Correlation and chi²

cdu tech SAARLAND UNIVERSITY

Lara Kataja

Room 1.13

l.kataja@edutech.uni-saarland.de

chi²

- Non-parametric test, for data on nominal (or ordinal) level
- Testing a possible association between two categorical variables
- In our case: We test if a categorical variable (e.g. "School Type") is equally distributed on our experimental groups: "Do group A and group B have the same proportion of different school types?"

chi² in SPSS

<u>A</u> naly	ze Direct <u>M</u> arketing	<u>G</u> raphs	<u>U</u> ti	ities	E <u>x</u> ter	nsions	<u>W</u> in	dow
F	Re <u>p</u> orts		•					A
C	escriptive Statistics		•	123 Er	equer	icies		
Т	a <u>b</u> les		۶.		escrip	tives		
C	Co <u>m</u> pare Means		۴.	A F	nlore			-
<u>c</u>	eneral Linear Model		•		roceta			
G	Generali <u>z</u> ed Linear Mode	ls	•			US		
N	li <u>x</u> ed Models		•	± 10	JRF AI	nalysis		
<u>c</u>	orrelate		•	<u>172</u> <u>R</u> a	atio			
E	egression		•	🞓 P-	P Plot	S		
L	. <u>og</u> linear		•	🤧 <u>Q</u> -	-Q Plo	ts		
N	leural Net <u>w</u> orks		•	Kr	ippen	dorff's al	pha	
c	lassi <u>f</u> y		۲	-		Problem	. 1 in	
	imension Reduction		•	0		Problem	- 1 in	P
s	c <u>a</u> le		•	0		Problem	n 1 in	n P
<u>N</u>	lonparametric Tests		•	0		Problem	n 2 in	n P
F	orecasting		•	0		Problen	n 2 in	n P.
<u>s</u>	urvival		•	0		Problen	n 2 in	n P.
N	l <u>u</u> ltiple Response		•	0		Problen	n 2 in	n P.
53 M	lissing Value Anal <u>v</u> sis			0		Problen	n 3 in	n P.
N	Iultiple Imputation		•	0		Problen	n 3 in	n P.
c	Complex Samples		•	0		Problen	n 3 in	n P.
₿₽	imulation			0		Problen	n 4 in	n P.
	uality Control		•	0		Problem	n 4 in	n P.
	OC Curve			0		Problem	n 4 in	n P.
	natial and Temporal Mo	delina		0		Problem	n 4 in	n P.
9	para and remporativo	denng		0		Problem	a 4 in	n P



Enter your two categorical variables here

H0: F1_FAC is independent from School Type H1: F1_FAC is not independent from School Type

chi² in SPSS

Crosstabs: Statistics ×	ta Crosstabs: Cell Display	×
Chi-square Correlations Contingency coefficient Phi and Cramer's V Lambba Contail Kendall's tau-b	Counts Observed Expected Hide small counts Less than 5	e column proportions p-values (<u>B</u> onferroni method)
Uncertainty coefficient Kendall's tau-c Nominal by Interval Kappa Eta Risk McNemar Cochran's and Mantel-Naenszel statistics Test common odds ratio rquals: 1	Percentages Residuals - Row Unstand Column Standard Total Adjusted Noninteger Weights Round cell counts Round cell counts Round case weights Truncate cell counts Truncate case weights	ardized ized standardized
Continue Cancel Help	O No adjust <u>m</u> ents	Help
Select "chi-square" ar and "expected" counts	nd "observed"	edu
		Lech

chi² in SPSS

F2_AGG * School_Type Crosstabulation

		School_Type				
			below- average	average	above- average	Total
F2_AGG	NO AGG	Count	14	40	28	82
		Expected Count	15,2	39,5	27,3	82,0
	AGG	Count	16	38	26	80
		Expected Count	14,8	38,5	26,7	80,0
Total		Count	30	78	54	162
		Expected Count	30,0	78,0	54,0	162,0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	,234 ^a	2	,890
Likelihood Ratio	,234	2	,890
Linear-by-Linear Association	,169	1	,681
N of Valid Cases	162		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 14,81.

There was no significant association between F2_AGG and School_Type: X^2 (2) = .234, p = .890

Fisher's Exact Test

• Like chi², but for small sample sizes (if the expected count of a cell is < 5)

F2_AGG * School_Type Crosstabulation

			School_Type			
			below- average	average	Total	
F2_AGG	NO AGG	Count	3	15	18	
		Expected Count	3,1	14,9	18,0	
	AGG	Count	3	14	17	
		Expected Count	2,9	14,1	17,0	
Total		Count	6	29	35	
		Expected Count	6,0	29,0	35,0	

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)	
Pearson Chi-Square	,006 ^a	1	,939			
Continuity Correction ^b	,000,	1	1,000			
Likelihood Ratio	006	1	939			
Fisher's Exact Test				1,000	,642	
Linear-by-Linear Association	,006	1	,940			
N of Valid Cases	35					
a. 2 cells (50,0%) have expected count less than 5. The minimum expected count is 2,91.						
h. Computed only for a 2x2 table						

There was no significant association between F1_FAC and Gender: p = 1.000 (Fisher's Exact Test, two-sided)

Spearman- vs. Pearson correlation

- Pearson correlation:
 - Normal distributed variables
 - Linear relationship
 - Interval scale
- Spearman / Kendall:
 - Monotonous relationship
 - Calculates with the ranks of the values
 - Ordinal scale

Interpretation

- Correlation coefficient r
 - Ranges from -1 to +1
 - -1 = perfect negative relationship ("the more X, the less Y")
 - +1 = perfect positive relationship ("the more X, the more Y")
 - $\square 0 = no relationship$
- ♦ If $p \le .05 \rightarrow Correlation$ is significant

Reporting correlations

"There was a significant correlation of X and Y with r(964) = .14, p = .000"

df = n-2) Correlations

		AGE OF RESPONDEN T	HOURS PER DAY WATCHING TV
AGE OF RESPONDENT	Pearson Correlation	1	.139**
	Sig. (2-tailed)		.000
2	N	1491	966
HOURS PER DAY	Pearson Correlation	.139**	1
WATCHINGTV	Sig. (2-tailed)	.000	
	N	966	973

**. Correlation is significant at the 0.01 level (2-tailed).

Interpretation of the correlation coefficient

- □ Small: .10 / -.10
- ^D Medium: .30 / -.30
- Large: .50 / -.50

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum.

Correlation and Causation (?)



cdi





Guess the correlation

http://guessthecorrelation.com/

