

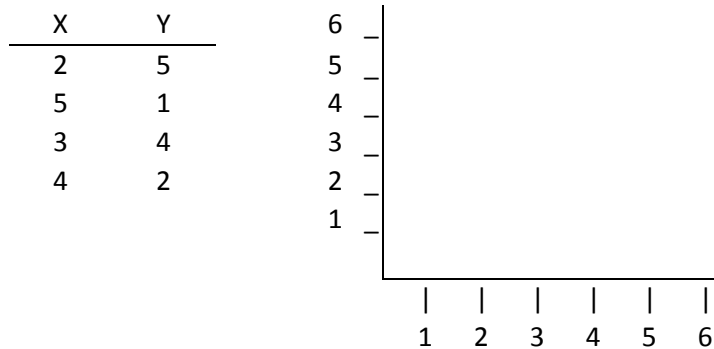
## Soc 102 - Final Exam – Correlation and Regression (Practice)

### Multiple Choice

Identify the choice that best completes the statement or answers the question.

- A Pearson correlation of  $r = -0.85$  indicates that a graph of the data would show \_\_\_\_\_.
  - points clustered close to a line that slopes up to the right
  - points clustered close to a line that slopes down to the right
  - points widely scattered around a line that slopes up to the right
  - points widely scattered around a line that slopes down to the right
- A positive value for a correlation indicates \_\_\_\_\_.
  - increases in X tend to be accompanied by increases in Y
  - increases in X tend to be accompanied by decreases in Y
  - a much stronger relationship than if the correlation were negative
  - a much weaker relationship than if the correlation were negative
- A scatterplot shows a set of data points that are widely scattered around a line that slopes up to the right. Which of the following values would be closest to the correlation for these data?
  - 0.80
  - 0.40
  - 0.40
  - 0.80
- A set of X and Y scores has  $SS_X = 10$ ,  $SS_Y = 20$ , and  $SP = 8$ . What is the slope for the regression equation?
  - 8/10
  - 8/20
  - 10/8
  - 20/8
- A set of  $N = 5$  pairs of X and Y values has  $SS_X = 10$ ,  $SS_Y = 40$  and  $SP = 10$ . For these data, the Pearson correlation is \_\_\_\_\_.
  - $r = 10/20 = 0.50$
  - $r = 10/400 = 0.025$
  - $r = -10/20 = 0.707$
  - $r = 40/100 = 0.40$
- For a hypothesis test for the Pearson correlation, the null hypothesis states that \_\_\_\_\_.
  - there is a non-zero correlation for the general population
  - the population correlation is zero
  - there is a non-zero correlation for the sample
  - the sample correlation is zero
- For which of the following correlations would the data points be clustered most closely around the regression line?
  - $r = 0.10$
  - $r = 0.50$
  - $r = -0.80$
  - There is no relationship between the correlation and how close the data points are to the regression line.
- If you wanted to measure the correlation between how people voted on a ballot measure (Yes or No) and how much money they earn, you should use \_\_\_\_\_.
  - Pearson's
  - Kendall's tau-b
  - Spearman's
  - Can't do it

9. Draw the scatterplot for these data. Which choice below best describes the Pearson correlation for these data?



- a. Positive  
b. Negative  
c. Zero  
d. Cannot be determined
10. In the general linear equation,  $Y = bX + a$ , what is the value of “a” called?  
a. the slope constant  
b. the Y-intercept  
c. the X-intercept  
d. the beta factor
11. The numerical value for a correlation \_\_\_\_\_.  
a. can never be greater than 1.00  
b. can never be less than -1.00  
c. can never be greater than 1.00 and can never be less than -1.00  
d. can be greater than 1.00 and can be less than -1.00
12. The Pearson correlation measures \_\_\_\_\_.  
a. the degree of relationship without regard to the form of the relationship  
b. the degree to which the relationship is consistently one directional  
c. the degree of linear relationship  
d. the degree of curvilinear relationship
13. The Pearson correlation is calculated for a sample of  $N = 25$  individuals. What value of df should be used to determine whether or not the correlation is significant?  
a. 23  
b. 24  
c. 25  
d. cannot be determined without additional information
14. The Pearson and the Spearman correlations are both computed for the same set of data. If the Pearson correlation is  $r = +1.00$ , then what can you conclude about the Spearman correlation?  
a. It will be positive.  
b. It will have a value of 1.00  
c. It will be positive and have a value of 1.00  
d. There is no predictable relationship between the Pearson and the Spearman correlations.
15. The regression equation is determined by minimizing \_\_\_\_\_.  
a. the total error between the X and Y values  
b. the total error between the predicted Y values and the actual Y values  
c. the total squared error between the X and the Y values  
d. the total squared error between the predicted Y values and the actual Y values

