CHI-SQUARE TEST OF INDEPENDENCE: HINTS

You will find in this file general approach on how to apply Chi-square test. Adapt the method to your data and your requests!

CHI-SQUARE TEST OF INDEPENDENCE: THEORETICAL BACKGROUND

In research, when we are interested in investigating the relationship between qualitative variable, the appropriate test is sometimes the chi-square test of independence. The steps necessary to be follow in order to carry out this test are:

- Calculate the observed frequencies for each cell of the cross tabulation (the Pivot Table can be used here);
- Calculate the expected frequencies for each cell of the cross tabulation (there is not a predefined formula to calculate expected frequencies);
- Compute the value of chi-square parameter;
- Evaluate the significance of obtained chi-square parameter.

The formula for the chi-square test of independence is give by:

$$\chi^2 = \sum \frac{\left(O_i - E_i\right)^2}{E_i}$$
 where O_i is an observed cell frequency and E_i is an expected cell frequency.

Assumptions for the Chi-Square Test of Independence

- 1. Subjects are randomly and independently sampled from the population of interest
- 2. Measurements are obtained from a single sample
- 3. Variables included in the analysis are measured on a qualitative scale
- 4. Expected cell frequencies are greater than or equal to five.

APPLYING CHI-SQUARE TEST: MICROSOFT EXCEL

To create the observed contingency table:

[Insert – Tables – Pivot Table – Pivot Table]

🗶 🔚 🥞	• (°" -	∓							In	k Tools		
File	Home	Insert	Page Layou	t Form	iulas D	Data	Review	Viev	v	Pens		
17					1 -+	1	*	0	=		:*:	0
PivotTable	Table	Picture C	lip Shapes S Art *	martArt Scr	reenshot *	Column	Line	Pie *	Bar *	Area	Scatter	Other Charts *
Pivot	Table		Illustratio	ons		<u> </u>		C	harts			19
PivotChart		• (f _x									

• Select data corresponding to all three variables as well as the cell where you want the results:

Create PivotTable	
Choose the data that you want to analyze	Select the data
Select a table or range	corresponding to all
Table/Range: 'Chi-Square'!\$A\$1:\$C\$111	three columns
© <u>U</u> se an external data source	
Choose Connection	
Connection name:	
Choose where you want the PivotTable report to be placed	
© <u>N</u> ew Worksheet	Click on an empty cell
Existing Worksheet Existing Works	I from where you want
Location: 'Chi-Square'!\$F\$1	to put the results
OK Cancel	



• An empty table like the one in the image bellow will appear:

• Place the "*BreastFeeding*"/"Sex" variable on rows and "*FerripriveAnemia*" on columns. Place one variable ("Sex"/"FerripriveAnemia") on Values cell:

Row Labels ▼ no yes Grand Total F 48 13 61 M 33 16 49 Grand Total 81 29 110 FerripriveAnemia (yes/no) FerripriveAnemia (yes/no) Row Labels Column Labels FerripriveAnemia Sex	F Count of FerripriveAnemia (yes	G	H bels 💌		PivotTable Field List
Drag fields between areas below:	Row Labels F M Grand Total	▼ no	yes Gri 48 13 33 16 81 29	and Total 61 49 110	Sex BreastFeeding (yes/no) FerripriveAnemia (yes/no)
Drag fields between areas below:					
Row Labels Σ Values Sex Count of Ferr					Drag fields between areas below: Report Filter Column Labels FerripriveAne
🔄 📃 Defer Layout Update Update					Row Labels Σ Values Sex Count of Ferr▼ □ Defer Layout Update Update

G

Н

1

• Rearrange the table as in the image bellow:

F

Count of FerripriveAnemia (yes/no	o) Column Label	s ↓		
Row Labels	✓ yes		no	Grand Total
F		13	48	61
M		16	33	49
Grand Total		29	81	110

Select the cell 'no' + right click and [Sort – Sort Z to A]

61

49

110

▼ (<i>f</i> x no						~
D E	Arial	- 10 - A A \$ - % , a	G	Н		PivotTable Field List 🛛 👻 🗙
Cou Row	B		an Labels 💌	yes	Grand Total ■	Choose fields to add to report:
F M Gra	■	<u>C</u> opy <u>F</u> ormat Cells <u>R</u> efresh	48 33 81	13 16 29	61 49 110	V Sex BreastFeeding (yes/no) FerripriveAnemia (yes/no)
A Sort A to Z		<u>S</u> ort •				
X↓ Sort Z to A More Sort Options	 Image: A start of the start of	Filter ▶ Subtotal "FerripriveAnemia (yes/no)" Expand/Collapse				
	* *	Group Ungroup Move ► Parous "FarrinriveAnamia (var (no)"				Drag fields between areas below: V Report Filter Column Labels FerripriveAne
	•	Field Settings PivotTable Options Hide Field List				Row Labels Σ Values Sex ▼ Count of Ferr ▼

Copy and paste the observed contingency table and change the table as in the image bellow: •

Observed table	Feriprive anemia = yes	Feriprive anemia = no	Grand Total
F	13	48	61
M	16	33	49
Grand Total	29	81	110

To create the expected contingency table:

Copy the GrandTotal structure of the observed table: •

Expected table	Feriprive anemia = yes	Feriprive anemia = no	Grand Total
F			61
M			49
Grand Total	29	81	110

