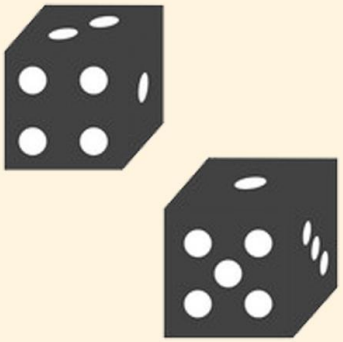


Biostatistics

Doctor 2017 | Medicine | JU

Number >>	10
Doctor	Hamza Alduradi
Done By	Nael Jadallah
CorrectedBy	...

Chi-squar X^2

- It is a number that you can calculate using certain formula
- Helps you take decision regarding to your hypothesis
- Helps you at the end of calculation to decide either to keep or reject the null hypothesis (determine if there is an association between dependent and in-dependent variables)
- We use it when we have dependent variable and in-dependent variable that are both nominal

Example:

Variable 1 → Do you belong to Democrat party or Republican Party?

Variable 2 → Are you favor or oppose of using weapons?

Also we may add neutral to variable 2

Research question → is there an association between your political party and your position of using weapons.

We use chi-square because both variables are nominal

Important condition:

1) Both variable must be nominal
2) Groups must be independence
3) Sample must contain at least 20 person
   Usually hundreds or thousands
4) Each cell must contain at least 5 observations
   Example

	Favor	Oppose
Democrat	IIII IIII II	IIII III
republican	IIII III	IIII IIII IIII I

1) we have more than 20 person
2) we have more than observation in each cell
So we can use chi-square

	Favor	Oppose
Democrat	II	IIII IIII III
Republican	IIII III	IIII IIII IIII II

1) we have more than 20 person

2) we don't have more than observation in all cell

So we can't use chi-square

	Favor	Oppose
Democrat	IIII	IIII II
Republican	II	IIII I

1) we don't have more than 20 person

2) we don't have more than observation in all cell

We can't use chi-square

How to calculate chi-square?

- We can calculate chi-square by this formula

$$X^2 = \frac{\Sigma (O - E)^2}{E}$$

X^2 = The value of chi square
O = The observed value
E = The expected value
$\Sigma (O - E)^2$ = all the values of (O – E) squared then added together

- observed value : is the number that i Collect from people

- Expected value :

$$\text{Expected frequency} = \frac{\text{row total} \times \text{column total}}{\text{Grand total}}$$

	Favor	Neutral	Oppose	f_{row}
Democrat	1 10	2 10	3 30	A 50
Republican	4 15	5 15	6 10	40
f_{column}	B 25	25	40	n = 90

So if I want to calculate the expected value for who are democrat and favor for weapons use

A*B/n → 25*50/90 = 13.9 (for cell 1) then I do it to all the 6 cells

So chi-square is →

$$\chi^2 = \underset{1}{\frac{(10-13.89)^2}{13.89}} + \underset{2}{\frac{(10-13.89)^2}{13.89}} + \underset{3}{\frac{(30-22.2)^2}{22.2}} + \underset{4}{\frac{(15-11.11)^2}{11.11}} + \underset{5}{\frac{(15-11.11)^2}{11.11}} + \underset{6}{\frac{(10-17.8)^2}{17.8}}$$

So the calculated chi-square is 11.03

Next step I use the calculated chi-square in a process of 6 step and at the end I can answer my research question.

Steps in test hypothesis

For the previous example (party and weapons)

First step: determine the appropriate test

1. Party Membership (2 levels) and Nominal

2. Voting Preference (3 levels) and Nominal

So I can use chi-square

Second step: establish level of significance

Usually we want our alpha to be 0.05

This mean that my result at the end (95% confidence) so I can say I am 95% confidence that there is association or there is no association

NOTE: some people may choose another alpha

Third step: determine the hypothesis

Write down your hypothesis

Ho (null hypothesis): There is no difference between Democrat & Republican in their opinion on gun control issue in the American population

Ha (my hypothesis): There is an association between responses to the gun control survey and the party membership in the American population.

