CHAPTER 6

Correlation and Regression

Summary

Correlation and ***linear regression*** are two different statistical methods that are closely related mathematically. Both methods provide information about bivariate distributions. A ***bivariate distribution*** has two variables whose scores are logically paired.

Correlation and regression are used for different purposes. A ***correlation coefficient*** (symbolized as *r),* is used to describe the *degree* and the *direction* of a relationship between two variables. A ***regression equation*** is used to draw the ***regression line*** (also known as the *line of best fit*) and to *predict scores* on one variable, if you have scores on the other variable. The textbook describes the simplest case for these two methods, the case that requires that the two variables have a *linear relationship.* You can make a visual check for linearity by constructing and examining a ***scatterplot*** of the two variables. This is a recurring theme in the text and the study guide; when in doubt, graph the data to gain insight into the nature of the relationship between variables.

To summarize several of the characteristics of *r* and its interpretation:

a. The *algebraic sign* of *r* gives the direction of the relationship. If the sign is positive, the relationship is direct (higher scores on one variable go with higher scores on the other variable). If the sign is negative, the relationship is inverse (higher scores on one variable go with lower scores on the other).
b. The closer the absolute value of *r* is to 1.00, the stronger the relationship, and the more confidence you can put into a prediction made from a regression equation based on the data.
c. Positive coefficients are not "better" than negative coefficients.
d. r^2, the ***coefficient of determination,*** gives the proportion of variance the two variables have in common.
e. Correlation coefficients, no matter how large, *are not sufficient evidence to claim a causal relationship between two variables.*
f. Low correlations do not necessarily mean that there is no relationship between the two variables; ***nonlinearity*** and ***truncated ranges*** both produce spuriously (artificially) low correlation coefficients.
g. The ***effect size index for r*** is the correlation coefficient itself. Depending on the reason that *r* was calculated, descriptive adjectives of *small, medium, and large* are appropriate for different values of *r.*
h. When the same measure is administered twice to the same subjects, a correlation coefficient of .80 or greater indicates that the measure has adequate ***reliability****.*

Besides the Pearson product-moment correlation coefficient, which is used for *two quantitative variables,* other kinds of correlation coefficients are used when the relationship between variables is examined. Any correlation coefficient expresses the strength and direction of the relationship between variables.

A *regression equation,* $\hat{Y} = a + bX,$ allows you to predict a value for *Y* at any value of *X.* The prediction will be more accurate when the relationship between *X* and *Y* is linear and the correlation coefficient is large.

To write a regression equation for a bivariate distribution, calculate values for the *two* ***regression coefficients****, a* and *b.* The regression coefficient *a* is the ***intercept*** of the regression line with the *Y* axis, and the coefficient *b* is the ***slope*** of the regression line.

A regression line can be presented as a graph, but its appearance will depend on the units used on the *X* and *Y* axes. Also, there are two regression lines for one set of bivariate data. The line that your calculations produce depends on which variable you designate as the *Y* variable.

Multiple-Choice Questions

1. To use the regression equation technique described in your text, you must have ________________.

    a. a logical pairing of the scores on the two variables
    b. a linear relationship between the two variables

c. both a. and b.
d. neither a. nor b.

2. Quantification is the idea that ________________.

a. all things can be counted
b. all physical things can be counted
c. the numerical representation of a phenomenon gives the most important picture
d. a phenomenon can be better understood if its important parts are expressed as numbers

3. Pearson correlation coefficient is appropriate to describe which of the situations below?

a. as X increases, Y decreases by the same amount
b. as X increases, Y goes up at first slowly and then faster
c. as X increases, Y goes up at first and then goes down
d. all of the above

4. A correlation of -.88 between television viewing time and grades in high school is best understood as demonstrating that ________________.

a. as television viewing time increases, grades increase
b. as television viewing time decreases, grades decrease
c. as television viewing time increases, grades decrease
d. an error was made in calculating the correlation (a value of -.88 is not possible)

5. Assume you conduct a study to evaluate the relationship between the amount of time a child is read to and reading ability at age 15. You find a correlation coefficient of .04. This correlation suggests ________________.

a. a very strong relationship; as reading to children increases, reading ability increases
b. a very weak relationship; as reading to children increases, reading ability decreases slightly
c. almost no relationship
d. cannot be determined from the information given

6. A linear relationship is described by which of the statements below?

a. the two variables are paired in some logical fashion
b. for every one-point increase in one variable, you get a four-point increase in the other variable
c. both a. and b.
d. neither a. nor b.

