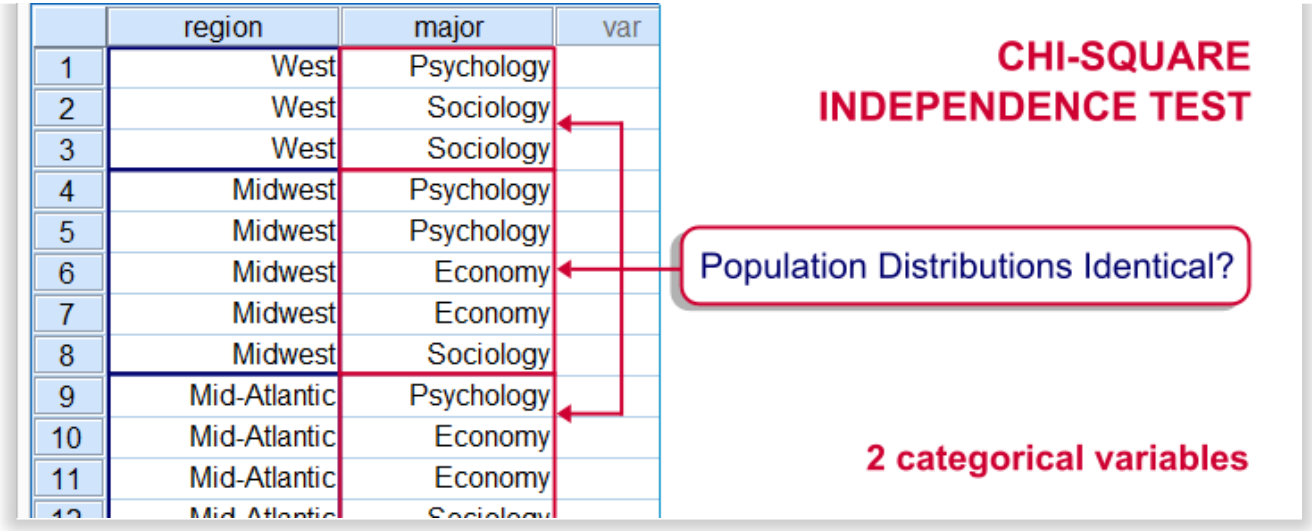
SPSS Chi-Square Independence Test

- Null Hypothesis
- Assumptions
- Chi-Square Independence Test in SPSS
- Output
- Reporting

Null Hypothesis Chi-Square Independence Test



A chi-square independence test evaluates if two categorical variables are associated in some population. We'll therefore try to refute the null hypothesis that

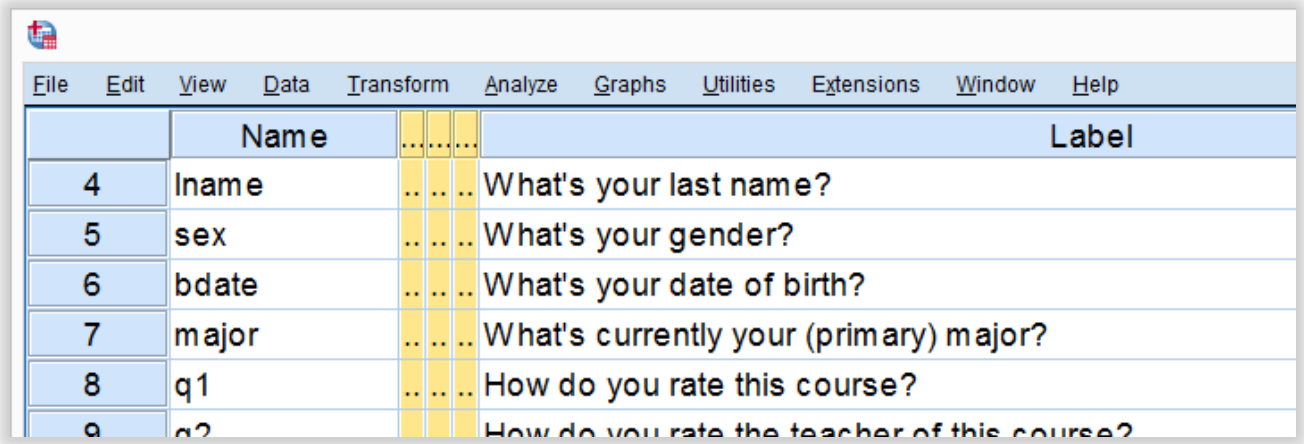
two categorical variables are (perfectly) independent in some population.

