

Tutorial 6

Research Methods for Political Science - PO3110

Andrea Salvi

5 & 6 November 2019

Trinity College Dublin,

<https://andrsalvi.github.io/research-methods/>

Table of contents

1. Chi Square Test
2. In-class Exercise with UCDP Data

Chi Square Test

A brief review of cross-tabs

- Two nominal or ordinal variables
- Keep ordering for ordinal variables
- Independent variables in the columns
- Dependent variables in the rows
- Calculate column percentages
- Compare percentages across the rows

Practical Example

	Republican	Democrat	Other
Male	26	13	5
Female	20	29	7

Practical Example

	Republican	Democrat	Other
Male	26	13	5
Female	20	29	7

1. Calculate totals for the rows;

Practical Example

	Republican	Democrat	Other
Male	26	13	5
Female	20	29	7

1. Calculate totals for the rows;
2. 44 males, 56 females, grand-total; of 100!

Practical Example

	Republican	Democrat	Other
Male	26	13	5
Female	20	29	7

1. Calculate totals for the rows;
2. 44 males, 56 females, grand-total; of 100!
3. Let's update our table accordingly

Practical Example

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
				100

Practical Example

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
				100

1. Calculate totals for the column;

Practical Example

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
				100

1. Calculate totals for the column;
2. 46 Republicans, 42 Democrats, 12 others;

Practical Example

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
				100

1. Calculate totals for the column;
2. 46 Republicans, 42 Democrats, 12 others;
3. Grand-total is obviously 100;

Practical Example

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
				100

1. Calculate totals for the column;
2. 46 Republicans, 42 Democrats, 12 others;
3. Grand-total is obviously 100;
4. Let's update our table once more.

Practical Examples

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

Practical Examples

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

1. Let's now calculate the expected frequencies

Practical Examples

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

1. Let's now calculate the expected frequencies
2. $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$

Practical Examples

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

1. Let's now calculate the expected frequencies
2. $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$
3. Let's start from Male Republicans. What's the row margin?
What's the column margin?

Practical Examples

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

1. Let's now calculate the expected frequencies
2. $f_e = \frac{\text{row margin} * \text{column margin}}{\text{total}}$
3. Let's start from Male Republicans. What's the row margin?
What's the column margin?

<u>Expected</u>	Republican	Democrat	Other
Male	20.24	18.48	5.28
Female	25.76	23.52	6.72

We are essentially doing hypothesis testing. Remember the five steps¹?

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;
2. Choose level of significance (α);

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;
2. Choose level of significance (α);
3. Find Critical Values;

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;
2. Choose level of significance (α);
3. Find Critical Values;
4. Find Test Statistics;

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;
2. Choose level of significance (α);
3. Find Critical Values;
4. Find Test Statistics;
5. Draw Conclusion.

¹Credits to Mathmeeting

We are essentially doing hypothesis testing. Remember the five steps¹?

1. State null (H_0) and alternative (H_1) hypothesis;
2. Choose level of significance (α);
3. Find Critical Values;
4. Find Test Statistics;
5. Draw Conclusion.

Let's adapt it to our test!

¹Credits to Mathmeeting

Practical Example

- H_0 : Political Party independent upon gender

Practical Example

- H_0 : Political Party independent upon gender
- H_1 : Political Party dependent upon gender

Practical Example

- H_0 : Political Party independent upon gender
- H_1 : Political Party dependent upon gender
- Level of significance $\alpha = 0.05$. As usual this defines the area of our rejections region.

Practical Example

- H_0 : Political Party independent upon gender
- H_1 : Political Party dependent upon gender
- Level of significance $\alpha = 0.05$. As usual this defines the area of our rejections region.
- We need to find the critical value that separates the tail from the rest of the area.

