proprep™

Workbook

Table of Contents

Measures of Association

Linear Measure of Association (Pearson)

Questions

1) The following table displays the final marks at the end of the course and the number of absences for six students:

No. of Absences	Mark
2	80
1	90
0	90
2	70
3	70
4	50

a. Draw a scatter plot of the data. What can be concluded from the diagram about the type of connection between a student's number of absences and his mark? Which is the independent variable and which is the dependent variable?

b. Calculate the Pearson Measure of Association. Is the result consistent with your answer in section a?

c. Explain without calculations how the correlation coefficient would change if a student was added with four absences and a mark of 80.

2) A medical study examined whether there was a relationship between the level of hormone X in the patient's blood and the level of hormone Y.
The levels of these two hormones were recorded for five patients.
The following table displays the results obtained:

X	Y
10	12
14	15
15	15
18	17
20	21

a. What is the average level of each hormone?

b. What is the correlation coefficient of the two hormones? What is the significance of the result?

3) Let X be a family's income in thousands of dollars. Let Y be a family's spending in thousands of dollars. 20 families were selected, and the following results were obtained:

$$\sum_{i=1}^{20} X_i = 240 \qquad \sum_{i=1}^{20} Y_i = 200 \qquad \sum_{i-1}^{20} \left(X_i - \overline{X}\right)^2 = 76\,, \qquad \sum_{i-1}^{20} \left(Y_i - \overline{Y}\right)^2 = 76$$

$$\sum_{i=1}^{20} \left(X_i - \overline{X}\right) \cdot \left(Y_i - \overline{Y}\right) = 60.8$$

a. Calculate the linear measure of association between X and Y.
Which is the dependent variable?

b. What is the significance of the result obtained in section a?

4) 20 families were selected, and the following results were obtained:

$$\sum_{i=1}^{20} X_i = 240\,, \qquad \sum_{i=1}^{20} Y_i = 200\,, \qquad \sum_{i=1}^{20} X_i^{\,2} = 2960\,, \qquad \sum_{i=1}^{20} Y_i^2 = 2080\,, \qquad \sum_{i=1}^{20} X_I Y_i = 2464$$

Calculate the linear measure of association between X and Y.

5) In a certain academic institution, the mark of suitability is calculated as follows:
The average matriculation mark is multiplied by 3, and two points are then subtracted.
It is known that for 40 candidates, the standard deviation of the average matriculation mark was 2.
What is the correlation coefficient of the mark of suitability and the average matriculation mark?

6) The following is a list of statements.
State whether each statement is correct or incorrect, and explain:

a. A real estate agent converted apartment prices from dollars to English Pounds.
Assume that the pound-dollar exchange rate is $£3.50 / \$$.
If the real estate agent calculates the Pearson Measure of Association between the apartment price in pounds and the apartment price in dollars, he will get 1.

b. A series of data gave the results $s_x = s_y = 1$ and $\overline{X} = \overline{Y} = 6$.
The Pearson Measure of Association is therefore 1.

c. If the covariance of X and Y is 0, then the Pearson Measure of Association must be 0.

Multiple Choice Questions

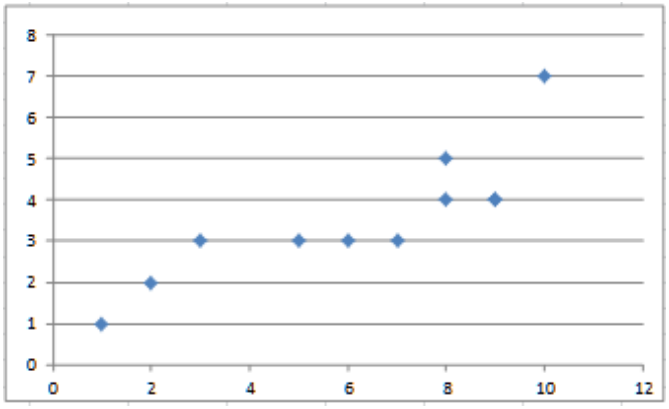
7) A negative correlation coefficient was found between examination marks in English and arithmetic.

a. This indicates that the marks on the exam in the class were negative.
b. As a student's mark in arithmetic decreases, his mark in English tends to decrease.
c. As a student's mark in arithmetic increases, his mark in English tends to decrease.
d. None of the above answers is correct.

8) 20 products were sampled, and their prices in dollars and Mexican Pesos were checked on a certain day (1 dollar was worth 4.20 Mexican Pesos).
What is the correlation coefficient between the price in dollars and the price in Pesos?

a. 1
b. 0
c. 4.2
d. The information is insufficient to answer the question.

