CHAPTER THREE
Investigating relationships among different sets of test scores

II. Conceptual exercises

A. Matching

1. c
2. h
3. g
4. e
5. f
6. d
7. a
8. b

B. True or false

1. F
2. F
3. T
4. T
5. F
6. T
7. F
8. F

C. Brief responses

1. A strong positive relationship means that there is much covariation between the two variables whereas a weak positive relationship will mean there is much less covariation. Another way of looking at this would be that in a strong positive relationship when values of one variable increase, the values of the second variable will also increase in more or less similar proportion. With a weak positive relationship, such a strong connection in increase in values between the two variables may not be noticed.

2. A strong negative relationship means that there is covariation between the two variables but the relationship is inverse. A weak negative relationship will mean very little covariation. Another way of looking at this would be that in a strong negative relationship when values of one variable increase, the values of the second variable decrease in more or less similar proportion. With a weak negative relationship, such an increase in the values of one variable does not correspond to a decrease in values of the other variable.

3. In a linear relationship between two variables, an increase or decrease in one variable results in a corresponding increase or decrease in the other variable. As a result, it is possible to draw a line through a scatterplot of a linear relationship between two variables to fit the relationship. In a curvilinear relationship, a change in one variable may not result in a similar or dissimilar change in the second variable; the rate of change in the second variable may vary depending on other factors and therefore if a line is drawn through a scatterplot of such a relationship between two variables, it will misrepresent the relationship.

4. It is appropriate to use and interpret the Pearson product-moment correlation coefficient when the following assumptions about the data are satisfied: (1) the relationship between the two variables is linear; (2) both variables constitute interval scales; and (3) both variables are normally distributed. It is appropriate to use the Spearman rank-order correlation coefficient when the following assumptions about the data are satisfied: (1) the relationship between the two variables is linear and (2) both variables constitute at least an ordinal scale.

5. Range restriction refers to the restriction of the range of scores on either or both variables. Unequal variation refers to the situation when the distributions of the two variables are different. One variable might be normally distributed and the other skewed, or both might be skewed.

D. Multiple choice

1. C. Explanation: Correlation values range from -1.0 to +1.0, and this value is out of this range.
2. C. Explanation: This is probably due to range restriction as most students missed many questions.
3. A. Explanation: When the points on a scatterplot are spread out, this is a clear indication that the two variables are not correlated.
4. A. Explanation: This is probably due to the result of combining groups of very different levels of ability.