If this is true and we draw a sample from this population, then we may see *some* association between these variables in our sample. This is because **samples tend to differ** somewhat from the populations from which they're drawn.

However, a ***strong association*** **between variables is unlikely** to occur in a sample if the variables are independent in the entire population. If we do observe this anyway, we'll conclude that the variables probably aren't

Example

A sample of 183 students evaluated some course. Apart from their evaluations, we also have their genders and study majors. The data are in course_evaluation.sav, part of which is shown below.



	Name	Label
4	lname	What's your last name?
5	sex	What's your gender?
6	bdate	What's your date of birth?
7	major	What's currently your (primary) major?
8	q1	How do you rate this course?
9	q2	How do you rate the teacher of this course?

We'd now like to know: **is study major associated with gender?** And -if so- how? Since study major and gender are nominal variables, we'll run a chi-square test to find out.

Assumptions Chi-Square Independence Test

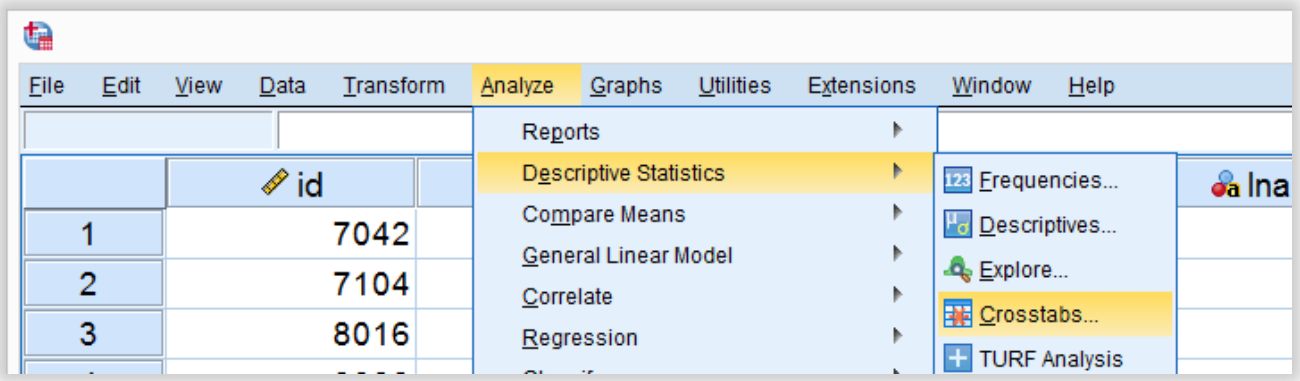
Conclusions from a chi-square independence test can be trusted if two