9) The following is a dispersion diagram:



What will the correlation coefficient between the variables be?

a. 1
b. 0.85
c. 0.15
d. 0

Answer Key

1) a. In the recording. b. Yes, it is: $r_s = -0.9325$.

c. The correlation between X and Y will become weaker, and the correlation coefficient will equals almost zero.

2) a. $\bar{x} = 15.4$ $\bar{y} = 16$ b. $r_{xy} = 0.96$ - a very strong positive correlation.

3) a. $r = 0.8$; Y is dependent.

b. There is a partial positive linear correlation between income and spending.

4) $r = 0.8$

5) $r_{xy} = 1$

6) a. True. b. False. c. True.

7) c

8) a

9) b

Effect of Linear Transformation on Pearson Measure of Association

Questions

1) A test is composed of a quantitative section and a verbal section.
The correlation coefficient between the marks on the two sections is 0.9.

a. If all the marks in the verbal section are increased by 20%, what will be the correlation coefficient between the new verbal mark and the mark in the quantitative section, and between the mark in the old verbal section and the new verbal section?

b. We define a new variable was the distance between the mark on the verbal section and the maximum possible mark on the exam – 150.
Calculate the correlation coefficient between the mark on the verbal exam and W and between W and the mark on the quantitative section.

2) The correlation coefficient between income and spending of 10 families was calculated and found to be 0.7.
If the income of the entire population increases by 5% and it's spending increases by 7%, what will be the correlation coefficient between the new income and the new spending?

3) An ice cream manufacturing company conducts a study to examine the relationship between the ice cream packages sold in a day and the temperature on that day.
10 days were examined and a 0.85 was found.
The company begins each day with 150 packages of ice cream. In addition, the company wants the temperature to be expressed in Fahrenheit instead of Centigrade.
What is the correlation coefficient between the number of containers left at the end of the day and the temperature in Fahrenheit?
The relationship between Centigrade ($C°$) and Fahrenheit ($F°$) is given by the formula:

$$F = \frac{9}{5}C + 32$$

Select the correct answer:

a. 0.85
b. -0.85
c. 1
d. The information is insufficient to find the answer.

4) The correlation coefficient between X and Y is 0.4. All the values of X are multiplied by 2, and the new correlation coefficient between the variables will therefore be:

a. 0.8
b. 0.4
c. -0.4
d. The information is insufficient to find the answer.

Answer Key

1) a. Between the old and new marks on the verbal exam: 1

Between the new mark on the verbal exam and the mark on the quantitative exam: 0.9.

b. Between W and the mark on the verbal exam: -1.

Between W and the mark on the quantitative exam: -0.9.

2) $r_{x',y'} = r_{x,y} = 0.7$

3) b

4) b

Linear Regression

Questions

1) Let X be the income of a family in thousands of dollars and Y the spending of a family in thousands of dollars. 20 families were selected, and the following results were obtained:

$\sum_{i=1}^{20} X_i = 240$, $\sum_{i=1}^{20} Y_i = 200$

$\sum_{i-1}^{20} \left(X_i - \overline{X}\right)^2 = 76$, $\sum_{i-1}^{20} \left(Y_i - \overline{Y}\right)^2 = 76$, $\sum_{i=1}^{20} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 60.8$

a. Calculate the measure of linear relationship between X and Y. Which is the dependent variable?
b. Find the regression line for predicting the spending of a family on the basis of its income. Explain the significance of the parameters of the regression line.
c. The Smiths earns $\$15,000$. What is its expected spending?

2) Let X be the education of a person (in years).
Let Y be his income in thousands of dollars.
The following results were obtained in a study:

$S_x = 2, \quad S_y = 5, \quad \overline{X} = 14 \quad \overline{Y} = 8, \ \mathrm{cov}\left(X, Y\right) = 7.5$

a. Calculate the Pearson Measure of Association between education and income.
b. What is the expected income of a person with 12 years of education?
c. What is the predicted number of years for a person whose income is $\$10,000$?

3) A researcher wishes to study the relationship between the mark on a statistics exam and the number of hours spent studying for the exam. A sample of 100 students in the stats course who took the exam yielded the following results: the average mark of the students was 65 with a standard deviation of 27. The average number of study hours was 30 with a standard deviation of 18. The correlation coefficient between the mark and the hours of study was 0.8.

a. According to the regression equation, how much will one hour of study improve a student's mark on the exam?
b. According to the regression equation, what mark will a student who did not study for the exam receive?
c. What is the equation for the regression line used to predict the mark of a student according to the number of hours he studied?

