$r = 0.8621$

The doctor believes that there is a positive correlation between head circumference and gestation period.

(b) Test the doctor's claim at the 1% significance level, stating your hypotheses clearly. **(3 marks)**

H_0: $\rho = 0$, H_1: $\rho > 0$
Sample size = 6
Significance level = 0.01
From the table, the critical value of r is 0.8822 and the critical region is $r > 0.8822$.
$0.8621 < 0.8822$, so the observed value of r is not in the critical region.
There is not enough evidence, at the 1% level, to reject H_0, so the data does not support the doctor's claim.

Sample vs population

You calculate a value of the PMCC for a **sample** and use this to make a claim about the whole **population**:

- ✓ Use r to denote the PMCC for the sample.
- ✓ Use ρ (rho) to denote the PMCC for the whole population.

Work out the sample PMCC, r, by entering the values into your calculator.

The doctor believes there is positive correlation, so you need to use a one-tailed test. Use the table to determine the critical value for a sample of size 6 and a significance level of 1% (or 0.01). If you assume H_0 to be true, then $P(r > 0.8822) = 0.01$. This means that the probability of incorrectly rejecting the null hypothesis is 0.01.

In part (c), you are testing for any evidence of linear correlation (positive or negative) so this is a two-tailed test. You need to halve the significance level to determine the probability in each tail.

Now try this

(a) State what is measured by the product moment correlation coefficient. Answer **(1 mark)**

In a chemistry experiment, the temperature, x °C, and the mass of the product, y mg, were recorded.

x (°C)	100	110	120	130	140	150	160	170
y (mg)	34	48	47	44	42	39	21	7

(b) Calculate the product moment correlation coefficient for these data. Answer **(2 marks)**

(c) Test, at the 5% significance level, whether these results show any evidence of a linear relationship between the temperature and the mass of the product. Answer **(3 marks)**