5. C. Explanation: The correlation would be lower due to range restriction of the admitted group.

III. Calculations with small data sets

A. Examining scatterplots

Figure 3.1 Scatterplot 1: positive

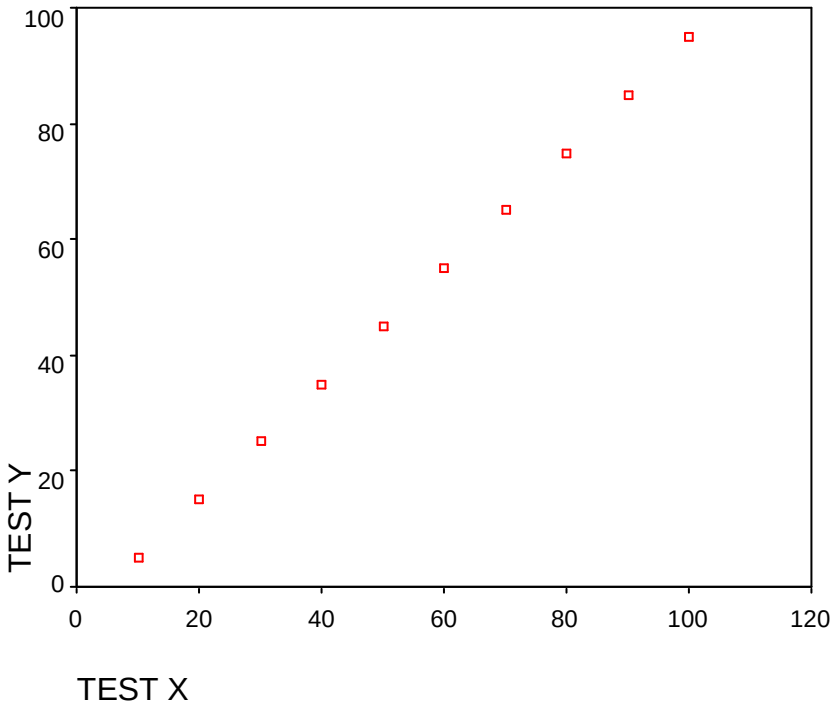
Figure 3.2 Scatterplot 2: negative

Figure 3.3 Scatterplot 3: no relationship

Figure 3.4 Scatterplot 4: curvilinear

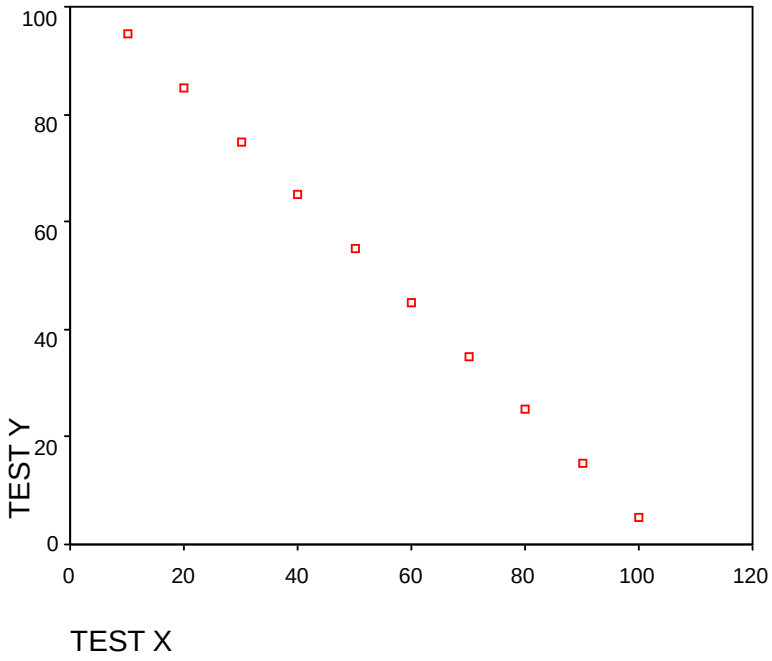
B. Plotting test scores

Data set 1



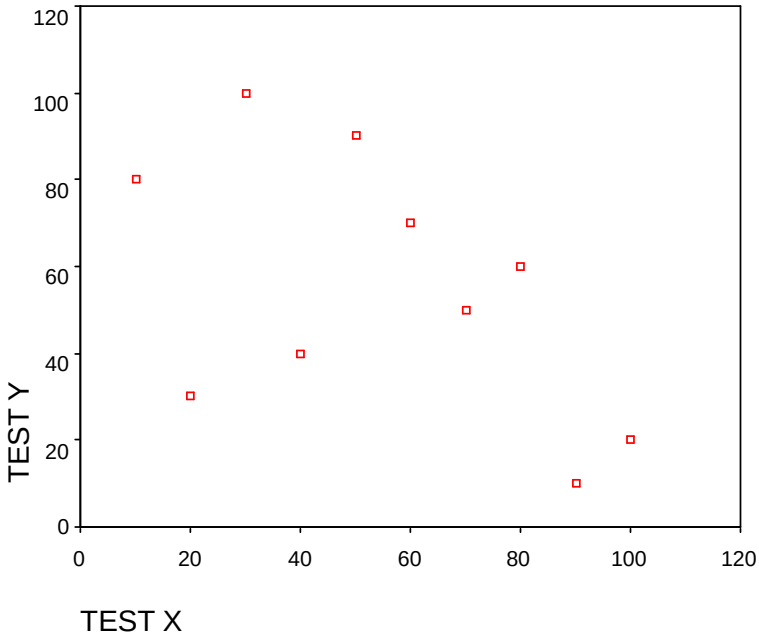
Description: perfect positive relationship

