Inferential statistics

Section No. (5)

Chapter (3): Correlation Coefficient

Presented by

Ghidaa abobakr Hamouda

Assistant Lecturer in Department of
Insurance, Statistics and Mathematics

Choose the correct answer

A researcher at the Institute of Genetic Engineering at Sadat City University wants to know if there is an association between eye color and gender. So he surveyed 70 individuals and obtained the following results:

	Blue	Green	Brown	Total
Male	9	13	15	**37**
Female	11	8	14	**33**
Total	**20**	**21**	**29**	**70**

1) The appropriate correlation coefficient between two variables

a) Cramer	b) spearman	c) phi	d) Pearson

2) The correlation coefficient between two variables =====

$$\varphi = \sqrt{\frac{x^2}{n}}$$

$$x^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$E_{ij} = \frac{total\ of\ row * total\ of\ colum}{total}$$

$$E_{11} = \frac{37 * 20}{70} = 10.57, E_{12} = \frac{37 * 21}{70} = 11.1, E_{13} = \frac{37 * 29}{70} = 15.3$$

$$E_{21} = \frac{33 * 20}{70} = 9.43, E_{22} = \frac{33 * 21}{70} = 9.9, E_{23} = \frac{33 * 29}{70} = 13.67$$

$$x^2 = \frac{(9 - 10.57)^2}{10.57} + \frac{(13 - 11.1)^2}{11.1} + \frac{(15 - 15.3)^2}{15.3} + \frac{(11 - 9.43)^2}{9.43} + \frac{(8 - 9.9)^2}{9.9} + \frac{(14 - 13.67)^2}{13.67} = 1.991$$

$$\varphi = \sqrt{\frac{x^2}{n}} = \sqrt{\frac{1.991}{70}} =$$

a) 0.08	b) 0.87	c) 0.97	d) 0.169

3) The strength of the relation between two variables

a) weak	b) strong	c) intermediate	d) No relation

4) State the null and alternate hypothesis

H0 : there is no relationship in the population. H1: there is a relationship in the population.	H0: $\rho_s \neq 0$ H1: $\rho_s = 0$	H0: $\rho_s \leq 0$ H1: $\rho_s > 0$	H0: $\rho_s \geq 0$ H1: $\rho_s < 0$

5) Choose the appropriate test to test the Significance of Association Between Two Variables

a) x^2 – distribution	b) t – distribution	c) F-distribution	d) Z-distribution

6) The value of the test statistics

$$x^2 = \sum_i \sum_j \frac{(o_{ij} - E_{ij})^2}{E_{ij}}$$

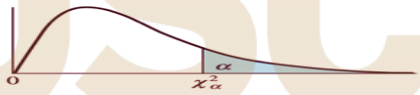
$$x^2 = \frac{(9-10.57)^2}{10.57} + \frac{(13-11.1)^2}{11.1} + \frac{(15-15.3)^2}{15.3} + \frac{(11-9.43)^2}{9.43} + \frac{(8-9.9)^2}{9.9} + \frac{(14-13.67)^2}{13.67} = 1.991$$

a) 0.095	b) 0.5	c) 0.95	d) 1.991

7) The critical value is

$$df = (r-1) * (c-1) = (2-1) * (3-1) = 2, \alpha = 0.05$$

(Chi-square table) χ^2 جدول توزيع مربع كاي



Degree of Freedom (df)	Right-Tail Area				
	0.10	0.05	0.02	0.01	0.001
1	2.7055	3.8415	5.4119	6.6349	10.8276
2	4.6052	5.9915	7.8240	9.2103	13.8155
3	6.2514	7.8147	9.8374	11.3449	16.2662
4	7.7794	9.4877	11.6678	13.2767	18.4668
5	9.2364	11.0705	13.3882	15.0863	20.5150

a) 0.095	b) 0.5	c) 0.95	d) 5.9915

8) The decision rule is

a) a significant relationship between the two variables	b) don't reject H1	c) no significant relationship between the two variables

One of the candidates for the elections wants to know if there is an association between the level of education and political party preference. A random sample was selected of 120 persons and the data was organized in the following contingency table:

	higher	middle	Illiterate	total
democrat	(18)23	(8.6)4	(3.3)3	30
independent	(5.4)2	(2.6)6	(1)1	9
Republican	(3.6)2	(1.7)3	(0.6)1	6
total	27	13	5	45