NOTE: we should determine the population level

Fourth step: calculating the statistics

We calculate the chi-square as we do

Calculated chi-square = 11.03

Fifth step: determine degree of freedom

$$Df = (Row - 1)(Column - 1)$$

In the example, we have 2 rows and 3 column

So the Df is → (2 – 1) (3 – 1) = 2

NOTE: it is an important number to calculate in any chi-square test

So up to now we have 3 thing

1) Alpha = 0.05 2) Df = 2 3) chi-square = 11.03

Then use this number in step number 6

Sixth step: compare computed test statistic against a tabled/critical value

In this step I compare my calculated chi-square with critical chi-square

If calculated chi-square > critical chi-square then → null hypothesis is rejected

If calculated chi-square < critical chi-square then → null hypothesis is accepted

There is no scenario they are become equal (if they equal reject null hypothesis)

Table 5.3. Chi-square value

Degrees of Freedom	Probability						Non-significant		Signi-ficant	Highly significant
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21

- Sometime instead of write degree of freedom they write (r)
- Critical chi-square is the number were alpha and Df meet each other
- Don’t memorize the table it will be given for you in the exam

	$P(X \leq x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21

In our example:

Df (r) → 2
Alpha → 0.05
So the critical chi-square → 5.991

So now, I have

1) Calculated chi-square → 11.03 2) critical chi-square → 5.99

Last thing

Calculated chi-square > critical chi-square → I will reject the null hypothesis

So we are 95% confidence that there is an association between your political party and your position of using weapons In the population level of US

NOTE: remember to write your confidence
Also you cannot be 100% confidence

To sum up

1. Determine the appropriate test
2. Establish the level of significance: α
3. Formulate the statistical hypothesis
4. Calculate the test statistic
5. Determine the degree of freedom
6. Compare computed test statistic against a tabled/critical value

Example:

I have a sample of 100 patient and I wonder if there is an association between smoking cigarette and developing lung cancer

	Cancer	No-cancer	row
Smoker	1 35	2 10	45
Non-smoker	3 5	4 50	55
column	40	60	N=100

First step → both are nominal **SO** I use chi-square

Second step → alpha is 0.05

Third step →

Ha: there is an association between smoking cigarette and developing lung cancer

Ho: there is **No** association between smoking cigarette and developing lung cancer

Fourth step → calculate chi-square

First, calculate expected value

Cell 1 → (40)*(45)/100 = 18 Cell 2 → (60)*)45)/100 = 27

Cell 3 → (40)*(55)/100 = 22 Cell 4 → (60)*(55)/100 = 33

Second Calculate chi-square for each cell then add them together

Cell 1 → $(35-18)^2/18 = 16.06$ cell 2 → $(10-27)^2/27 = 10.70$

Cell 3 → $(5-22)^2/22 = 13.14$ cell 4 → $(50-33)^2/33 = 8.76$

Chi-square→ 16.06 + 10.70 + 13.14 + 8.76 = 48.66

Fifth step → degree of freedom

Df→ (2-1)*(2-1) = 1

Sixth step → find the critical chi-square

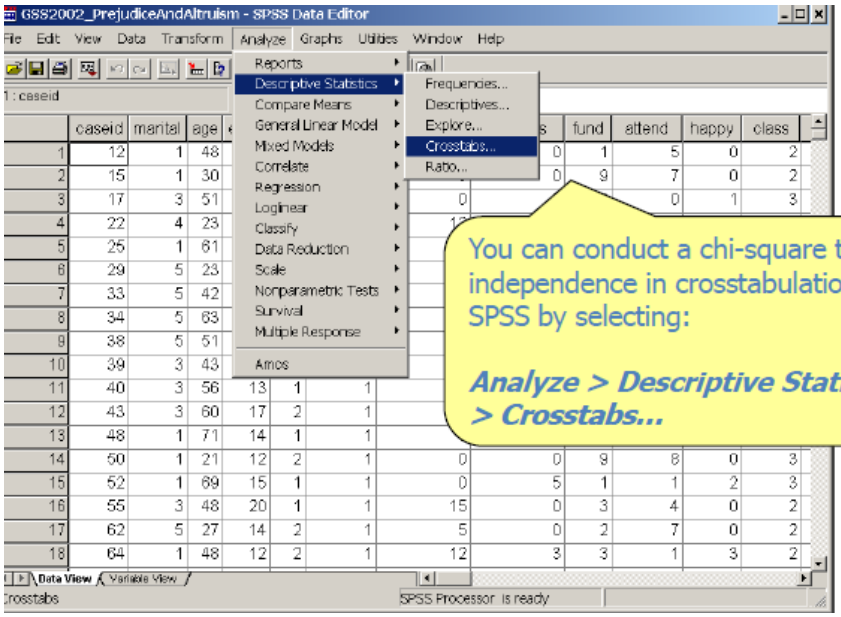
Critical chi-square → 3.841

Since calculated chi-square > critical chi-square → null hypothesis is rejected

So I am 95% confidence that there is an association between smoking and developing lung cancer

	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21

How SPSS calculate chi-square?

