

2A The Product Moment Correlation Coefficient

The product moment correlation coefficient (PMCC) measures:

The PMCC can take values between:

$r = 1$ means

$r = -1$ means

The PMCC was designed for data that is:

The product moment correlation coefficient, r , is given by:

$r =$

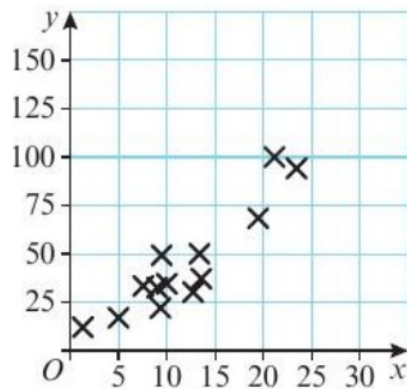
Effect of linear coding on the PMCC, for example

$$p = ax + b \quad \text{or} \quad p = \frac{x - a}{b}$$

Why:

1. The number of vehicles, x millions, and the number of accidents, y thousands, in 15 different countries were recorded. The following summary statistics were calculated and a scatter graph of the data is given below:

$$\Sigma x = 176.9, \quad \Sigma y = 679, \quad \Sigma x^2 = 2576.47, \quad \Sigma y^2 = 39\,771, \quad \Sigma xy = 9915.3$$



- a) Calculate the product moment correlation coefficient between x and y .
- b) With reference to your answer to part a and the scatter graph, comment on the suitability of a linear regression model for these data.

2. Data are collected on the amount of dietary supplement, d grams, given to a sample of 8 cows and their milk yield, m litres. The data were coded using

$$x = \frac{d}{2} - 6 \quad \text{and} \quad y = \frac{m}{20}$$

The following summary statistics were obtained:

$$\Sigma d^2 = 4592, \quad S_{dm} = 90.6, \quad \Sigma x = 44, \quad S_{yy} = 0.05915$$

- a) Use the formula for S_{yy} to show that $S_{mm} = 23.66$.

- b) Find the value of the product moment correlation coefficient between d and m .

2B Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient, r_s , is:

To rank two sets of data:

You might use Spearman's rank correlation coefficient instead of the PMCC when:

- -
 -
1. Two tea tasters were asked to rank nine blends of tea in their order of preference. The tea they liked best was ranked 1. Their orders of preference are shown in the table:

Blend	A	B	C	D	E	F	G	H	I
Taster 1 (x)	3	6	2	8	5	9	7	1	4
Taster 2 (y)	5	6	4	2	7	8	9	1	3

Calculate Spearman's rank correlation coefficient for these data.

x^2									
y^2									
xy									

A Spearman's rank correlation coefficient of:

+1 means

-1 means

0 means

If there are no tied ranks, Spearman's rank correlation coefficient is calculated using:

$r_s =$ where d is and n is

2. During a cattle show, two judges ranked ten cattle for quality according to the following table.

Cattle	A	B	C	D	E	F	G	H	I	J
Judge A	1	5	2	6	4	8	3	7	10	9
Judge B	3	6	2	7	5	8	1	4	9	10
d										
d ²										

Find Spearman's rank correlation coefficient between the two judges and comment on the result.

Tied ranks: equal data values are given a rank equal to:

Data value	200	350	350	400	700	800	800	800	1200
Rank									

With tied ranks, the formula for r_s :

3. The marks of eight pupils in French and German tests were as follows:

	A	B	C	D	E	F	G	H
French, f %	52	25	86	33	55	55	54	46
German, g %	40	48	65	57	40	39	63	34
Rank, f								
Rank, g								
d								
d²								

- a) Use the formula below to find an estimate for Spearman's rank correlation coefficient, showing clearly how you deal with tied ranks. Give your answer to 2 decimal places.

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

- b) Without recalculating the correlation coefficient, state how your answer to part a would change if:

i) pupil H's mark for German was changed to 38%

ii) a ninth pupil was included who scored 95% in French and 89% in German.

The teacher collects extra data from other students in the class and finds that there are now many tied ranks.

- c) Describe how she would now find a measure of the correlation.

2C Hypothesis Testing for Zero Correlation

A hypothesis test for zero correlation is used to:

Notation: r and r_s are used for

ρ and ρ_s are used for

For a one-tailed test, use either:

$H_0:$ $H_1:$ or $H_0:$ $H_1:$

For a two-tailed test, use:

$H_0:$ $H_1:$

Finding the critical region:

Watch out:

1. A chemist observed 20 reactions, and recorded the mass of the reactant, x grams, and the duration of a reaction, y minutes. She summarised her findings as follows:

$$\Sigma x = 20, \quad \Sigma y = 35, \quad \Sigma xy = 65, \quad \Sigma x^2 = 35, \quad \Sigma y^2 = 130$$

Test, at the 5% significance level, whether these results show evidence of any correlation between the mass of the reactant and the duration of the reaction.

2. The popularity of 16 subjects at a comprehensive school was found by counting the number of boys and the number of girls who chose each subject and then ranking the subjects. The results are shown in the table below.

Subject	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Boys' ranks, b	2	5	9	8	1	3	15	16	6	10	12	14	4	7	11	13
Girls' ranks, g	4	7	11	3	6	9	12	16	5	13	10	8	2	1	15	14
d																
d^2																

- a) Calculate Spearman's rank correlation coefficient.
- b) Using a suitable test, at the 1% level of significance, test the assertion that boys' and girls' choices are positively correlated.