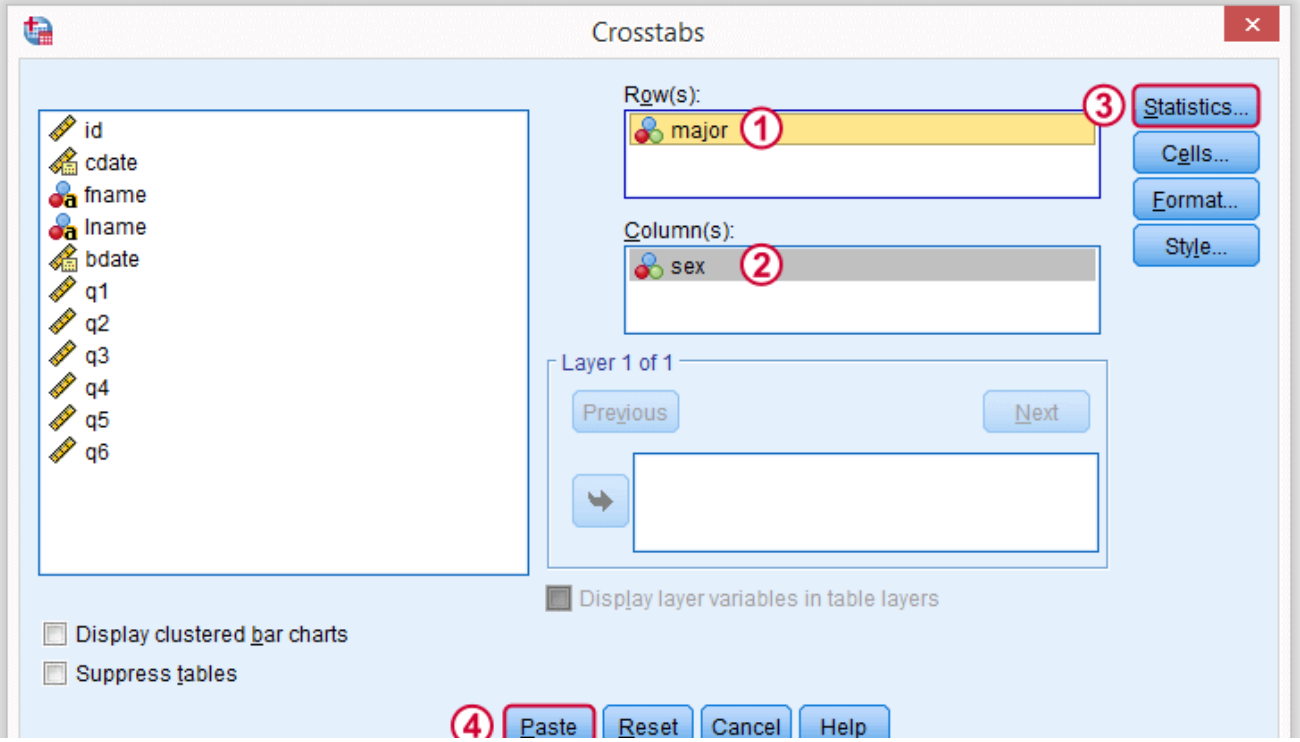
- **independent observations**. This usually -not always- holds if each case in SPSS holds a unique person or other statistical unit. Since this is that case for our data, we'll assume this has been met.
- For a 2 by 2 table, all **expected frequencies > 5**.* For a larger table, no more than 20% of all cells may have an expected frequency < 5 and all expected frequencies > 1.

  SPSS will test this assumption for us when we'll run our test. We'll get to it later.

Chi-Square Independence Test in SPSS

In SPSS, the chi-square independence test is part of the CROSSTABS procedure which we can run as shown below.





In the main dialog, we'll enter one variable into the Row(s) box and the other into Column(s) . Since sex has only 2 categories (male or female), using it as our column variable results in a table that's rather narrow and high. **It will fit more easily** into our final report than a wider table resulting from using major as our column variable. Anyway, both options yield identical test results.

③ Under Stastistics we'll just select Chi-Square . Clicking Paste results in the syntax below.

SPSS Chi-Square Independence Test Syntax

```
*Crosstabs with Chi-Square test as pasted from menu.

CROSSTABS
/TABLES=major BY sex
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ
/CELLS=COUNT
/COUNT ROUND CELL.
```

You can use this syntax if you like but I personally prefer a shorter version

is much faster than clicking through the menu. Both versions yield **identical results**.

```
*Crosstabs with Chi-Square test - short version.

crosstabs major by sex
/statistics chisq.
```

Output Chi-Square Independence Test

Case Processing Summary

If many missings, find out why.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
major What's currently your (primary) major? * sex What's your gender?	183	100.0%	0	0.0%	183	100.0%

First off, we take a quick look at the **Case Processing Summary** to see if any cases have been excluded due to missing values. That's not the case here. With other data, if many cases are excluded, we'd like to know why and if it makes sense.

major What's currently your (primary) major? * sex What's your gender?

Count

Observed frequencies

Marginal frequencies

		sex What's your gender? 0 female	1 male	Total
major What's currently your (primary) major?	1 Psychology	54	8	62
	2 Economy	7	28	35
	3 Sociology	12	21	33
	4 Anthropology	15	22	37
	5 Other	4	12	16
Total		92	91	183

Next, we inspect our contingency table. Note that its **marginal frequencies** -the frequencies reported in the *margins* of our table- show the frequency distributions of either variable separately.
Both distributions look plausible and since there's no “no answer” categories, there's no need to specify any user missing values.

Significance Test

“The association is statistically significant, $\chi^2(4) = 54.50$, $p = 0.000$”

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	54.504[a]	4	.000
Likelihood Ratio	59.758	4	.000
Linear-by-Linear Association	25.597	1	.000
N of Valid Cases	183		

① First off, **our data meet the assumption** of all expected frequencies > 5 that we mentioned earlier. Since this holds, we can rely on our significance test for which we use Pearson Chi-Square.