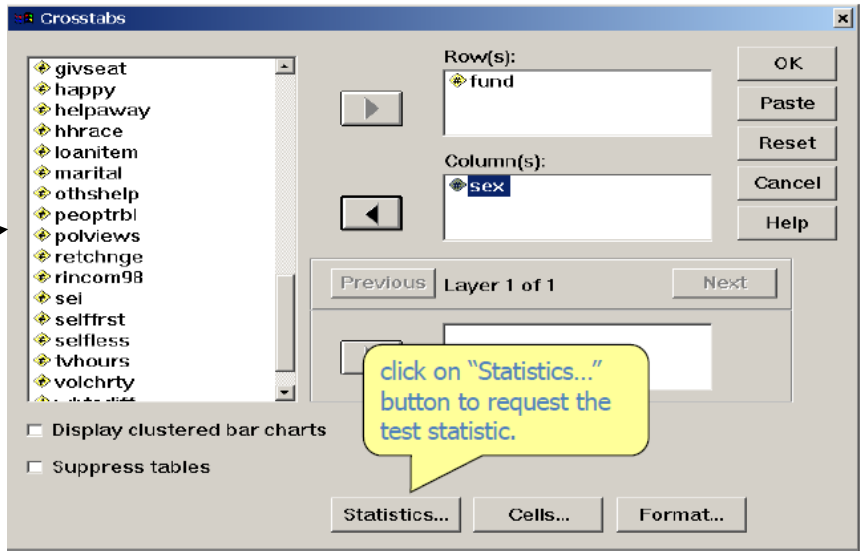


First

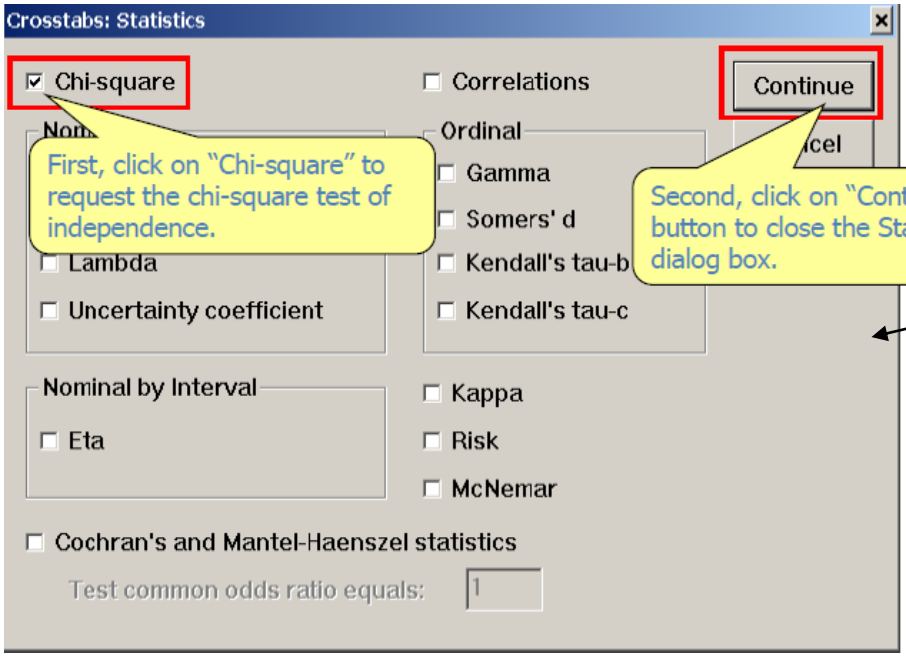
You can conduct a chi-square test of independence in crosstabulation of SPSS by selecting:

Analyze > Descriptive Statistics > Crosstabs...

Second



click on "Statistics..." button to request the test statistic.

