6

Biostatistics

Doctor 2018 | Medicine | JU

◉ Sheet ○ Slides

Done By

Hadeel Abdullah

Contributed in the Scientific Correction

Hadeel Abdullah

Contributed in the Grammatical Correction

Hadeel Abdullah

Doctor

Hamza Alduraidi

→ Study this sheet carefully and try to understand each and every point because doctor Hamza said that -in the exam- there are more than 6 questions on this topic.

→ After that, check the photos in dr.Hamza's slides and please, don't hesitate to ask.

A quick recapitulation:

We said previously that several conditions of validity must be met so that the result of a parametric test is reliable. For example, we use t-test and ANOVA (parametric tests) to compare mean differences between groups assuming that:

-Dependent variables are continuous (intervals / ratios).

- Groups are randomly drawn from normally distributed populations.

- Groups are homogeneous.

- Samples are large enough to represent populations (n>30).

On the other hand, when the assumptions are violated; like when subjects are not randomly sampled, dependent variables are discrete (ordinal (ranked) / nominal (categorized)), groups are drawn from a greatly skewed population, or when we have no knowledge about the population distribution, we use tests that are equivalent to parametric tests, these tests are called "non-parametric" tests.

For clarification; one assumption for the one way ANOVA is that the data comes from a normal distribution. If your data isn't normally distributed, you can't run an ANOVA, but you can run the non-parametric alternative—the Kruskal-Wallis test.

Chi-Squared Test ($\chi 2$):

-Chi-Square is used when both variables are measured on a nominal scale; dependent and independent variables are nominal.

-It can be applied to interval or ratio data that have been categorized into a small number of groups.

-It assumes that the observations are randomly sampled from a population.

-All observations are independent (an individual can appear only once in a table and there are no overlapping categories).

-It does not make any assumptions about the shape of the distribution nor about the homogeneity of variances.

-A sufficiently large sample size is required (n>20); the total number of observations must be greater than 20.

-Data must be in **raw frequencies**; the data must be in the form of frequencies/ number of occurrences. (Actual **count** data; not percentages)

-The frequency data must have a precise numerical value and must be organized into **categories** or groups.

-Does **not** prove causality.

-The **expected frequency** in any one cell of the table must be **greater than five.** E frequency / E value is discussed in the following page; just know that it must be greater than five.

The **purpose** of Chi-squared Test is to determine if two variables of interest independent (not related) or are related (dependent).

When the variables are independent, we are saying that knowledge of one gives us **no** information about the other variables. When they are dependent; we are saying that knowledge of one variable is helpful in predicting the value of the other variables.

How to do a Chi-Squared Test:

1. Establish level of significance (α).

- The alpha level (α) is a predetermined value. The usual alpha level is 0.05 (5%), but you could also have other levels like 0.01 or 0.10.

2. Determine the hypotheses (H_o and H_1).

-A null hypothesis (H_o) is a hypothesis that says there is **no** statistical significance between the two variables (the two variables are **independent**). It is the hypothesis a researcher will try to disprove (we as researchers will try to reject the null hypothesis). An alternative hypothesis (H_1) is one that states **there is** a statistically significant relationship between two variables (the two variables are **associated**).

3. Draw a Chi-Square Table (**Contingency Table**).

-Draw a table that displays the frequency distribution of the variables.

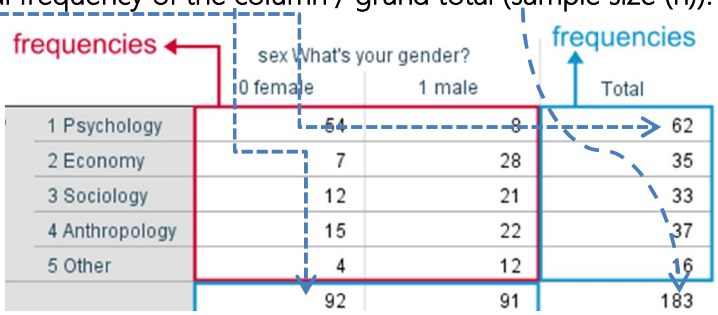
-Contrasts **observed frequencies** in each cell of a contingency table with **expected frequencies.**

An **expected frequency (E)** is a theoretical predicted **(calculated)** frequency obtained from an experiment presumed to **be true** (H_o) until statistical evidence in the form of a hypothesis test indicates otherwise. It represents the number of cases that would be found in the cell if H_o were true.