Right, we usually say that the association between two variables is statistically significant if **Asymptotic Significance (2-sided) < 0.05** which is clearly the case here.

Significance is often referred to as "**p**", short for probability; it is the probability of observing our sample outcome if our variables are independent in the entire population. This probability is 0.000 in our case.

Conclusion: we reject the null hypothesis that our variables are independent in the entire population.

Understanding the Association Between Variables

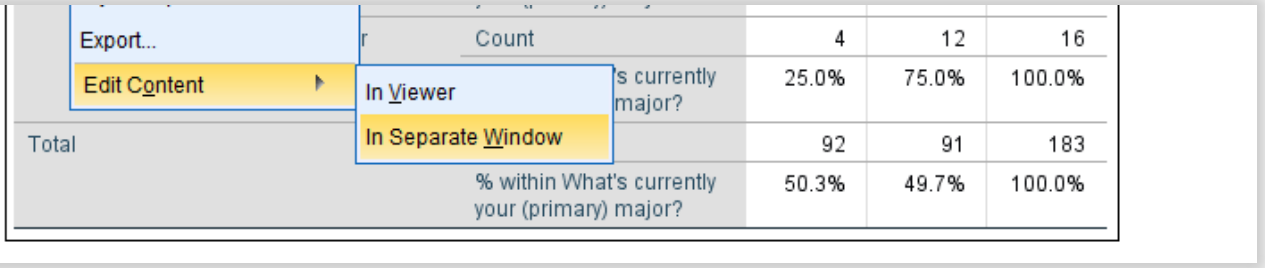
We conclude that our variables are associated but what does this association look like? Well, one way to find out is inspecting either column or **row percentages**. I'll compute them by adding a line to my syntax as shown below.

```
*Show only variable/value labels in output.

set tvars labels tnumbers labels.
```

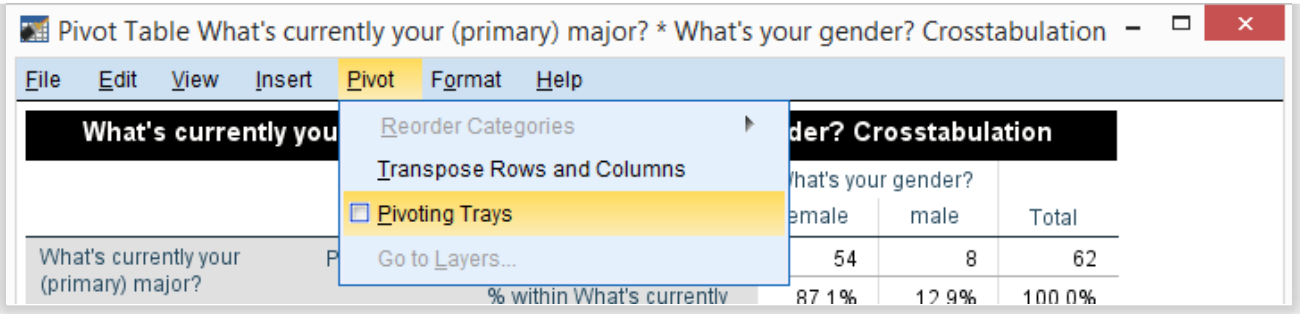
```
*Crosstabs with frequencies and row percentages.

crosstabs major by sex
/cells count row
/statistics chisq.
```

