

Correlation regression and chi square

[Science](#)



1). what information is provided by the numerical value of the pearson correlation?

Pearson correlation is a measure of the degree of association between two variables. The degree of association becomes stronger as it approaches the values 1 and -1. It also suggest something about the direction of the association, whether it is downward or upward.

2). in the following data, there are three scores (x, y, and z) for each of the n = 5 individuals.

a. Sketch a graph showing the relationship between X and Y, compute the Pearson correlation between X and Y.

$$\begin{aligned} \text{Pearson correlation } (r) &= (\sum XY - (\sum X \sum Y)/N) / (\sqrt{(\sum X^2 - (\sum X)^2/N)(\sum Y^2 - (\sum Y)^2/N)}) \\ &= (36 - (10 \cdot 15)/5) / (\sqrt{(30 - 100/5)(55 - (225/5))}) \\ &= (36 - 30) / (\sqrt{(30 - 20)(55 - 45)}) \\ &= 6 / (\sqrt{10 \cdot 10}) \\ &= 6/10 \\ &= 0.6 \end{aligned}$$

b. Sketch a graph showing the relationship between Y and Z. compose the pearson correlation between Y and Z.

$$\begin{aligned} \text{Pearson correlation } (r) &= (\sum YZ - (\sum Y \sum Z)/N) / (\sqrt{(\sum Z^2 - (\sum Z)^2/N)(\sum Y^2 - (\sum Y)^2/N)}) \\ &= (66 - (20 \cdot 15)/5) / (\sqrt{(90 - (400)/5)(55 - (225/5))}) \\ &= (66 - 60) / (\sqrt{(90 - 80)(55 - 45)}) \\ &= 6 / (\sqrt{10 \cdot 10}) \end{aligned}$$

$$= 6/10$$

$$= 0.6$$

c. Given the result of parts A and B, what would you predict for the correlation between X and Z?

The Pearson correlation in parts A and B are equal. Then we would predict that the correlation between X and Z are the same with parts A and B.

d. Sketch a graph showing the relationship between X and Z. compute the Pearson correlation for these data.

$$\text{Pearson correlation } (r) = (\sum XZ - (\sum X)(\sum Z)/N) / (\sqrt{(\sum X^2 - (\sum X)^2/N)(\sum Z^2 - (\sum Z)^2/N)})$$

$$= (38 - (10 \cdot 20)/5) / (\sqrt{(30 - 100/5)(90 - (400)/5)})$$

$$= (38 - 40) / (\sqrt{(30 - 20)(90 - 80)})$$

$$= -2 / (\sqrt{10 \cdot 10})$$

$$= -2/10$$

$$= -0.2$$

e. What general conclusion can you make concerning relationship among correlations? If X is related to Y and Y is related to Z, does this necessarily mean that X is related to Z?

Transitivity is not applicable relationship among correlations. When one speaks of transitivity, this only means that for any variables a, b and c, if a = b and b = c then a = c. This definition is not applicable to correlations of variables. One cannot predict correlation between two variables based from correlations of other set of variables. For example, X is related to Y, and Y is related to Z. Then, one cannot predict or say that X is related to Z based on

the given correlations.

3) Sketch a graph showing the line for the equation $Y = 2x - 3$ on the same graph, show the line for $Y = -2x + 8$

4) A set of scores produces a regression equation of $F = 7x - 2$. Use the equation to find the predicted value of Y for each of the following X scores: 0, 2, 5, 8, 10

$$F = 7x - 2$$

$$F(0) = 7(0) - 2$$

$$= 0 - 2 = -2$$

$$F(2) = 7(2) - 2$$

$$= 14 - 2 = 12$$

$$F(5) = 7(5) - 2$$

$$= 35 - 2 = 33$$

$$F(8) = 7(8) - 2$$

$$= 56 - 2 = 54$$

$$F(10) = 7(10) - 2$$

$$= 70 - 2 = 68$$

5) For the following data:

a. find the regression equation for predicting Y from X

$$? = bX + A$$

$b = r(Sy/Sx)$, where r is the pearson correlation, Sy and Sx are the standard deviation of Y and X .

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

$$= \frac{(170 - (25 \cdot 30)/5)}{(\sqrt{(135 - (625)/5)(346 - (900)/5)}}$$

$$= \frac{(170 - 150)}{(\sqrt{(135 - 125)(346 - 180)}}$$

$$= \frac{(20)}{\sqrt{(10)(166)}}$$

$$= \frac{20}{\sqrt{1660}}$$

$$= \frac{20}{40.7431}$$

$$= 0.4909$$

$$b = 0.4909(6.4420/1.5811)$$

$$= 2.0001$$

$$A = My - bMx$$

$$= 6 - 2.0001(5)$$

$$= 6 - 10.0005$$

$$= -4.0005$$

$$Y = 2.0001X - 4.0005$$

b. use the regression equation to find a predicted Y for each X

$$Y = 2.0001X - 4.0005$$

X

?