7. Which of the following can be determined from a correlation coefficient?

a. degree of causation
b. degree of reliability
c. proportion of variance shared between X and Y
d. both b. and c.

8. Which of the following statements is (are) true?

a. correlations range from 0 to +1
b. correlations show causal relationships
c. correlations quantify the strength of a relationship
d. both a. and c.

9. Psychologists have demonstrated that number of hours spent in class is correlated with grades in that class. The correlation between number of hours in class and grades is ________________.

a. positive
b. negative
c. zero
d. cannot be determined from the information given

10. Which of the following describes a negative correlation?

a. the more I study, the better grades I get
b. the more I meditate, the more calm I am
c. the less I sleep, the less I can focus
d. the more caffeine I drink, the less I nap

11. The coefficient of determination allows you to ________________.

a. determine the proportion of variance two variables have in common
b. draw cause-and-effect statements
c. predict *X* scores, given *Y* scores
d. quickly determine the regression coefficients

12. Given a Pearson correlation coefficient of zero, which conclusion is correct?

a. a Pearson correlation coefficient is not proper for the data
b. there is no relationship between the two variables
c. correlation coefficients of zero cannot be interpreted
d. all of the above

13. Which of the following could **not** be a correlation coefficient?

a. -.65
b. .00
c. -1.50
d. -1.00

14. Pearson product-moment correlation coefficients can be used to establish the degree of relationship ________________.

a. even if the two variables are measuring different things
b. even if one of the variables is measured on a nominal scale
c. even if the relationship is not linear
d. all of the above

15. The regression coefficient, *a,* tells us ________________.

a. the slope of the regression line
b. what to predict for *X* when $Y = 0$
c. the proportion of variance shared between variable *X* and variable *Y*
d. the place where the regression line crosses the *y*-axis

16. A correlation coefficient may be small because of ________________.

a. a truncated range
b. a nonlinear relationship
c. both a. and b.
d. neither a. nor b.

17. Suppose you calculated the regression coefficients for the line that predicts the number of offspring a woman will have from the number of offspring of her maternal grandmother and found a = 2.3 and b = -1.0. Knowing this, you can conclude that the correlation coefficient for these data is ________________.

a. positive
b. negative
c. perfect
d. cannot be determined from the information given

18. Two ________________ exist for one set of bivariate data.

a. correlation coefficients
b. regression lines
c. coefficients of determination
d. all of the above

19. Suppose you had the exam scores for the first exam in General Psychology for 100 students. A Pearson correlation coefficient could be calculated if the scores were divided according to the variable ________________.

a. gender identity (male, female, non-binary)
b. where a person sits in the class (front or back)
c. both a. and b.
d. neither a. nor b.

20. The difference between correlation coefficients and regression equations is that correlation coefficients ________________.

a. allow us to infer causation, whereas regression equations do not
b. allow for prediction of scores, whereas regression equations do not
c. quantify the relationship between two variables, whereas regression equations can predict scores
d. require bivariate distributions, whereas regression equations do not

Short-Answer Questions

1. "The self-confidence of that group of recruits is negatively correlated with their success in the obstacle course." Tell what this statement means.

2. Describe the statistical method of linear regression. Tell what it is good for and what its limitations are.

3. A study of 4,138 students in 25 law schools found a correlation coefficient of .36 between first-year law school grades and scores on the Law School Admission Test (LSAT). Interpret the meaning of this correlation.

4. Two regression coefficients can be calculated from a bivariate distribution. Name these two coefficients and describe what each tells us.

5. Label each scenario as describing a positive correlation or a negative correlation.

a. As the speed of the car goes up, the amount of fuel consumed increases.
b. The longer I wait, the angrier I feel.
c. As I get older, I go to bed earlier.
d. If I cook at home more, I'll spend less money.
e. The less I watch Netflix, the more homework I get done.
f. If I drink less water, I feel less energetic.

Problems

1. Look at the below data and write down an estimate of what the correlation coefficient would be. Then, calculate r using the Pearson product-moment definition formula.

X	Y
1	1
3	3
4	5
5	9
8	12

2. Look at the data that follow and write down an estimate of what the correlation coefficient would be. Then, calculate r using the Pearson product-moment definition formula.