4) Let X and Y be random variables, where the average of X is 1.5, X and Y both have a variance of 4, and the regression line of Y on the basis of X is $Y = -0.2X + 0.5$. Calculate the correlation coefficient between X and Y.

Answer Key

1) a. $r_{x,y} = 0.8$ b. $\hat{y} = 0.8X + 0.4$ c. $\$12.4\text{K}$

2) a. $r_{x,y} = 0.75$ b. $\hat{y}_{x=12} = \$4.25\text{K}$ c. $\hat{x}_{y=10} = 14.6$ years.

3) a. 1.2 marks per hour. b. 29 c. $\hat{y} = 1.2x + 29$

4) $r = -0.2$

Regression - Explained and Unexplained Variance

Questions

1) A positive relationship having a power of 0.7 was found between the area of an apartment and its price. Also given is the standard deviation of apartment prices - 200.

a. What percentage of the variance of apartment prices is explained by the area of the apartment?
b. What percentage of the variance of apartment prices is not explained by the area of the apartment?
c. What are the explained variance and the unexplained variance of the apartment prices?

2) The following is a list of statements.
Determine whether each statement is true or false and explain.

a. If the variance of error (the unexplained variance) is 0, then the Pearson correlation coefficient is 1.
b. If the Pearson correlation coefficient between two variables is 1, then the variance of error (the unexplained variance) is 0.
c. If the covariance of X and Y is 0, then the Pearson correlation coefficient is 0.

Multiple Choice Questions

3) $r^2 = 0.64$ was obtained for the connection between two variables, therefore:

a. Without an exception, as the values of one variable increase the other variable increases.
b. 64% of the variance of one variable is explained by the other variable.
c. The power of the connection between the two variables is 0.64.
d. Both answers are correct.

4) If r^2 is increased, what can be said?

a. The percentage of explained variance will decrease.
b. The percentage of explained variance will increase.
c. The percentage of explained variance will be unchanged.
d. The standard deviation will change.
e. Nothing can be concluded.

5) In the Introduction to Economics course, two exams are given in the course of a year: an exam at the end of the first semester (X) and an exam at the end of the second semester (Y).

When a regression line is constructed for the mark on the exam at the end of the second semester according to the mark on the exam at the end of the first semester, the variance of error is 80 and the variance of predictions is 20.

According to these data, the correlation coefficient between the exam at the end of the first semester and the exam at the end of the second semester is:

a. 0.44
b. -0.44
c. The power of the linear connection is 0.44, but its sign cannot be determined.
d. The correlation coefficient cannot be calculated.
e. 0.35

Answer Key

1) a. 49% b. 51% c. Explained variance: $\$19,600$; unexplained variance: $\$20,400$.

2) a. False. b. True. c. True.

3) b

4) b

5) c

Cramer's Measure of Association

Questions

6) The following table displays the results of a study, examining the relationship between gender and education.
The gender and education of each subject is recorded, and the results are as follows:
Is there a relationship between gender and education? Explain.

Gender \ Education	Low	High School	High
Man	120	40	20
Woman	20	20	80

7) 200 people are sampled, 60 of whom stated that they exercised regularly.
Of those who exercised regularly, 50 were found to be in good health.
Of those who did not exercise regularly, 90 were found to be in good health.

a. Construct a joint frequency table.
b. Is there a relationship between exercise and health? Calculate the value of this relationship according to Cramer's Measure of Association.

Answer Key

1) Yes, there is some connection between Gender and Education, according the result of Cramer Measure of Association $(r_c = 0.595)$.

2) a.

Exercise \ Health	Healthy	Not Healthy	$f(X)$
Yes	50	10	60
No	90	50	140
$f(Y)$	140	60	$n = 200$

b. A very low connection between them: $r_c = 0.19$.

The Phi Measure of Association

Summery

The phi measure of association is a short way of calculating Cramer's measure of association. This measure is relevant only when the joint frequency table is 2×2, meaning that the two variables are dichotomous.

The formula is: $\phi = \sqrt{\frac{(a\cdot d - b\cdot c)^2}{e\cdot f\cdot r\cdot k}}$

	Y_1	Y_2	Total
X_1	a	b	e
X_2	c	d	f
Total	R	k	

Example (Solution in the recording)

A factory works in two shifts: a day shift and a night shift.
300 products are sampled from the day shift and 200 from the night shift.
10 of the products from the day shift were faulty, and 150 of the products sampled at night were properly made.
Is the type of shift related to the quality of the product?

Questions

1) The following table displays the results of a study examining the relationship between gender and education.
The gender and education of each subject is recorded and shown below.
Can the phi measure of association be calculated in this case?
If so, calculate it.

