5.2 Spearman Correlation

The Spearman rank correlation computes the product-moment correlation (Pearson *r*) on ordinal measurements or ranked scores. The Spearman rank correlation is often used in the place of the Pearson correlation when the sample size is small, or the distributions of scores are non-normal, or the data types are ordinal scales. The Spearman correlation measures consistency of the ranks, rather than the shape of the relationship between the variables being compared. Many data collection in social science and psychological research involving human subjects are based on preferences or ranks of observations rather than interval or ratio measurements. Most human subjects find it easier to rank categories of certain characteristics rather than place numerical values on these categories of variables. Associations of ranked observations or survey opinions from lowest to highest, with larger numbers being considered the higher rank, is the subject of the Spearman correlation analysis. Because the Spearman correlation coefficient involves ordinal rather than interval or ratio measurements, its formula (a modification of Pearson *r*'s formula), below, is simpler than that for the Pearson *r*.

$$r_s = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$$

where ΣD^2 is the sum of the squared difference between ranks,
and N is the number of pairs of ranks.

Table 5.2.1 shows a computation of the Spearman correlation coefficient, $\boldsymbol{r_s}$, commonly refer to as Spearman's rho (pronounced "roe"). The variables being compared are ordinal values.

Table 5.2.1 *Spearman Correlation: Two Variables with Ordinal Scales*

Subject	**X**	**Y**	**D**	D^2
1	6	4	2	4
2	1	1	0	0
3	3	2	1	1
4	10	9	1	1
5	7	6	1	1
6	2	5	-3	9
7	8	8	0	0
8	4	3	1	1
9	5	7	-2	4
10	9	10	-1	1
				$\sum D^2 = 22$

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(22)}{10(100-1)} = 1 - \frac{132}{990} = 1 - 0.1333 = 0.867$$

The SPSS procedure for computing the Spearman rank correlation coefficient is similar to computing the Pearson r; however, the Spearman choice for correlation is selected instead of Pearson's. (See Figure 5.2.1). The SPSS Spearman rho for the data in Table 5.2.1 is shown in Table 5.2.2.

Table 5.2.2 *SPSS Spearman Correlation Output 1*

			Y
Spearman's rho	X	Correlation Coefficient	0.867(**)
		Sig. (2-tailed)	0.001
		N	10