X	Y
2	8
4	6
5	2
6	5
8	4

3. Huynh, Oakes, Shay, and McGregor (2017) studied the relationship between the pursuit of virtue and one's ability to reason wisely about social conflicts. The researchers assessed pursuit of virtue by asking participants their agreement with statements like, "I would like to do what I believe in." They assessed participants' endorsement of wise reasoning strategies by asking them how useful they found strategies like "making an effort to take the other person's perspective" when facing a social conflict. Participants answered the wise reasoning questions about either their own conflicts (e.g., how useful they think perspective-taking is for their *own* conflicts) or other people's conflicts (e.g., how useful they think perspective-taking is for *other people's* conflicts). Below are summary data from Study 1 of Huynh et al. (2017; retrieved from https://osf.io/4chuy/ on July 6, 2018) that will allow you to closely replicate correlations reported in Table S1 of the supplemental material for this paper. Calculate the correlation between pursuit of virtue and wise reasoning for both conflicts involving the self and conflicts involving other people. Write a short interpretation of your findings.

Self	
Pursuit of Virtue Scores	Wise Reasoning Scores
$\Sigma X = 573.6$	$\Sigma Y = 479.74$
$\Sigma X^2 = 2{,}567.76$	$\Sigma Y^2 = 1{,}833.70$
$\Sigma XY = 2{,}138.37$	
$N = 130$	

Other	
Pursuit of Virtue Scores	Wise Reasoning Scores
$\Sigma X = 563$	$\Sigma Y = 504.16$
$\Sigma X^2 = 2{,}557.32$	$\Sigma Y^2 = 2{,}064.73$
$\Sigma XY = 2{,}259.12$	
$N = 126$	

4. The idea of using tests to predict who will do well in college began to emerge around 1900. Many (including Galton) assumed that people with quick reaction times and keen sensory abilities would be quick thinkers with keen intellects (who would, of course, make good grades). James McKeen Cattell at Columbia University gathered data on this assumption. The summary statistics below are representative of his findings, as reported by Clark Wissler (for an overview, see Sokal, 1982).

Sensory Ability Score	Grade Point Average
$\Sigma X = 3{,}500$	$\Sigma Y = 115$
$\Sigma X^2 = 250{,}000$	$\Sigma Y^2 = 289$
$\Sigma XY = 8{,}084$	
$N = 50$	

a. Calculate the correlation coefficient between sensory ability and first-year grade point average.
b. Calculate *a* and *b* and then construct the regression equation that predicts GPA.
c. Predict the GPA of a person with a sensory ability score of 100.
d. Write one sentence that summarizes how confident you are in the accuracy of your above prediction.

5. Is creativity related to humor? Each student in this data set has a score on a test of creativity and a score based on the number of puns produced while looking at a list of "wise sayings." Raw data are presented below.

Student	Creativity Test Score	Number of Puns
1	60	28
2	57	32
3	52	24
4	46	16
5	41	21
6	38	14
7	32	18
8	29	11
9	25	9
10	19	12

a. Calculate the correlation coefficient between creativity test score and number of puns.
b. Calculate the coefficient of determination.
c. Calculate *a* and *b* and then construct the regression equation that predicts number of puns.
d. Predict the number of puns for a student whose creativity test score was 35.
e. Draw a scatterplot of the data and plot the regression line on your scatterplot.
f. Write an explanation of what your analysis shows.