Gender \ Education	Low	High School	High
Man	120	40	20
Woman	20	20	80

2) 200 people are sampled, 60 of whom stated that they exercised regularly.
Of those who exercised regularly, 50 were found to be in good health.
Of those who did not exercise regularly, 90 were found to be in good health.
Can the ϕ measure of association be calculated?
If so, calculate it and explain its significance.

Answer Key

1) No

2) Yes, $\theta = 0.19$.

Lambda Measure of Association

Questions

1) The following table displays the results of a study examining the relationship between gender (X) and level of education (Y):

a. Calculate the lambda measure of association for predicting education on the basis of gender.

b. Calculate the lambda measure of association for predicting gender on the basis of education.

Gender \ Education	Low	High School	High
Man	120	40	20
Woman	20	20	80

2) There are four neighborhoods in a city.
The economic situation of each family (Y) in each neighborhood (X) is examined.
The following table displays the joint frequency table obtained:
Calculate the lambda measures of association and explain the findings.

Neighborhood \ Economic Situation	Low	Medium	High
A	40		
B		70	
C		80	
D			40

Answer Key

1) a. $\lambda_{x|y} = 0.375$ b. $\lambda_{y|x} = 0.5$

2) a. $\lambda_{x|y} = 0.533$ b. $\lambda_{y|x} = 1$

Spearman Measure of Association

Questions

1) Two judges assigned marks to the contestants in a beauty contest:
Is there a connection between the evaluations of the two judges? Explain.

Contestant No.	1	2	3	4	5	6	7
Mark by Judge A	7	8	6	8	9	5	6
Mark by Judge B	8	8	7	8	9	5	7

2) A manager wanted to examine whether there was an association between the motivation of its employees and the number of absences during a month of work.
The following results were obtained:
Is there a connection between the motivation of an employee and the number absences?
Calculate using the most appropriate measure of association and explain.

Number of Absences	Motivation
0	High
4	Low
2	Medium
5	Low
1	High

3) If $r_S = 1$, this means that the values of X are always equal to the values of Y.
Is this statement true? Explain.

Answer Key

1) Yes, there is a very high correlation between them: $r_s = 0.973$.

2) Yes, there is a strong negative correlation: $r_s = -0.85$.

3) Incorrect.

Eta Measure of Association

Background

The Eta measure of association is asymmetrical, i.e. there are two measures, in fact:

$\eta_{y|x}$ - Eta of Y given X – Where Y is a Ratio or an Interval variable, and X may be any type of variable.

$\eta_{x|y}$ - Eta of X given Y – Where X is a Ratio or an Interval variable, and Y may be any type of variable.

The Eta measure of association measures the strength of the relationship between X and Y, and not its direction (positive \ negative).

As such, the values that η can have are between 0 and 1, i.e. $0 \le \eta \le 1$.

Example (solution in the recording)

The following is a joint frequency table of students, where:
X is Gender, and Y is the number of courses to which a student registered in the current semester.

X \ Y	1	2	3	$f(x)$
Male	5	5	5	15
Female	10	5	0	15
$f(y)$	15	10	5	30

a. Which measure of Eta can be calculated in the above tables: $\eta_{y|x}$ or $\eta_{x|y}$?
b. Calculate the value of the relevant Eta measure of association.

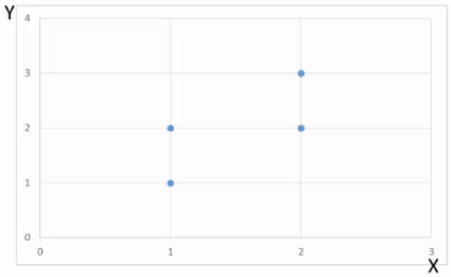
Questions

1) A survey was conducted to check the number of Tv sets per family, X, according to geographical area in the state of New York, Y.
The following is the joint frequency table displaying the results:

X \ Y	Mid	North	South	$f(x)$
0	5	15	0	20
1	5	5	20	30
2	5	0	0	5
$f(y)$	15	20	20	55

a. What Eta measure of association can be calculated?
b. Calculate the value of Eta measure.

2) The following is a scatter plot of four observations:
a. Construct the joint frequency table for the data.
b. Calculate the values for $\eta_{x|y}$ and $\eta_{y|x}$.



Answer Key

1) a. $\eta_{x|y}$ b. $\eta_{x|y} = 0.585$

2) a.

X \ Y	1	2	3	f(X)
1	1	1	0	2
2	0	1	1	2
f(Y)	1	2	1	4

b. $\eta_{y|x} = 0.707, \quad \eta_{x|y} = 0.707$