

Pairwise Correlation Coefficients Exercise

Consider 4 characteristics which can be measured for each item in a sample of 6. Determine the pairwise correlation coefficient matrix for the 4 characteristics and comment on the values.

	A	B	C	D
Item 1	6	1	5	2
Item 2	3	2	4	2
Item 3	5	3	4	3
Item 4	1	4	3	4
Item 5	4	5	3	5
Item 6	2	6	2	5

We are going to need the mean and standard deviation of each characteristic in order to calculate the covariances and correlations so we do that first.

	A	B	C	D
Mean of the Characteristic (μ)	3.5	3.5	3.5	3.5
Variance of the Characteristic (σ^2)	2.916	2.916	0.917	1.667
Standard Deviation of the Characteristic (σ)	1.708	1.708	0.958	1.291

Note that the values for characteristics A and B are the same but they are distributed differently between the 6 items. These characteristics will therefore have the same mean and standard deviation.

We can now calculate the Variance-Covariance matrix for the characteristics using -

$$Cov(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

Variance-Covariance Matrix

	A	B	C	D
A	2.916	-1.75	1.25	-1.25
B	-1.75	2.916	-1.583	2.083
C	1.25	-1.583	0.917	-1.083
D	-1.25	2.083	-1.083	1.667

Note that the main diagonal gives the variance for each characteristic.

The signs (+ and -) of the covariances give us an indication of whether characteristics are positively or negatively related but this matrix isn't very helpful in determining the strength of the relationships between the characteristics because the values are not normalised.

We can now calculate the Pairwise Correlation Coefficient matrix for the characteristics using -

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

Pairwise Correlation Coefficient Matrix

	A	B	C	D
A	1	-0.6	0.764	-0.57
B	-0.6	1	-0.967	0.945
C	0.764	-0.967	1	-0.876
D	-0.57	0.945	-0.876	1

Note that the main diagonal is all 1s indicating that each characteristic is maximally correlated with itself.

Comments on the results

We can see that characteristic B is very negatively correlated (-0.967) with characteristic C. I.e. items which have a high value for B consistently have a low value for C and vice versa.

We can also see that B and D tend to vary together (0.945). I.e. items with a high value for B consistently have a high value for D and vice versa.

C is also negatively correlated with D (-0.876) but not as strongly as with B (-0.967) and positively correlated with A (0.764) but this is not a compelling value.

A is even less strongly correlated with B (-0.6) and D (-0.57).

Overall, B and D appear to be related and C is complementary to them.