ANSWERS

Multiple-Choice Questions

1. c
   Explanation: Just having scores for two variables is not sufficient—each score on the *X* variable must be logically paired with a score on the *Y* variable. For example, scores on *X* and *Y* could come from the same person or two people who are related. If a relationship between two variables is not linear, then a straight line (which is what the regression equation produces) will not accurately represent the relationship between *X* and *Y*.
2. d
3. a
   Explanation: Both b. and c. describe nonlinear relationships between *X* and *Y*. If that's hard to picture through written words, try sketching out each description in a quick graph.
4. c
   Explanation: Both a. and b. describe positive correlations, because *X* and *Y* "move together" (increases are associated with increases and decreases are associated with decreases). The sign of the correlation coefficient indicates that the correlation is negative (and -.88 *is* a possible value for *r*).
5. c
6. b
   Explanation: Scores on two variables are required to be paired in some logical fashion in order to calculate a correlation coefficient, but this requirement is unrelated to whether the relationship between these variables is linear or nonlinear.
7. d
   Explanation: The correlation coefficient for two administrations of the same test tells us about the reliability of that test. If we square a correlation coefficient, we get the coefficient of determination (which tells us c.). We can *never* tell the degree of causation from a correlation coefficient.
8. c
   Explanation: Correlation coefficients have a possible range of -1 to +1. And, you should never conclude a causal relationship from a correlation coefficient. There *could* be a causal relationship between *X* and *Y*, but only an experimental research design can tell us about causality. Return to the section of the chapter titled, "To Establish Causation—NOT!," if you want a refresher on correlations and causality.
9. d
   Explanation: Though it's tempting to assume that spending more hours in class would be related to *higher* grades in that class (a positive correlation), that level of detail is not provided in the question.
10. d
11. a
12. b
    Explanation: A correlation coefficient of zero represents no relationship between two variables. However, it's always a good idea to make a scatterplot of these data to check for *nonlinearity* (in which case, a Pearson correlation would not be appropriate). If you suspect one or both of your variables might have a *truncated range*, the correlation coefficient will be spuriously low.
13. c
14. a
    Explanation: Both b. (nominal scale) and c. (nonlinearly) describe situations in which a Pearson correlation coefficient would not be appropriate. The Pearson product-moment correlation coefficient can only be calculated for two quantitative variables that are linearly related.
15. d
    Explanation: The regression coefficient, *a*, gives us the *y*-intercept. It's also accurate to say that *a* represents what score to predict for *Y* when $X = 0$ (you can see this by looking at the formula for *a*).
16. c
17. b
    Explanation: The regression coefficient, *b*, tells us the slope of the line that best represents the relationship between variable *X* (here, number of offspring of maternal grandmother) and variable *Y* (here, the number of offspring a woman will have). Because the slope of this line is negative, we know that the correlation coefficient for these two variables is also negative.
18. b
    Explanation: For any set of bivariate data, you could construct the regression line that predicts Variable 1 from Variable 2 and the regression line that predicts Variable 2 from Variable 1. In the first case, Variable 1 would

be Y; in the second case, Variable 2 would be Y.

19. d

Explanation: Pearson correlation coefficients can only be calculated between two *quantitative* (continuous or discrete) variables; both a. and b. represent categorical variables.

20. c

Short-Answer Questions

1. This statement means that higher feelings of self-confidence in recruits are associated with worse performance on the obstacle course. It also means that *lower* feelings of self-confidence in recruits are associated with *better* performance on the obstacle course.
2. Linear regression is a statistical method by which a straight line is constructed that best represents (or "fits") a bivariate relationship. If two variables are correlated, linear regression can be used to predict scores on one variable, given scores on another variable. The stronger the correlation between the two variables, the more confidence you can have in your predictions. If two variables are not linearly related to one another, linear regression is not appropriate.
3. There is a moderate positive correlation between first-year law school grades and scores on the LSAT, such that higher LSAT scores are associated with higher first-year grades. LSAT performance accounts for only 13% of the variance in first-year law school grades (r^2 = .13). Because of a truncated range for each variable, though, this correlation is likely spuriously low. Only individuals who score high on the LSAT get admitted to law school and grade distributions in graduate and professional programs tend to be rather high (mostly As and Bs). Such restrictions on variability always reduce the size of r.
4. The two regression coefficients are a and b. The coefficient, a, tells us where the regression line crosses the y-axis and is also called the intercept. The coefficient, b, tells us the slope of the regression line. Another way to interpret b is how much we expect Y to change (and in what direction) for every 1-unit increase in X.
5.
   a. positive
   b. positive
   c. negative
   d. negative
   e. negative
   f. positive

Problems

1. For every pair of scores, higher X values are associated with higher Y values. Thus, these data will show a strong positive r value. The calculated value is r = .97.

$$\bar{X} = \frac{21}{5} = 4.2 \quad \bar{Y} = \frac{30}{5} = 6$$

$$S_X = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} = \sqrt{\frac{115 - \frac{(21)^2}{5}}{5}} = 2.32$$

$$S_Y = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}} = \sqrt{\frac{260 - \frac{(30)^2}{5}}{5}} = 4$$

X	Y	z_X	z_Y	$z_X z_Y$
1	1	-1.38	-1.25	1.73
3	3	-0.52	-0.75	0.39
4	5	-0.09	-0.25	0.02

5	9	0.35	0.75	0.26
8	12	1.64	1.5	2.46

$$r = \frac{\sum(z_X z_Y)}{N} = \frac{4.86}{5} = .97$$

2. For most pairs of scores, higher *X* values are associated with smaller *Y* values, so these data will show an overall negative correlation. The difference between *X* and *Y* pairs varies throughout the bivariate distribution, so the correlation is not likely to be close to -1.00. The calculated value is $r = -.65$.