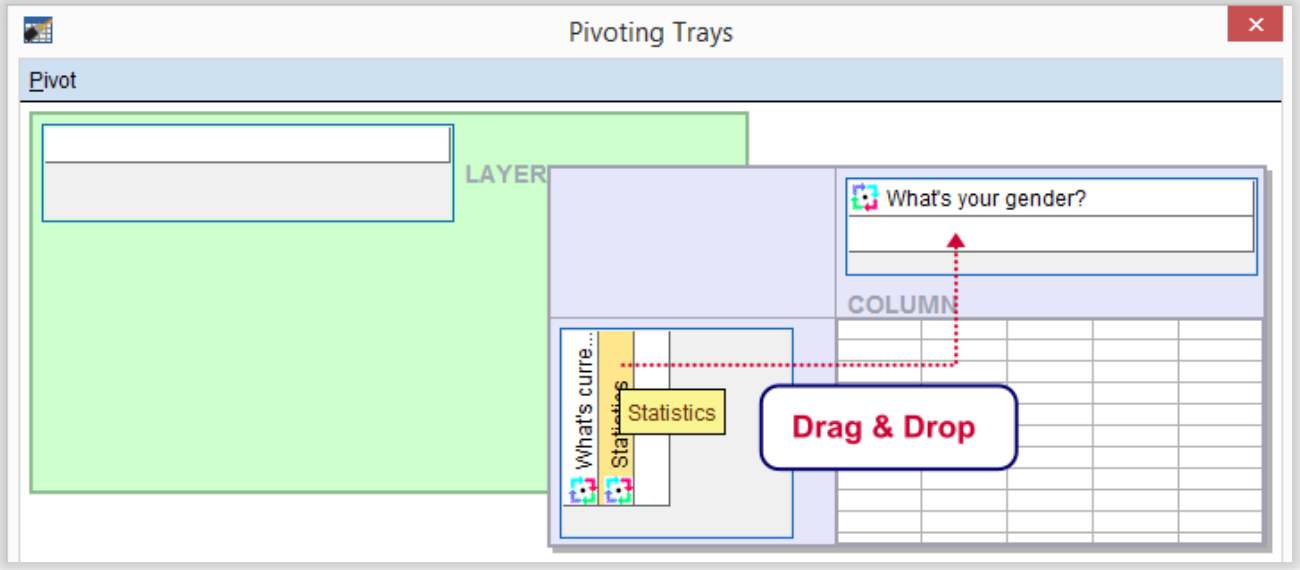
Adjusting Our Table

Since I'm not too happy with the format of my newly run table, I'll right-click it and select Edit Content ▸ In Separate Window



We select Pivoting Trays and then drag and drop Statistics right underneath “What's your gender?”. We'll close the pivot table editor.


Pivot Table What's currently your (primary) major? * What's your gender? Crosstabulation
File Edit View Insert Pivot Format Help
Reorder Categories
Transpose Rows and Columns
Pivoting Trays
Go to Layers...
What's currently your (primary) major?
What's your gender?
female
male
Total
54
8
62
87.1%
12.9%
100.0%


Pivoting Trays
Pivot
LAYER
What's your gender?
COLUMN
Statistics
Drag & Drop

Result

		female		male		Total	
		Count	% within What's currently your (primary) major?	Count	% within What's currently your (primary) major?	Count	% within What's currently your (primary) major?
What's currently your (primary) major?	Psychology	54	87.1%	8	12.9%	62	100.0%
	Economy	7	20.0%	28	80.0%	35	100.0%
	Sociology	12	36.4%	21	63.6%	33	100.0%
	Anthropology	15	40.5%	22	59.5%	37	100.0%
	Other	4	25.0%	12	75.0%	16	100.0%
Total		92	50.3%	91	49.7%	183	100.0%

Roughly half of our sample if female. Within psychology, however, a whopping 87% is female. That is, females are highly overrepresented among psychology students. Like so, **study major "says something" about gender**: if I know somebody studies psychology, I know she's probably female.

The opposite pattern holds for economy students: some 80% of them are male. In short, our **row percenages describe the association** we established with our chi-square test.

We could quantify the **strength of the association** by adding Cramér's V to our test but we'll leave that for another day.

Reporting a Chi-Square Independence Test

“an association between gender and study major was observed, $\chi^2(4) = 54.50$, $p = 0.000$.

Further, I suggest including our final contingency table (with frequencies and row percentages) in the report as well as it gives a lot of insight into the nature of the association.

So that's about it for now. Thanks for reading!

Let me know what you think!

Your name*

Your email address*

Your website

Your comment*

Done!

**Required field. Your comment will show up after approval from a moderator.*

This tutorial has 60 comments

Hi Sevi!

I've been working on a tutorial explaining exactly that but I haven't finished it yet. The chi-squ

Expand comment | all comments

By sevi on December 28th, 2017

How to find further which all major subjects are significantly associated with one particular gender?

By Muhammad irshad on October 20th, 2017

amazing bro
keep it up
may b there is no simple explanation like that.
really appreciated

By sami ullah on October 19th, 2017

very simple and well-explained. useful for the beginners.

By Charles Senessie on September 22nd, 2017

Very interesting and well laid out description of the tutorial.

Popular Pages

Basics

Regression

T-Test

Chi-Square Test

ANOVA

Get In Touch!

Ruben Geert van den Berg

LinkedIn

Facebook

Need More SPSS Help?

Sigma Plus Consulting

www.sigma-plus-consulting.com

info@sigma-plus-consulting.com