7

10.0002

5

6. 0000

6

8. 0001

3

1. 9998

4

3. 9999

c. find the difference between the actual Y value and the predicted Y value for each individual, square the differences, and add the squared values to obtain SS_{residual} .

?

Y

Y

-

?

 $(Y - ?)^2$

10. 0002

16

5. 9998

35. 9976

6. 0000

2

-4.0000

16.0000

8.0001

1

-7.0001

49.0014

1.9998

2

-0.0002

0.0000

3.9999

9

5.0001

25.0010

$$SS_{\text{residual}} = \sum (Y - \hat{Y})^2 = 126$$

d. calculate the Pearson correlation for these data. Use r^2 and $SS_{\hat{Y}}$ to compute SS_{residual} , with equation 15.13. You should obtain the same value as in part c.

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

$$= \frac{(170 - (25 \cdot 30)/5)}{\sqrt{((135 - (625)/5)(346 - (900)/5))}}$$

$$= \frac{(170 - 150)}{\sqrt{(135 - 125)(346 - 180)}}$$

$$= \frac{(20)}{\sqrt{(10)(166)}}$$

$$= 20 / \sqrt{1660}$$

$$= 20 / 40.7431$$

$$= 0.4909$$

$$SS_y = \sum (Y - \bar{Y})^2$$

$$= 166$$

$$SS_{\text{residual}} = (1 - r^2)SS_y$$

$$= (1 - 0.4909^2)(166)$$

$$= (1 - 0.2410)(166)$$

$$= (0.759)(166)$$

$$= 125.994 \quad \text{or} \quad 126$$

Chapter 16.

6) A professor noticed that the representatives on the college student government consist of 31 males and only 9 females. The general college population on the other hand, consists of 55% females and 45% males. Is the gender distribution for student government representatives significantly different from the distribution for the population? Test at the 0.5 level of significance.

Null Hypothesis: The gender distribution for student government representatives is not significantly different from the distribution for the population.

Alternative: The gender distribution for student government representatives is significantly different from the distribution for the population.

Observed

Expected

Male

31

18

Female

9

22

χ^2 (computed) = $\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

$$= \frac{(31-18)^2}{18} + \frac{(9-22)^2}{22}$$

$$= 9.3889 + 7.6818$$

$$= 17.0707$$

$$\chi^2$$
 (critical) = 3.841

Decision rule, reject null hypothesis if $\chi^2 \geq 3.841$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $\chi^2 = 17.0707 \geq 3.841$, then we reject the null hypothesis.

We are 95% confident that the gender distribution for student government representatives is significantly different from the distribution for the population.

7) Data from the department of motor vehicles indicate that 80% of all licensed drivers are older than 25.

a. In a sample of $n = 50$ people who recently received speeding tickets, 32

were older than 25 years and the other 18 were age 25 or younger. Is the age distribution for this sample significantly different from the distribution for the population licensed divers? Use $\alpha = .05$.

Null Hypothesis: The age distribution for this sample is not significantly different from the distribution for the population licensed divers.

Alternative: The age distribution for this sample is not significantly different from the distribution for the population licensed divers.

Licensed	Drivers
----------	---------

Observed	
----------	--

Expected	
----------	--

> 25 years old	
----------------	--

32	
----	--

40	
----	--

? 25 years old	
----------------	--

18	
----	--

10	
----	--

$$\chi^2 (\text{computed}) = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= \frac{(32-40)^2}{40} + \frac{(18-10)^2}{10}$$

$$= 1.6 + 6.4$$

$$= 8$$

$$\chi^2 (\text{critical}) = 3.841$$

Decision rule, reject null hypothesis if $\chi^2 \geq 3.841$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $\chi^2 = 8 \geq 3.841$, then we reject the null hypothesis.

We are 95% confident that the age distribution for this sample is significantly different from the distribution for the population licensed drivers.

b. In a sample of $n = 50$ people who recently received parking tickets. 38 were older than 25 years and the other 12 were age 25 or younger. Is the age distribution for this sample significantly different from the distribution for the population of licensed drivers? Use $\alpha = 0.05$.