$$\bar{X} = \frac{25}{5} = 5 \quad \bar{Y} = \frac{25}{5} = 5$$

$$S_X = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} = \sqrt{\frac{145 - \frac{(25)^2}{5}}{5}} = 2$$

$$S_Y = \sqrt{\frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N}} = \sqrt{\frac{145 - \frac{(25)^2}{5}}{5}} = 2$$

X	*Y*	z_X	z_Y	$z_X z_Y$
2	8	-1.5	1.5	-2.25
4	6	-0.5	0.5	-0.25
5	2	0	-1.5	0
6	5	0.5	0	0
8	4	1.5	-0.5	-0.75

$$r = \frac{\sum(z_X z_Y)}{N} = \frac{-3.25}{5} = -.65$$

3. Self: $r = .40$; Other: $r = .14$
Pursuit of virtue is positively correlated with wise reasoning scores for conflicts involving the self, $r = .40$, and for conflicts involving other people, $r = .14$. The correlation between virtue and wise reasoning is stronger for conflicts involving the self than conflicts involving others.

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

$$= \frac{130(2{,}138.37) - (573.6)(479.74)}{\sqrt{[130(2{,}576.76) - (573.6)^2][130(1{,}833.70) - (479.74)^2]}}$$

$$= \frac{277{,}988.1 - 275{,}178.864}{\sqrt{[334{,}978.8 - 329{,}016.96][238{,}381 - 230{,}150.468]}}$$

$$= \frac{2{,}809.236}{\sqrt{[5{,}961.84][8{,}230.532]}} = \frac{2{,}809.236}{\sqrt{49{,}069{,}114.9}} = \frac{2{,}809.236}{7{,}004.935} = .401 = .40$$

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

$$= \frac{126(2{,}259.12) - (563)(504.16)}{\sqrt{[126(2{,}557.32) - (563)^2][126(2{,}064.73) - (504.16)^2]}}$$

$$= \frac{284{,}649.12 - 283{,}842.08}{\sqrt{[322{,}222.32 - 316{,}969][260{,}155.98 - 254{,}177.306]}}$$

$$= \frac{807.04}{\sqrt{[5{,}253.32][5{,}978.674]}} = \frac{807.04}{\sqrt{31{,}407{,}887.7}} = \frac{807.04}{5{,}604.274} = .144 = .14$$

4.

a. $r = .10$
b. $b = 0.01$; $a = 1.6$; $\hat{Y} = 1.6 + 0.01X$
c. $\hat{Y} = 2.6$
d. Because the degree of relationship between sensory ability and GPA is so slight (a correlation of only .10), confidence in predicting GPA from sensory ability should be pretty low. Predictions from this regression equation are likely to have low accuracy.

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

$$= \frac{50(8{,}084) - (3{,}500)(115)}{\sqrt{[50(250{,}000) - (3{,}500)^2][50(289) - (115)^2]}}$$

$$= \frac{404{,}200 - 402{,}500}{\sqrt{[12{,}500{,}000 - 12{,}250{,}000][14{,}450 - 13{,}225]}}$$

$$= \frac{1{,}700}{\sqrt{[250{,}000][1{,}225]}} = \frac{1{,}700}{\sqrt{306{,}250{,}000}} = \frac{1{,}700}{17{,}500} = .097 = .10$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} = \frac{50(8{,}084) - (3{,}500)(115)}{50(250{,}000) - (3{,}500)^2}$$

$$= \frac{404{,}200 - 402{,}500}{12{,}500{,}000 - 12{,}250{,}000}$$

$$= \frac{1{,}700}{250{,}000} = 0.0068 = 0.01$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{3{,}500}{50} = 70$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{115}{50} = 2.3$$

