# INFERENTIAL STATISTICS III - CHI-SQUARE TEST OF INDEPENDENCE: HINTS

## 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:

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

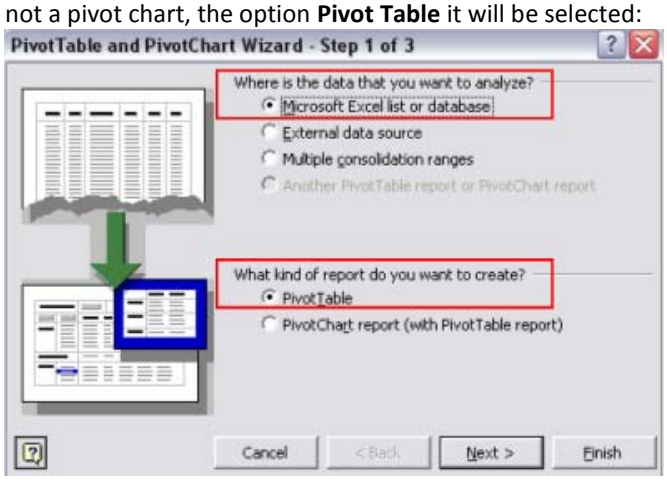
***Assumptions for the Chi-Square Test of Independence***

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.

## MICROSOFT EXCEL

To create the observed contingency table:

- **[Data – Pivot Table and Pivot Chart Report]:**
  - Pivot Table and Pivot Chart Wizard-Step 1 of 3. The date are into the ***Breastfeed.xls*** file, so the **Microsoft Excel List of Database** is already been selected. Because we want to create a pivot table not a pivot chart, the option **Pivot Table** it will be selected:




  - Pivot Table and Pivot Chart Wizard-Step 2 of 3. Enter the range of the data (from A1 to C111) in the space provided (click on cell A1 in the worksheet Cross Tabulation and drag to cell B111).


  - Pivot Table and Pivot Chart Wizard-Step 3 of 3. Put the report in the same worksheet as the data. Select **Existing worksheet.** Click in the Existing worksheet window and then click on cell F2 in order to place the report two columns right beside data.


- Place by drag and drop the "*BreastFeeding*"/"*Sex*" variable on rows and "*FerripriveAnemia*" on columns. Drag and drop any variable ("*FerripriveAnemia*" or "*BreastFeeding*"/"*Sex*") on the area labelled **Drop Data Items Here.** The obtained table look like in the image bellow:

| F | G | H | I |
|---|---|---|---|
| | | | |
| Count of FerripriveAnemia | FerripriveAnemia | | |
| Sex | NO | YES | Grand Total |
| F | 48 | 13 | 61 |
| M | 33 | 16 | 49 |
| Grand Total | 81 | 29 | 110 |

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

| **Observed** | FerripriveAnemia = NO | FerripriveAnemia =YES | Grand Total |
|---|---|---|---|
| Sex = F | 48 | 13 | 61 |
| Sex = M | 33 | 16 | 49 |
| Grand Total | 81 | 29 | 110 |

To create the expected contingency table:

- Copy the GrandTotal structure of the observed table:

| **Expected** | FerripriveAnemia = NO | FerripriveAnemia =YES | Grand Total |
|---|---|---|---|
| Sex = F | | | 61 |
| Sex = M | | | 49 |
| Grand Total | 81 | 29 | 110 |

- Apply the following formula to calculate the expected cell frequencies:

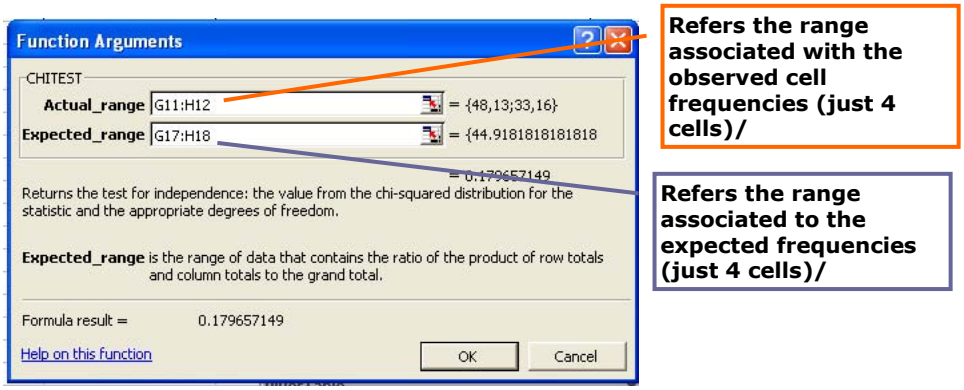
$$E_i = \frac{(\text{Column Grand Total}) \cdot (\text{Row Grand Total})}{\text{Overall Grand Total}}$$