Null Hypothesis: The age distribution for this sample is not significantly different from the distribution for the population licensed drivers.

Alternative: The age distribution for this sample is not significantly different from the distribution for the population licensed drivers.

Licensed

Drivers

Observed

Expected

> 25 years old

38

40

≤ 25 years old

12

10

χ^2 (computed) = $\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

$$= (38-40)^2/40 + (12-10)^2/10$$

$$= 0.1 + 0.4$$

$$= 0.5$$

$$\chi^2 (\text{critical}) = 3.841$$

Decision rule, reject null hypothesis if $\chi^2 \geq 3.841$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $\chi^2 = 0.5 < 3.841$, then we fail to reject the null hypothesis.

We are 95% confident that the age distribution for this sample is not significantly different from the distribution for the population licensed divers.

8) A researcher obtained a random sample of $n = 60$ students to determine whether there were any significant preferences among three leading brands of colas. Each student tasted all the brands and then selected his or her favorite. The resulting frequency distribution is as follows:

BRAND	A	BRAND	B	BRAND	C
	28		14		18

Are the data sufficient to indicate any preferences among the three brands?

Test with $\alpha = 0.05$.

Null hypothesis: There are no significant preferences among three leading brands of colas.

Alternative: There are significant preferences among three leading brands of colas.

Brand

Observed

Expected

A

28

20

B

14

20

C

18

20

χ^2 (computed) = $\sum (\text{observed} - \text{expected})^2 / (\text{expected})$

$$= (28-20)^2/20 + (14-20)^2/20 + (18-20)^2/20$$

$$= 3.2 + 1.8 + 0.2$$

$$= 5.2$$

χ^2 (critical) = 5.991

Decision rule, reject null hypothesis if $\chi^2 \geq 5.991$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $\chi^2 = 5.2 < 5.991$, then we fail to reject the null hypothesis.

We are 95% confident that there are no significant preferences among three leading brands of colas.

<https://assignbuster.com/correlation-regression-and-chi-square/>

9) A social psychologist suspects that people who serve on juries tends to be much older than citizens in the general population. Jurors are selected from the list of registered voters, so the ages for jurors should have the same distribution as the ages for voters. The psychologist obtains voters registration records and finds that 20% of registered voters are between 18 and 29 years old, and 35% are age 50 or older. The psychologist also monitors jury composition over several weeks and observes the following distribution of ages for actual juries.

Age categories	for jurors
18-29	12
30-49	36
50 and over	32

Are the data sufficient to conclude that the age distribution for the jurors is significantly different from the distribution for the population of registered voters? Test with ? = .05.

Null hypothesis: The age distribution for the jurors is not significantly different from the distribution for the population of registered voters.

Alternative: The age distribution for the jurors is not significantly different from the distribution for the population of registered voters

Age

Observed

Expected

18-29

12

16

30-49

36

36

50 and over

32

28

χ^2 (computed) = $\sum (\text{observed} - \text{expected})^2 / (\text{expected})$

$= (12-16)^2/16 + (36-36)^2/36 + (32-28)^2/28$

$= 1 + 0 + 0.5714$

$= 1.5714$

χ^2 (critical) = 5.991

Decision rule, reject null hypothesis if $\chi^2 \geq 5.991$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $\chi^2 = 1.5714 < 5.991$. Thus, we fail to reject the null hypothesis.

We are 95% confident that the age distribution for the jurors is not significantly different from the distribution for the population of registered voters.

10) A psychology professor is trying to decide which test-book to use for next year's introductory class. To help make the decision, the professor asks the current students to make to review three texts and identify which one they prefer. The distribution preference for the current class is as follows:

Book	1	book	2	book	3
------	---	------	---	------	---

52

41

27

Do the data indicate any significant preferences among the three books?

Test with $\alpha = .05$.

Null hypothesis: There are no significant preferences among the three books.

Alternative: There are significant differences among the three books.

Book

Observed

Expected

1

52

40

2

41

40

3

27

40

χ^2 (computed) = $\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

= $\frac{(52-40)^2}{40} + \frac{(41-40)^2}{40} + \frac{(27-40)^2}{40}$

= $3.6 + 0.025 + 4.225$

= 7.85

χ^2 (critical) = 5.991

<https://assignbuster.com/correlation-regression-and-chi-square/>

Decision rule, reject null hypothesis if $X^2 \geq 5.991$. Otherwise, fail to reject the null hypothesis.

At $\alpha = 0.05$, $X^2 = 7.85 \geq 5.991$. Thus, we reject the null hypothesis.

We are 95% confident that there are significant differences among the three books.