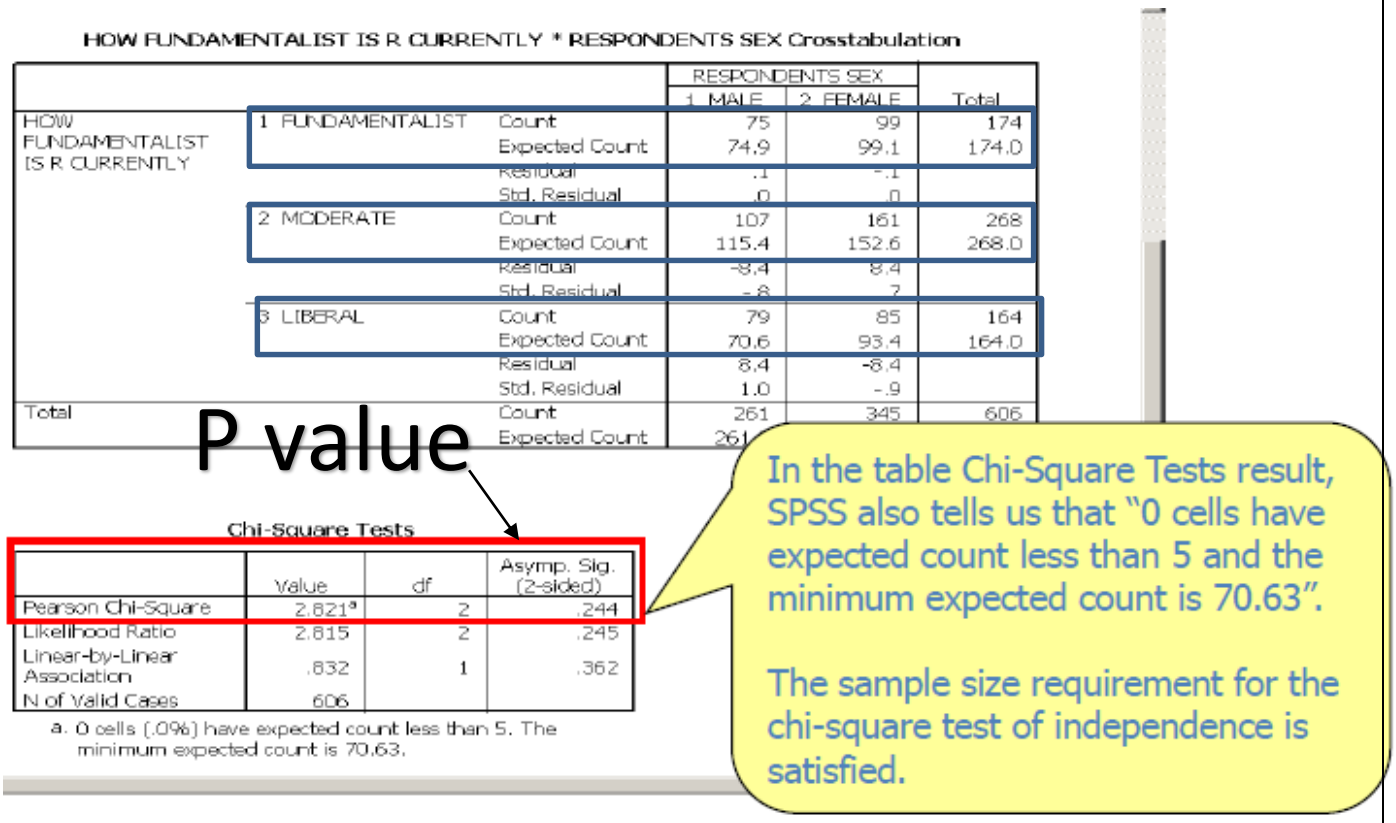


First, click on "Chi-square" to request the chi-square test of independence.

Second, click on "Continue" button to close the Statistics dialog box.

Third

I end up with



HOW FUNDAMENTALIST IS R CURRENTLY * RESPONDENTS SEX Crosstabulation

			RESPONDENTS SEX		
			1 MALE	2 FEMALE	Total
HOW FUNDAMENTALIST IS R CURRENTLY	1 FUNDAMENTALIST	Count	75	99	174
		Expected Count	74.9	99.1	174.0
		Residual	.1	-.1	
		Std. Residual	.0	.0	
	2 MODERATE	Count	107	161	268
		Expected Count	115.4	152.6	268.0
		Residual	-8.4	8.4	
		Std. Residual	-.8	.7	
	3 LIBERAL	Count	79	85	164
		Expected Count	70.6	93.4	164.0
		Residual	8.4	-8.4	
		Std. Residual	1.0	-.9	
Total		Count	261	345	606
		Expected Count	261		

P value

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.821[a]	2	.244
Likelihood Ratio	2.815	2	.245
Linear-by-Linear Association	.832	1	.362
N of Valid Cases	606		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 70.63.

In the table Chi-Square Tests result, SPSS also tells us that "0 cells have expected count less than 5 and the minimum expected count is 70.63".

The sample size requirement for the chi-square test of independence is satisfied.

He did not give the critical value but he give me a better thing, which is P value

P value →

If alpha value > p value → null hypothesis is rejected
If alpha value < p value → null hypothesis is accepted

Alpha = 0.05 (as usual) P value = 0.244
So If alpha value < p value → null hypothesis is accepted
So I am 95% confidence that there is no association

(for better understanding watch video 10 from 39:44 to the end)

Best of luck