** Correlation is significant at the 0.01 level (2-tailed).

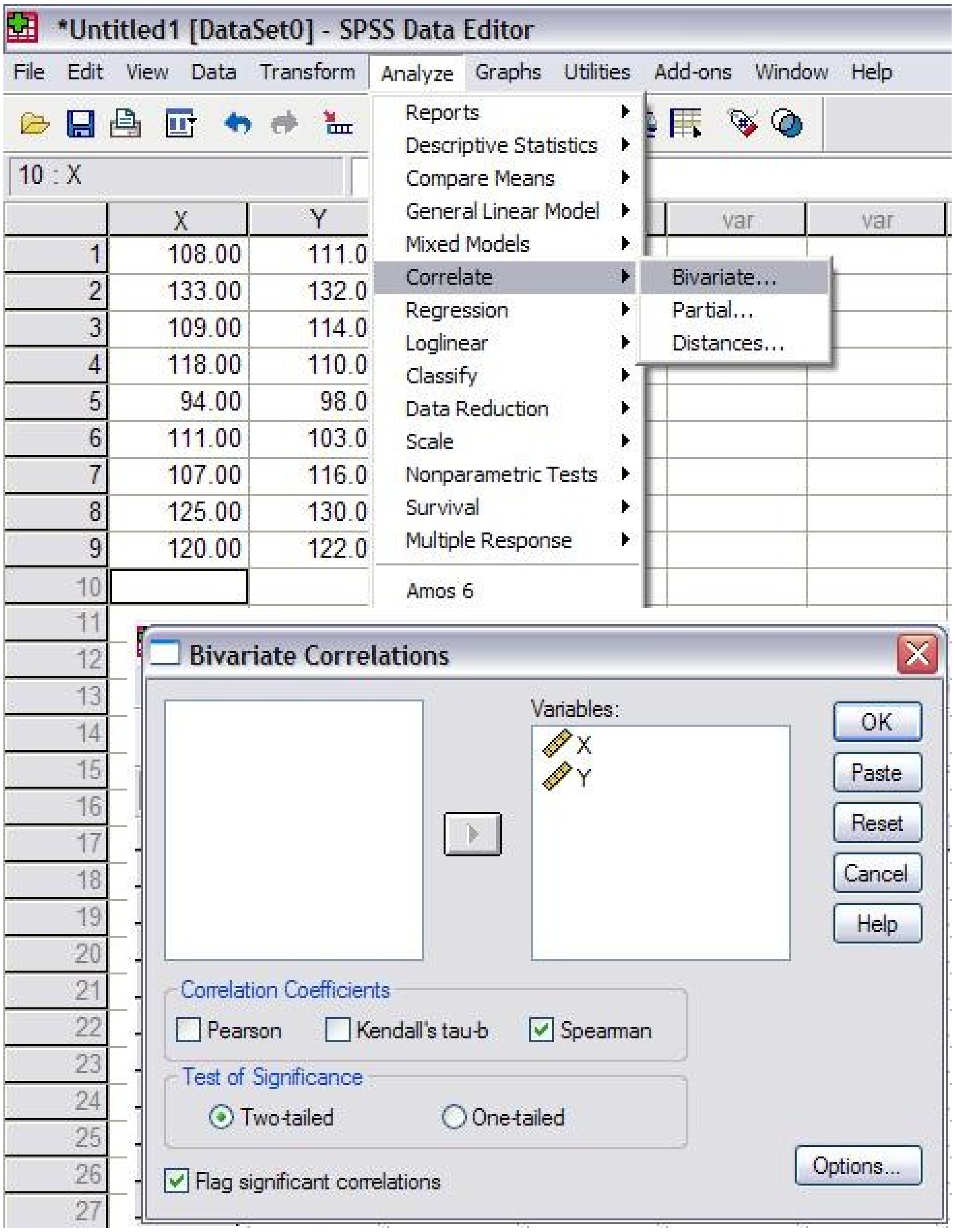

Figure 5.2. 1 SPSS Spearman rho Correlation Procedure

When Spearman correlation coefficient between interval or ratio variables, we first rank the data for each, the lowest score receives a rank of 1, the next higher a rank of 2, and so on. The two variables here are the first nine scores of the Verbal and Quant for the HWJ100 dataset. When the values are all different (***Untied***) for a variable, we calculate the Spearman rho on the ranked untied scores. Table 5.2.3 shows the SPSS Spearman rho output for data in Table 5.2.3. We may also rank the scores in reverse order (i.e., the highest receives a rank of 1). Ranking the scores in regular or reverse order (**reverse rank**ing), does not affect the value of the Spearman rho, but we must be consistent in our ranking procedure for each variable.

Table 5.2.3 *Spearman rho: Ranked Untied Interval Scores*

Subject	**X**	**Y**	**X Rank**	**Y Rank**	**D**	**D^2**
1	108.00	111.00	3	4	-1	1
2	33.00	132.00	9	9	0	0
3	09.00	114.00	4	5	-1	1
4	18.00	110.00	6	3	3	9
5	94.00	98.00	1	1	0	0
6	11.00	103.00	5	2	3	9
7	07.00	116.00	2	6	-4	16
8	25.00	130.00	8	8	0	0
9	20.00	122.00	7	7	0	0
						$\sum D^2 = \mathbf{36}$

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(36)}{9(81-1)} = 1 - \frac{216}{720} = 1 - 0.3 = 0.7$$

Table 5.2. 4 *SPSS Spearman rho Output 2*

			Y
Spearman's rho	X	Correlation Coefficient	0.700*
		Sig. (2-tailed)	0.036
		N	9

* Correlation is significant at the 0.05 level (2-tailed).

When ranking interval or ratio scales that are ***tied*** for computation of the Spearman correlation coefficient, we first rank the data for each, the lowest score receives a rank of 1, the next higher a rank of 2, and so on. When the values are the same for a variable (tied), we assign the average rank for the tied values; e.g. if subjects 4 and 6 tied for *X* with values = 1, so ranks 9 and 10 are averaged and a new rank of 9.5 is assigned.

Table 5.2. 5 *Spearman rho: Ranked Tied Interval Scores*

Subject	X	Y	X Rank	Y Rank	D	D^2
1	11	111.00	4.5	4	0.5	0.25
2	13	132.00	7	9	-2	4
3	10	114.00	2.5	5	-2.5	6.25
4	11	110.00	4.5	3	1.5	2.25
5	9	98.00	1	1	0	0
6	12	103.00	6	2	4	16
7	10	116.00	2.5	6	-3.5	12.25
8	14	130.00	8.5	8	0.5	0.25
9	14	122.00	8.5	7	1.5	2.25
						$\sum D^2 = \boldsymbol{43.5}$

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(43.5)}{9(81-1)} = 1 - \frac{261}{720} = 1 - 0.3625 = 0.638$$

Table 5.2. 6 *SPSS Spearman rho Output 3*

			Y
Spearman's rho	X	Correlation Coefficient	0.633
		Sig. (2-tailed)	0.067
		N	9

Coefficient of Determination

The ***coefficient of determination***, r^2, describes the degree of common variance between distributions. Knowing the degree of common variance between two distributions provides the researcher with an insight of how well one distribution can be predicted from another. This concept belongs to the chapter on regression analysis, but is introduced here because the coefficient of determination is simply the square of the correlation coefficient, i.e., $(r)^2$.

The correlation coefficient computed from Table 5.2.1 was $r = 0.867$. Since this statistics, r is high, it indicates that the variables are linearly related and the coefficient of determination, r^2 *is* 0.752. Since the coefficient of determination tells us how much of the variation in the dependent variable, *Y*, is due to change in the independent variable, *X*. A r^2 of 0.752 tells us that ***75.2% of the variation in Y scores is associated with changes in X***. That means that 24.8% is caused by other factors. This 24.8% is known as the *coefficient of nondetermination* ($1 - r^2$) and is the complement of the coefficient of determination.