Apply the following formula to calculate the expected cell frequencies:

 $E_{i} = \frac{(\text{Column Grand Total}) \cdot (\text{Row Grand Total})}{(\text{Column Grand Total})}$

Overall Grand Total

To apply the formula:

Expected table	Feriprive anemia = ves	Feriprive anemia = no	Grand Total
F	=I15*G17/I17		61
Μ			49
Grand Total	29	81	110
Expected table	Ferinrive anemia = ves	Ferinrive anemia = no	Grand Total
F	16	=115*H17/117	61
M			49
Grand Total	29	81	110
Expected table	Feriprive anemia = yes	Feriprive anemia = no	Grand Total
F	16	45	61
-			
M	=I16*G17/I17		49
M Grand Total	= 16*G17/ 17 29	81	49 110
M Grand Total Expected table	= 16*G17/ 17 29 Feriprive anemia = yes	81 Feriprive anemia = no	49 110 Grand Total
M Grand Total Expected table F	= 16*G17/ 17 29 Feriprive anemia = yes 16	81 Feriprive anemia = no 45	49 110 Grand Total 61
M Grand Total Expected table F M	= 16*G17/ 17 29 Feriprive anemia = yes 16 13	81 Feriprive anemia = no 45 = 16*H17/ 17	49 110 Grand Total 61 49
M Grand Total Expected table F M Grand Total	= 16*G17/ 17 Feriprive anemia = yes 16 13 29	81 Feriprive anemia = no 45 = 16*H17/ 17 81	49 110 Grand Total 61 49 110

Feriprive anemia = yes Feriprive anemia = no Grand Total Expected table F 16 45 Μ 36 13 Grand Total 29 81

	Chi-square test of independence: sex vs ferriprive anemia
H ₀ hypothesis	Sex and Feriprive anemina are independent
H _a hypothesis	Sex and Feriprive anemina are dependent
alpha	0.05
df (degrees of freedom)	=(2-1)*(2-1), where 2 = number of rows and 2 = number of columns
Critical Chi-square	=CHISQ.INV.RT ((value of alpha),(value of degrees of freedom))
p-values	=CHISQ.TEST((the 4 values from observed table),(the 4 values from expected table))
Observed Chi-square	=CHISQ.INV.RT((select the cell where the p-value is),(select the cell where df are))

To compute chi-square statistics:

Statistical conclusion	
Clinical conclusion	

- *Alpha*: use alpha equal to 0.05 for this analysis (significance level of 5%).
- *df*: the formula for the degree of freedom (df) for the chi-square test of independence is df = (r 1)·(c 1), where r = the number of rows in the cross tabulation and c = the number of columns in the cross tabulation. Because our cross tabulation has 2 rows and 2 columns, df = $(2 1) \cdot (2 1) = 1$.
- Critical chi-square: the CHIINV function will be use in order to find the critical chi-square value.
 Select the corresponding cell and [Insert Function... Function category (Statistical) CHIINV]:
 - Click in the probability window of the CHIINV dialog box and enter the value of alpha
 - Deg_freedom: click in the Deg_freedom window and enter the df value by selecting the corresponding cell
- *p-value:* this is the probability associated with the observed value of chi-square. The CHITEST function will be use to find the p-value. Select the cell where you want the result and [Insert Function... Function category (Statistical) CHISQ.TEST]. Fill the CHISQ.TEST dialog box with requested information:

	Observed table	Feriprive anemia = yes	F	eriprive anemia = no	Grand Total	
	F		13	48	61	
	M		16	33	49	
	Grand Total		29	81	110	
	Expected table	Feriprive anemia = yes	F	eriprive anemia = no	Grand Total	
			16	45	61	
	M		13	36	49	
	Grand Total		29	81	110	
_						
F	unction Arguments		2	Ref	fers the rang	e associated
1 c				wit	th the observ	/ed cell
	CHISQ.TEST			fre	auencies (ius	st 4 cells)/
	Actual_range G9:H10	I = {13,48;16,33}			1	
ł	Expected_range G15:H16	16.08181818181	82,44.9181818	181		
		= 0.179657057		Refer	rs the range	associated
	Returns the test for independence: the value fro	m the chi-squared distribution for the statistic a	ind the appropri	iate to th	e expected f	requencies
	legrees of freedom.			(inct		equencies
	Actual_range is the	range of data that contains observations to te	st against expe	cted	4 censj/	
	Value	5.				
	Formula result = 0.179657057					
	Help on this function	ОК	Car	ncel		
L						

- *Observed chi-square*: the CHIINV function will be use to find the observed chi-square value.
 - *Probability*: enter here p-value
 - *Deg_freedom*: enter the df value.
- The results will look like in the example bellow:

Chi-square test of independence: sex vs ferriprive anemia				
H _o hypothesis	Sex and Feriprive anemina are independent			
H _a hypothesis	Sex and Feriprive anemina are dependent			
alpha	0.05			
df (degrees of freedom)	1			
Critical Chi-square	3.8415			
p-values	0.1797			
Observed Chi-square	1.8005			
Statistical conclusion	Null hypothesis could not be rejected			
Clinical conclusion	Sex and feriprive anemia are independent			