16. The results from a research study indicate that adolescents who watch more violent content on television also tend to engage in more violent behavior than their peers. The correlation between amount of TV violent content and amount of violent behavior is an example of \_\_\_\_\_.
- a positive correlation
  - a negative correlation
  - a correlation near zero
  - a correlation near one
17. What is indicated by a positive value for a correlation?
- Increases in X tend to be accompanied by increases in Y
  - Increases in X tend to be accompanied by decreases in Y
  - A much stronger relationship than if the correlation were negative
  - A much weaker relationship than if the correlation were negative
18. What is the value of SP for the following set of data?

| X | Y |
|---|---|
| 1 | 5 |
| 3 | 3 |
| 5 | 1 |

- 0
  - 8
  - +8
  - 19
19. Which of the following Pearson correlations shows the greatest strength or consistency of relationship?
- 0.90
  - +0.74
  - +0.85
  - 0.33

## Soc 102 - Final Exam – Correlation and Regression (Practice)

### Answers

1. A Pearson correlation of  $r = -0.85$  indicates that a graph of the data would show \_\_\_\_\_.
- points clustered close to a line that slopes up to the right
  - points clustered close to a line that slopes down to the right
  - points widely scattered around a line that slopes up to the right
  - points widely scattered around a line that slopes down to the right

ANS: B

The value of  $r$  is negative, so the line will slope down from left to right ( $\searrow$ ), and the correlation is strong, so the points will fall close to the line.

2. A positive value for a correlation indicates \_\_\_\_\_.
- increases in X tend to be accompanied by increases in Y
  - increases in X tend to be accompanied by decreases in Y
  - a much stronger relationship than if the correlation were negative
  - a much weaker relationship than if the correlation were negative

ANS: A

The sign only tells us about direction, not strength. Negative means that high values in X are associated with low values in Y, and low values in X with high values in Y.

3. A scatterplot shows a set of data points that are widely scattered around a line that slopes up to the right. Which of the following values would be closest to the correlation for these data?
- 0.80
  - 0.40
  - 0.40
  - 0.80

ANS: B

“Widely scattered” suggests a low value for  $r$ , and the line sloping up to the right ( $\nearrow$ ) a positive slope.

4. A set of X and Y scores has  $SS_X = 10$ ,  $SS_Y = 20$ , and  $SP = 8$ . What is the slope for the regression equation?
- 8/10
  - 8/20
  - 10/8
  - 20/8

ANS: A

Slope is  $SP/SS_X$ , so 8/10.

5. A set of  $N = 5$  pairs of  $X$  and  $Y$  values has  $SSX = 10$ ,  $SSY = 40$  and  $SP = 10$ . For these data, the Pearson correlation is \_\_\_\_\_.
- a.  $r = 10/20 = 0.50$
  - b.  $r = 10/400 = 0.025$
  - c.  $r = -10/20 = 0.707$
  - d.  $r = 40/100 = 0.40$

ANS: A

The correlation  $r$  is equal to  $r = \frac{SP}{\sqrt{SS_X SS_Y}} = \frac{10}{\sqrt{10 * 40}} = \frac{10}{\sqrt{400}} = \frac{10}{\sqrt{400}} = \frac{10}{20}$

6. For a hypothesis test for the Pearson correlation, the null hypothesis states that \_\_\_\_\_.
- a. there is a non-zero correlation for the general population
  - b. the population correlation is zero
  - c. there is a non-zero correlation for the sample
  - d. the sample correlation is zero

ANS: B

The hypotheses are always about a population (that's why we have so far used  $\mu$  in one way or another). Since the null hypothesis is always that there isn't whatever we are looking for, in the case of a correlation it would be that there is no correlation ( $r = 0$ ).

7. For which of the following correlations would the data points be clustered most closely around the regression line?
- a.  $r = 0.10$
  - b.  $r = 0.50$
  - c.  $r = -0.80$
  - d. There is no relationship between the correlation and how close the data points are to the regression line.

ANS: C

The regression line aims to minimize the distance between predictions and actual values. If the correlation is high, the difference between predicted values and actual values will be low, and if the correlation is low, the distances will be bigger. In this case, we just need to look for the highest correlation, and not worry about the sign, since direction does not affect how well a regression line fits.