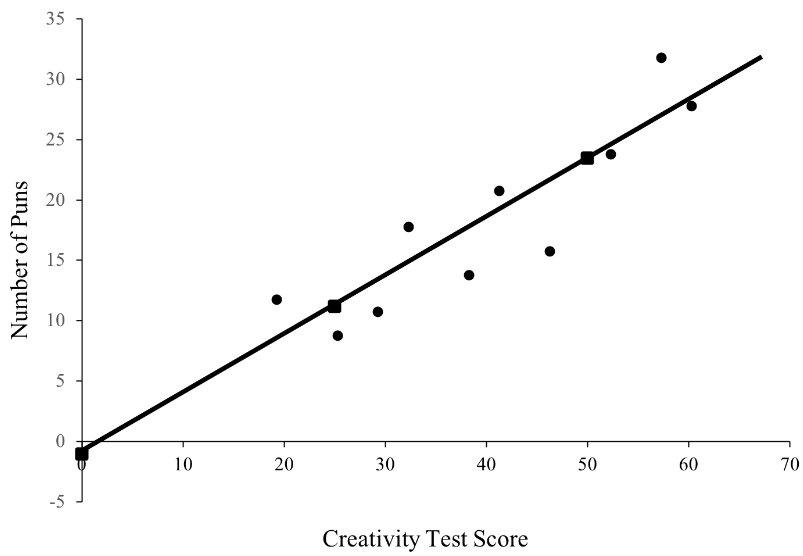
$$a = \bar{Y} - b\bar{X} = 2.3 - (0.01)(70) = 2.3 - 0.7 = 1.6$$

$$\hat{Y} = 1.6 + 0.01X$$

$$\hat{Y} = 1.6 + 0.01X = 1.6 + 0.01(100) = 1.6 + 1 = 2.6$$

5.

a. $r = .89$
b. $r^2 = .79$
c. $b = 0.49$; $a = -1.05$; $\hat{Y} = -1.05 + 0.49X$
d. $\hat{Y} = 16.1$
e.



Values used to draw regression line: $X = 0$, $Y = -1.05$ (y-intercept); $X = 25$, $Y = 11.2$; $X = 50$, $Y = 23.45$

f. The scatterplot shows that the relationship between scores on the creativity test and number of puns generated is linear. There is a very strong positive correlation between creativity test scores and number of puns generated, $r = .89$, such that higher creativity scores are associated with more puns. With a coefficient of determination of .79, almost 80% of the variance in number of puns generated is explained by variance in creativity scores. Thus, these two tests seem to be measuring similar abilities.

Student	Creativity Test Score (X)	X^2	Number of Puns (Y)	Y^2	XY
1	60	3,600	28	784	1,680
2	57	3,249	32	1,024	1,824
3	52	2,704	24	576	1,248
4	46	2,116	16	256	736
5	41	1,681	21	441	861
6	38	1,444	14	196	532

7	32	1,024	18	324	576
8	29	841	11	121	319
9	25	625	9	81	225
10	19	361	12	144	228
	$\Sigma X = 399$	$\Sigma X^2 = 17{,}645$	$\Sigma Y = 185$	$\Sigma Y^2 = 3{,}947$	$\Sigma XY = 8{,}229$

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

$$= \frac{10(8{,}229) - (399)(185)}{\sqrt{[10(17{,}645) - (399)^2][10(3{,}947) - (185)^2]}}$$

$$= \frac{82{,}290 - 73{,}815}{\sqrt{[176{,}450 - 159{,}201][39{,}470 - 34{,}225]}}$$

$$= \frac{8{,}475}{\sqrt{[17{,}249][5{,}245]}} = \frac{8{,}475}{\sqrt{90{,}471{,}005}} = \frac{8{,}475}{9{,}511.625} = .891 = .89$$

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} = \frac{10(8{,}229) - (399)(185)}{10(17{,}645) - (399)^2}$$

$$= \frac{82{,}290 - 73{,}815}{176{,}450 - 159{,}201} = \frac{8{,}475}{17{,}249} = 0.491 = 0.49$$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{399}{10} = 39.9$$

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{185}{10} = 18.5$$

$$a = \bar{Y} - b\bar{X} = 18.5 - (0.49)(39.9) = 18.5 - 19.551 = -1.051 = -1.05$$

$$\hat{Y} = -1.05 + 0.49X$$

$$\hat{Y} = -1.05 + 0.49X = -1.05 + 0.49(35) = -1.05 + 17.15 = 16.1$$

References

Huynh, A. C., Oakes, H., Shay, G. R., & McGregor, I. (2017). The wisdom in virtue: Pursuit of virtue predicts wise reasoning about personal conflicts. *Psychological Science, 28*(12), 1848-856. doi: 10.1177/095679761772621

Sokol, M. M. (1982). James McKeen Cattell and the failure of anthropometric mental testing, 1890-1901. In W. R. Woodward and M. G. Ash (Eds.). *The problematic science: Psychology in nineteenth-century thought.* New York: Praeger.