Click on 'Bookmarks' in the Left-Hand menu and then click on the required section.

12.7 Spearman's rank correlation (Web)

12.7.1 Introduction

Spearman's Rank correlation coefficient, r_S (sometimes written as ρ, 'rho'), is the non-parametric equivalent of Pearson's product moment coefficient, r, (13.1.2), and is used as a measure of association between two variables.

When using the *parametric* correlation coefficient, r, it is required that the data values for the two variables are derived from populations with *normal* distributions. There is no such requirement for the Spearman's Rank correlation, and it can be used with any data whose values can be ranked in order.

12.7.2 Spearman's Rank Correlation Coefficient

The first step is to rank the x-values and the y-values **separately**. The correlation coefficient, r_S, can then be calculated in one of two ways:

1. The value of r_S equals the value of the *parametric* correlation coefficient, r, between the *rank values* of the 'x' and 'y' data, and can be calculated in EXCEL using the functions CORREL or PEARSON.

2. Alternatively, the differences, d_i, in rank can be calculated between each pair of data values:

$$d_i = \text{Rank of } x_i - \text{Rank of } y_i, \qquad [12.11]$$

and then using the formula:

$$r_S = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \qquad [12.12]$$

where n is the number of data pairs.

The two above methods for calculating, r_S, show slight differences if there are values of equal rank (ties) in either of the data sets.

A selected set of critical values, $r_{S(CRIT)}$, for the Spearman's correlation coefficient are given in Appendix V.

If $r_S \geq r_{S(CRIT)}$ we would accept that there is a significant correlation.

Example 12.15
Referring to the same data as in Example 13.2, use a *non-parametric* method to investigate whether the data shows a *1-tailed* correlation between performance in the number of 'press-ups' and 'sit-ups' achieved by 7 children (subjects).

Subject:	1	2	3	4	5	6	7
Press Ups:	10	8	2	6	7	3	5
Sit Ups:	24	27	12	14	21	16	22

See following text for calculations:

The ranks are calculated *separately* for the data values for 'press-ups' and 'sit-ups' in Example 12.15 - see Table 12.15

Subject:	1	2	3	4	5	6	7	Total
Number of 'press-ups':	**10**	**8**	**2**	**6**	**7**	**3**	**5**	
Ranks for 'press-ups':	7	6	1	4	5	2	3	
Number of 'sit-ups':	**24**	**27**	**12**	**14**	**21**	**16**	**22**	
Ranks for 'sit-ups':	6	7	1	2	4	3	5	
Difference in Ranks, d:	1	-1	0	2	1	-1	-2	
d^2:	1	1	0	4	1	1	4	12

Table 12.15 Ranking Data for Spearman's Correlation in Example 12.15

Using method 1 above, the correlation coefficient between the *ranks* for the two sets of data in EXCEL is found to be $r_S = 0.7857$.

Using method 2, taking the differences between each of the ranks, squaring the differences, and then taking the sum of the differences gives $\Sigma d^2 = 12$. Substituting this value, together with $n = 7$ into [12.12] gives $r_S = 0.7857$.

The critical 1-tailed value for $\alpha = 0.05$ from Appendix V is $r_{S(CRIT)} = 0.714$.

Since $r_S > r_{S(CRIT)}$ we accept the Proposed Hypothesis that the number of 'press-ups' shows a positive correlation with the number of 'sit-ups'.

Note that the results obtained for Example 12.15 using the non-parametric method are consistent with those obtained for Example 13.2 using the parametric method.

12.7.3 Using the *p*-value in EXCEL

The 2-tailed *p*-value can be calculated in EXCEL by first calculating the separate ranks (in 12.7.2) of the two data sets and then using:

Tools > Data Analysis > Regression

Example 12.16

Use EXCEL to derive the *p*-value for example Example 12.15.

Entering the rank values into two columns in EXCEL, and using

Tools > Data Analysis > Regression

for the data in the two columns gives p(2-tailed) = 0.036.
Calculating the 1-tailed value: p(1-tailed) = 0.036/2 = 0.018.
Since $p < 0.05$ we accept the Proposed Hypothesis (as in Example 12.15).

Q12.11 Ten trainees attended different lengths of training, and their individual progress was assessed on a 5-point scale, from 0 (poor) to 5 (excellent).
Using the data given below assess (at 0.05) whether their assessment score increased with the number of days training.

Trainee	1	2	3	4	5	6	7	8	9	10
Days	8	3	4	6	7	5	9	10	6	5
Assessment	2	1	4	1	4	2	5	3	5	4

i) Calculate Spearman's Rank correlation coefficient and compare with the critical value to investigate possible correlation

ii) Use the *p*-value method to investigate possible correlation.