An **observed frequency**, on the other hand, is the **actual frequency** that is obtained from the experiment.

(E) Value = Total frequency of the row x Total frequency of the column / grand total (sample size (n)).

– For the example beside, the (E) value for female psychologists is (62*92 / 183).



-We are not required to calculate (E) value.

4. Calculate Test Statistic ($\chi 2$ calc.)

$$\chi^2 = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

χ2: calculated Chi-Square value.

O: observed frequency.

E: expected frequency.

- For each category, you find the difference between the observed number (O) and expected number (E), **square** it, and then divide it by the expected number (E). Finally, **add up** the figures for each category (total differences).

5. Determine the "Degree of Freedom" → df= (R (rows) – 1) x (C (columns) – 1)

df = (number of levels in row variable – 1) x (number of levels in column variable – 1)

6. Compare the computed (**calculated**) test statistic against a tabled / **critical value**.

-The calculated value of the Pearson Chis-Square statistic is compared with the critical value to determine if the calculated value is **improbable**.

-The critical values are based on sampling distributions of the Pearson Chi-Square statistic.

- If $\chi 2_{calc.}$ is **greater** than $\chi 2_{cv}$ (cv: critical value), then we reject the null hypothesis (H_o).

7. Interpret Results.

- According to our decision in the previous step (whether to reject or to keep H_o), we express our results as follows:

-If H_o is rejected:

We are ((1-α) %) confident that there is a statistically significant **association** between (Variable1) and (Variable2) in the population. ($\chi 2_{calc}$= ( ), α = ( )).

-If H_o is not rejected (kept):

We are ((1-α) %) confident that there is **NO** statistically significant association between (Variable1) and (Variable2) in the population. (χ^2_{calc}= ( ), α = ( ))

Chi-Square Test of Independence in SPSS

Chi-Square test is an option within the "Crosstabs" procedure which creates a **contingency table** that summarizes the distribution of two categorical (nominal) variables.

To perform **a chi-square test** of independence:

-Click **Analyze > Descriptive Statistics > Crosstabs.**

-Enter at least one row variable, and one column variable.

- Click on **statistics** button to request the **test statistic.**

This opens the **"Crosstabs: Statistics"** window, which contains different inferential statistics for comparing categorical (nominal) variables. To request the **Chi-Square Test of Independence**, make sure that the **Chi-square** box is checked off. **[√]**

- Click "**continue**" to close the statistics dialog box.

Output:

1) Contingency Table. (Discussed earlier)→

Count

		Gender		
		Male	Female	Total
Do you smoke cigarettes?	Nonsmoker	149	148	297
	Past smoker	13	24	37
	Current smoker	31	37	68
Total		193	209	402

2) Significance Test (Chi-square test) table:

- This table shows three important statistics:

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	3.171[a]	2	.205
Likelihood Ratio	3.217	2	.200
Linear-by-Linear Association	1.106	1	.293
N of Valid Cases	402		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 17.76.

1. χ^2_{calc}: the computed (calculated) value of Chi-Square test.

2. Degree of freedom:

df = (no. of rows – 1) x (no. of columns – 1)

3. **P- value** or **Sig**-nificance:

Significance is often referred to as "**p-value**", short for probability; it is the probability of observing our sample outcome if our variables are independent (not associated (H_o)) in the entire population.

- If the p-value is lesser than our chosen significance level (usually; α = 0.05), we reject the null hypothesis.

So, reject H_o when: p-value < **α** and **X^2 calc.** > X^2 cv.