Data set 2



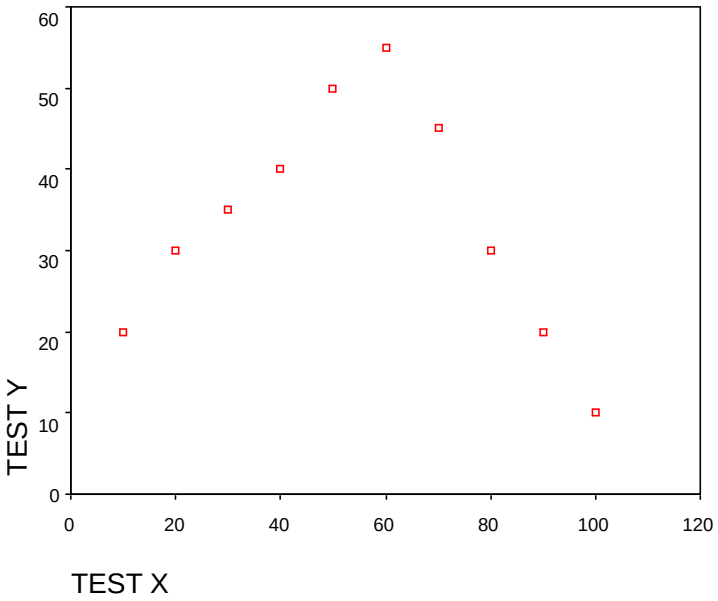
Description: perfect negative relationship

Data set 3



Description: no relationship

Data set 4



Description: curvilinear relationship

C. Pearson product-moment correlation coefficient

Table 3.1 *Deviation Score Method*

Student	**X**	$\overline{X}_X$	**x**	**Y**	$\overline{X}_Y$	**y**	**xy**
1	47	43.6	3.4	46	43.4	2.6	8.84
2	47	43.6	3.4	47	43.4	3.6	12.24
3	46	43.6	2.4	47	43.4	3.6	8.64
4	46	43.6	2.4	46	43.4	2.6	6.24
5	46	43.6	2.4	46	43.4	2.6	6.24
6	46	43.6	2.4	45	43.4	1.6	3.84
7	46	43.6	2.4	46	43.4	2.6	6.24
8	45	43.6	1.4	46	43.4	2.6	3.64
9	45	43.6	1.4	44	43.4	0.6	0.84
10	45	43.6	1.4	45	43.4	1.6	2.24
11	45	43.6	1.4	46	43.4	2.6	3.64
12	44	43.6	0.4	42	43.4	-1.4	-0.56
13	44	43.6	0.4	44	43.4	0.6	0.24
14	44	43.6	0.4	45	43.4	1.6	0.64
15	42	43.6	-1.6	38	43.4	-5.4	8.64
16	41	43.6	-2.6	38	43.4	-5.4	14.04
17	40	43.6	-3.6	40	43.4	-3.4	12.24
18	38	43.6	-5.6	40	43.4	-3.4	19.04
19	38	43.6	-5.6	37	43.4	-6.4	35.84
20	37	43.6	-6.6	39	43.4	-4.4	29.04
Σ							181.8
$\overline{X}$	43.6			43.4			
S	3.07			3.31			

The formula is:

$$r_{xy} = \frac{\sum xy}{N(S_X)(S_Y)} = \frac{181.8}{20(3.07)(3.31)} = \frac{181.8}{203.234}$$

r_{xy} = .895 (rounded to three decimals)

Table 3.2 *Raw Score Method*

Student	**X**	**X^2**	**Y**	**Y^2**	**XY**
1	47	2209	46	2116	2162
2	47	2209	47	2209	2209
3	46	2116	47	2209	2162
4	46	2116	46	2116	2116
5	46	2116	46	2116	2116
6	46	2116	45	2025	2070
7	46	2116	46	2116	2116
8	45	2025	46	2116	2070
9	45	2025	44	1936	1980
10	45	2025	45	2025	2025
11	45	2025	46	2116	2070
12	44	1936	42	1764	1848
13	44	1936	44	1936	1936
14	44	1936	45	2025	1980
15	42	1764	38	1444	1596
16	41	1681	38	1444	1558
17	40	1600	40	1600	1600
18	38	1444	40	1600	1520
19	38	1444	37	1369	1406
20	37	1369	39	1521	1443
∑	872	38208	867	37803	37983