9) The appropriate correlation coefficient between two variables

a) Cramer	b) spearman	c) phi	d) Pearson

10) The correlation coefficient between two variables =

$$v = \sqrt{\frac{x^2}{n * \min(r-1), (c-1)}}$$

$$x^2 = \frac{(23-18)^2}{18} + \frac{(4-8.6)^2}{8.6} + \frac{(3-3.3)^2}{3.3} + \frac{(2-5.4)^2}{5.4} + \frac{(6-2.6)^2}{2.6} + \frac{(1-1)^2}{1} + \frac{(2-3.6)^2}{3.6} + \frac{(3-1.7)^2}{1.7} + \frac{(1-0.6)^2}{0.6} = 12.43$$

$$v = \sqrt{\frac{12.43}{45 * 2}} = 0.37$$

a) 0.08	b) - 0.97	c) 0.37	d) 0.882

11) The strength of the relation between two variables

a) weak	b) strong	c) intermediate	d) No relation

12) State the null and alternate hypothesis

H0 : there is no relationship in the population. H1: there is a relationship in the population.	H0: $\rho_s \neq 0$ H1: $\rho_s = 0$	H0: $\rho_s \leq 0$ H1: $\rho_s > 0$	H0: $\rho_s \geq 0$ H1: $\rho_s < 0$

13) The value of the test statistics

$$x^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$x^2 = \frac{(23-18)^2}{18} + \frac{(4-8.6)^2}{8.6} + \frac{(3-3.3)^2}{3.3} + \frac{(2-5.4)^2}{5.4} + \frac{(6-2.6)^2}{2.6} + \frac{(1-1)^2}{1} + \frac{(2-3.6)^2}{3.6} + \frac{(3-1.7)^2}{1.7} + \frac{(1-0.6)^2}{0.6} = 12.43$$

a) 0.095	b) 0.5	c) 0.95	d) 12.43

14) The critical value is

$$df = (r-1) * (c-1) = (3-1) * (3-1) = 4, \alpha = 0.05$$

(Chi-square table) جدول توزيع مربع كاي χ^2

Degree of Freedom (df)	Right-Tail Area				
	0.10	0.05	0.02	0.01	0.001
1	2.7055	3.84[illegible]	5.4119	6.6349	10.8276
2	4.6052	5.99[illegible]	7.8240	9.2103	13.8155
3	6.2514	[illegible]	9.8374	11.3449	16.2662
4	[illegible]	9.4877	11.6678	13.2767	18.4668
5	9.2364	[illegible]	13.3882	15.0863	20.5150

a) 0.095	b) 0.5	c) 0.95	d) 9.48

15) The decision rule is

a) a significant relationship between the two variables	b) don't reject H1	c) no significant relationship between the two variables

- One of the car dealers wants to know if there is a relationship between the car brand and the gender and he selected a sample of 100 people and got the following results Test is there a correlation at a level of 0.05

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
l car brand and the gender	10	100.0%	0	0.0%	10	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.333[a]	3	.149
Likelihood Ratio	7.271	3	.064
Linear-by-Linear Association	1.552	1	.213
N of Valid Cases	10		

a. 8 cells (100.0%) have expected count less than 5. The minimum expected count is 1.00.

car brand and the gender **Crosstabulation**

Count

		car brand				Total
		kia	tereos	toyota	lanser	
ender	nale	0	2	1	2	5
	emale	3	0	1	1	5
Total		3	2	2	3	10

Symmetric Measures

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Nominal by Nominal	Phi	.730			.149
	Cramer's V	.630			.149
	Contingency Coefficient	.590			.149
Interval by Interval	Pearson's R	-.415	.265	-1.291	.233[c]
Ordinal by Ordinal	Spearman Correlation	-.431	.291	-1.351	.214[c]
N of Valid Cases		10			

16) The appropriate correlation coefficient between two variables

a) Cramer	b) spearman	c) phi	d) Pearson

17) The correlation coefficient between two variables =

a) 0.63	b) - 0.431	c) 0.73	d) - 0.415

18) The strength of the relation between two variables

a) weak	b) strong	c) intermediate	d) No relation

19) p-value

a) 0.241	b) 0.233	c) 0.149	d) - 0.415

20) The decision rule is

$$p\text{-value} > 0.05$$

a) a significant relationship between the two variables	b) don't reject H1	c) no significant relationship between the two variables