8. If you wanted to measure the correlation between how people voted on a ballot measure (Yes or No) and how much money they earn, you should use \_\_\_\_\_.
- a. Pearson's                      c. Spearman's  
b. Kendall's tau-b                d. Can't do it

ANS: B

Pearson's needs at least interval data, Spearman's is for rank-ordered (ordinal) data, so only Kendall's tau-b can be used with a dichotomous (2 values) variable.

9. Draw the scatterplot for these data. Which choice below best describes the Pearson correlation for these data?



- a. Positive                      c. Zero  
b. Negative                      d. Cannot be determined

ANS: B



10. In the general linear equation,  $Y = bX + a$ , what is the value of “a” called?
- the slope constant
  - the Y-intercept
  - the X-intercept
  - the beta factor

ANS: B

11. The numerical value for a correlation \_\_\_\_\_.
- can never be greater than 1.00
  - can never be less than  $-1.00$
  - can never be greater than 1.00 and can never be less than  $-1.00$
  - can be greater than 1.00 and can be less than  $-1.00$

ANS: C

12. The Pearson correlation measures \_\_\_\_\_.
- the degree of relationship without regard to the form of the relationship
  - the degree to which the relationship is consistently one directional
  - the degree of linear relationship
  - the degree of curvilinear relationship

ANS: C

Pearson's is only good for linear relationships.

13. The Pearson correlation is calculated for a sample of  $N = 25$  individuals. What value of df should be used to determine whether or not the correlation is significant?
- 23
  - 24
  - 25
  - cannot be determined without additional information

ANS: A

For a correlation  $df = N - 2$ .

14. The Pearson and the Spearman correlations are both computed for the same set of data. If the Pearson correlation is  $r = +1.00$ , then what can you conclude about the Spearman correlation?
- It will be positive.
  - It will have a value of 1.00
  - It will be positive and have a value of 1.00
  - There is no predictable relationship between the Pearson and the Spearman correlations.

ANS: C

If the correlation is perfect ( $r = +1.00$ ), it will be perfect with either Pearson's or Spearman's.

15. The regression equation is determined by minimizing \_\_\_\_\_.
- the total error between the X and Y values
  - the total error between the predicted Y values and the actual Y values
  - the total squared error between the X and the Y values
  - the total squared error between the predicted Y values and the actual Y values

ANS: D

16. The results from a research study indicate that adolescents who watch more violent content on television also tend to engage in more violent behavior than their peers. The correlation between amount of TV violent content and amount of violent behavior is an example of \_\_\_\_\_.
- |                           |                            |
|---------------------------|----------------------------|
| a. a positive correlation | c. a correlation near zero |
| b. a negative correlation | d. a correlation near one  |

ANS: A

When values in one variable are like the values in another variable (high with high, low with low) correlations are positive. In this case high values of violent TV watching are found with high levels of violent behavior, so a positive correlation.

17. What is indicated by a positive value for a correlation?
- Increases in X tend to be accompanied by increases in Y
  - Increases in X tend to be accompanied by decreases in Y
  - A much stronger relationship than if the correlation were negative
  - A much weaker relationship than if the correlation were negative

ANS: A



18. What is the value of SP for the following set of data?

|                  | X | Y | X <sup>2</sup> | Y <sup>2</sup> | XY |
|------------------|---|---|----------------|----------------|----|
|                  | 1 | 5 | 1              | 25             | 5  |
|                  | 3 | 3 | 9              | 9              | 9  |
|                  | 5 | 1 | 25             | 1              | 5  |
| <hr/> $\Sigma =$ | 9 | 9 | 35             | 35             | 19 |

- a. 0
- b. -8
- c. +8
- d. 19

ANS: B

We use this formula  $SP = \sum XY - \frac{(\sum X)(\sum Y)}{N}$

We need to calculate  $\sum X$ ,  $\sum Y$  and  $\sum XY$

$$SP = \sum XY - \frac{(\sum X)(\sum Y)}{N} = 19 - \frac{9*9}{3} = 19 - \frac{81}{3} = 19 - 27 = -8$$

19. Which of the following Pearson correlations shows the greatest strength or consistency of relationship?

- a. -0.90
- c. +0.85
- b. +0.74
- d. -0.33

ANS: A

Since strength does not depend on direction, we can ignore the signs and look for the biggest absolute value.