The formula is:

$$r_{xy} = \frac{[\sum XY - (\sum X)(\sum Y) / N]}{\sqrt{[\sum X^2 - (\sum X)^2 / N][\sum Y^2 - (\sum Y)^2 / N]}}$$

$$r_{xy} = \frac{[37983 - (872)(867)/20]}{\sqrt{[38208 - (872)^2/20][37803 - (867)^2/20]}}$$

$$r_{xy} = \frac{[37983 - 37801.2]}{\sqrt{[38208 - 38019.2][37803 - 37584.45]}}$$

$$r_{xy} = \frac{181.8}{\sqrt{[188.8][218.55]}}$$

r_{XY} = .895 (rounded to three decimals)

Interpretation: The correlation coefficient of .895 indicates that the two variables are very highly correlated.

D. Spearman rank-order correlation coefficient

Student	**X**	**Y**	**Rx**	**Ry**	**d (Rx-Ry)**	**d^2**
1	47	46	1.5	5.5	-4	16
2	47	47	1.5	1.5	0	0
3	46	47	5	1.5	3.5	12.25
4	46	46	5	5.5	-0.5	0.25
5	46	46	5	5.5	-0.5	0.25
6	46	45	5	10	-5	25
7	46	46	5	5.5	-0.5	0.25
8	45	46	9.5	5.5	4	16
9	45	44	9.5	12.5	3	9
10	45	45	9.5	10	-0.5	0.25
11	45	46	9.5	5.5	4	16
12	44	42	13	14	-1	1
13	44	44	13	12.5	.5	.25
14	44	45	13	10	3	9
15	42	38	15	18.5	-3.5	12.25
16	41	38	16	18.5	-2.5	6.25
17	40	40	17	15.5	1.5	2.25
18	38	40	18.5	15.5	3	9
19	38	37	18.5	20	-1.5	2.25
20	37	39	20	17	3	9
						$\sum$= 146.5

$$r_S = 1 - \frac{6\Sigma d^2}{N(N^2-1)} = 1 - \frac{6(146.5)}{20((20)^2-1)} = 1 - \frac{79}{7980} = 1 - .1102$$

r_S = .890

Interpretation: The correlation coefficient of .892 indicates that the two variables are very highly correlated.

Which coefficient to report and why

While both the Pearson product-moment and the Spearman rank-order correlation coefficients can be interpreted in essentially the same way, as indicators of the strength and directionality of relationships between two variables, the Pearson r can be used appropriately with large samples where we can assume normal distributions. The Spearman r_S, on the other hand, does not assume that variables are normally distributed, and should thus be used to investigate relationships among variables with small sample sizes, as is typically the case with classroom-based assessments. (Because r_S is a special case of the product-moment correlation coefficient, as the size of the sample increases, the value of r_S will approach that of r.)

VI. SPSS exercises

Calculating and interpreting correlations:

A. *Reading Total Score* (readtot) and *Listening Total Score* (listot)

1. Pearson product-moment.
2. Correlation r = .432.
3. This is a somewhat weak positive relationship between the two variables.

B. *Speaking Total Score* (stotav) and *Writing Total Score* (wtotav)

1. Pearson product-moment.
2. Correlation r = .575.
3. This is a moderate positive relationship between the two variables.

C. *Reading Total Score* (readtot) and *Writing Total Score* (wtotav)

1. Pearson product-moment.
2. Correlation r = .287.
3. This is a weak relationship between the two variables.

D. *Listening Total Score* (listot) and *Speaking Total Score* (stotav)

1. Pearson product-moment.

2. Correlation r = .552.

3. This is a moderate positive relationship between the two variables.