Practical Example

- H_0 : Political Party independent upon gender
- H_1 : Political Party dependent upon gender
- Level of significance $\alpha = 0.05$. As usual this defines the area of our rejections region.
- We need to find the critical value that separates the tail from the rest of the area.
- In order to do that we resort to a χ^2 table
- To use such a tool we need to know *the area in the tail* and *our degrees of freedom*

Practical Example

- H_0 : Political Party independent upon gender
- H_1 : Political Party dependent upon gender
- Level of significance $\alpha = 0.05$. As usual this defines the area of our rejections region.
- We need to find the critical value that separates the tail from the rest of the area.
- In order to do that we resort to a χ^2 table
- To use such a tool we need to know *the area in the tail* and *our degrees of freedom*
- $df = (\text{rows} - 1) * (\text{columns} - 1)$

Practical Example - Critical Values

Degrees of Freedom	Chi-Square (χ^2) Distribution							
	Area to the Right of Critical Value							
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	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	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

Practical Example

We now need to calculate the test statistic. In this case: χ^2 !

Practical Example

We now need to calculate the test statistic. In this case: χ^2 !

- We know that the formula is: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
- In our case:

Practical Example

We now need to calculate the test statistic. In this case: χ^2 !

- We know that the formula is: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
- In our case:

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

<u>Expected</u>	Republican	Democrat	Other
Male	20.24	18.48	5.28
Female	25.76	23.52	6.72

Practical Example

We now need to calculate the test statistic. In this case: χ^2 !

- We know that the formula is: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$
- In our case:

	Republican	Democrat	Other	Total
Male	26	13	5	44
Female	20	29	7	56
	46	42	12	100

<u>Expected</u>	Republican	Democrat	Other
Male	20.24	18.48	5.28
Female	25.76	23.52	6.72

Plugging everything into the formula we get a test statistic of **5.86**

Practical Example

Time to draw conclusions! What to we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

Practical Example

Time to draw conclusions! What do we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

The value is **NOT** in our rejection region thus **we fail to reject the null hypothesis!**

Practical Example

Time to draw conclusions! What do we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

The value is **NOT** in our rejection region thus **we fail to reject the null hypothesis!**

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.

Practical Example

Time to draw conclusions! What do we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

The value is **NOT** in our rejection region thus **we fail to reject the null hypothesis!**

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).

Practical Example

Time to draw conclusions! What do we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

The value is **NOT** in our rejection region thus **we fail to reject the null hypothesis!**

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).
- We need a measure of association! Let's take Cramers' V.
- $$V = \sqrt{\frac{\chi^2}{N * \text{MIN of } r-1, c-1}}$$

Practical Example

Time to draw conclusions! What do we conclude comparing our test statistic $\chi^2 = 5.86$ with the critical value of 5.991?

The value is **NOT** in our rejection region thus **we fail to reject the null hypothesis!**

- Chi squared does not tell you how strong a relationship is, only whether it is statistically significant.
- If N is large, you are likely to find a significant relationship (but it might be a weak one).
- We need a measure of association! Let's take Cramers' V.
- $$V = \sqrt{\frac{\chi^2}{N * \text{MIN of } r-1, c-1}}$$

In-class Exercise with UCDP Data

In-class Exercise

- Download "UCDP.sav" <https://tinyurl.com/ucdp-mt5>
- You can work in pairs

In-class Exercise

- Create a new syntax file and store your output there!
- Paste the following into the syntax file: what does that do?

```
DATASET ACTIVATE DataSet1.  
COMPUTE duration=end_year - start_year.  
EXECUTE.
```

In-class Exercise 2

- Subset the data in order to select just the following conflicts (hint: "type_of_conflict"): "Internal armed conflict occurs between the government of a state and one or more groups (no int)".
- Subset the data in order to select just conflict occurring in Africa and Middle-East
- Plot a histogram of your choice that conveys meaningful information.
- Have a look at the visualization options in SPSS. Any hints on which ones suits our data?

In-class Exercise 2

- Get rid of cases with missing values in the duration variable.
- "Split" the data-set based on the "region" variable (Hint: Data -> Split File). Now try to calculate the mean, standard deviation and Standard Error of the Mean for the "duration" variable. What happened?
- Conduct an independent sample t-test to compare the duration in Africa and in the Middle East. Are they significantly different?
- Pick two variables of interests and produce a crosstab!
- Calculate the χ^2 and the Cramer's V