- To begin, click on the Sex = F/FerripriveAnemia = NO expected frequency cell (e.g. G17): =I17*G19/I19
  - Press [Enter]. Cell G17 should contain an expected frequency of 45.
  - Repeat the above steps to calculate the expected frequency for each of the remaining cells.

| **Expected** | FerripriveAnemia = NO | FerripriveAnemia =YES | Grand Total |
|---|---|---|---|
| Sex = F | 45 | 16 | 61 |
| Sex = M | 36 | 13 | 49 |
| Grand Total | 81 | 29 | 110 |

To compute chi-square statistics:

- *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)·(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) - CHITEST]**. Fill the CHITEST dialog box with requested information:

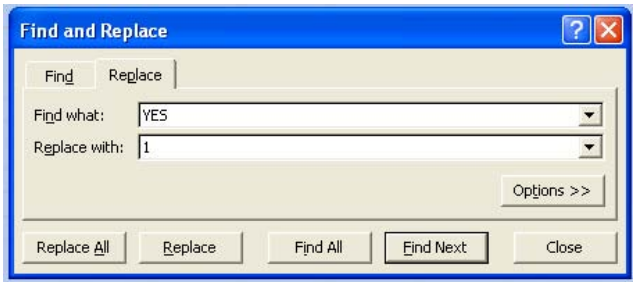



- *Observed chi-square*: the CHIINV function will be use to find the observed chi-square value.
  - *Probability*: enter here p-value
  - *Deg_freedom*: enter the df value.
- The results will look like in the example bellow:

| L | M |
|---|---|
| **Chi-square test of independence: sex vs ferriprive anemia** | |
| H0 | Sex and ferriprive anemia are independent |
| H1 | Sex and ferriprive anemia are dependent |
| alpha | 0.05 |
| df | 1 |
| Critical chi-square | 3.8415 |
| p-value | 0.1797 |
| Observed chi-square | 1.8005 |
| | |
| Conclusion | |
| Statistical | H0 is accepted |
| Clinica | Sex and ferriprive anemia are independent |

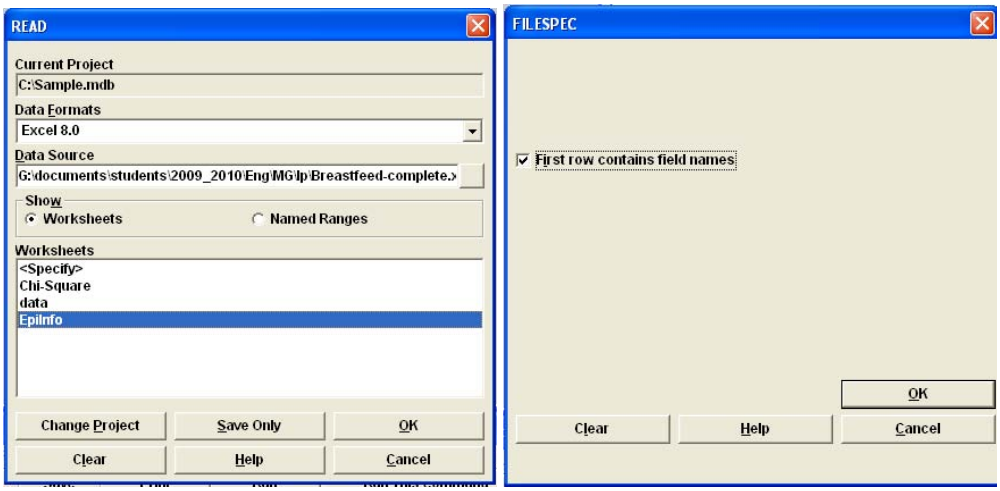
## EPIINFO

To replace text: Ctrl+A – Ctrl + H

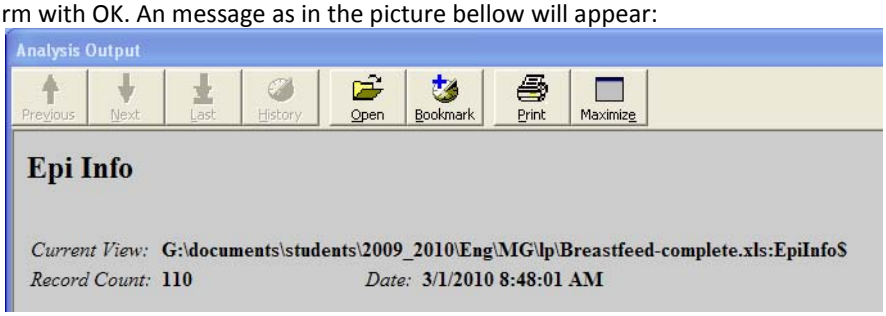



To work with EpiInfo:

- Open Analysis module of EpiInfo ([Start – All Programs – Epi Info - Analysis]). Activate Read(Import) option.
  a. Define the type of document which you wand to open (Excel file) & data source (find the **BreastFeed.xls** file on your partition on the server) & Worksheet (select here the *EpiInfo* work sheet):

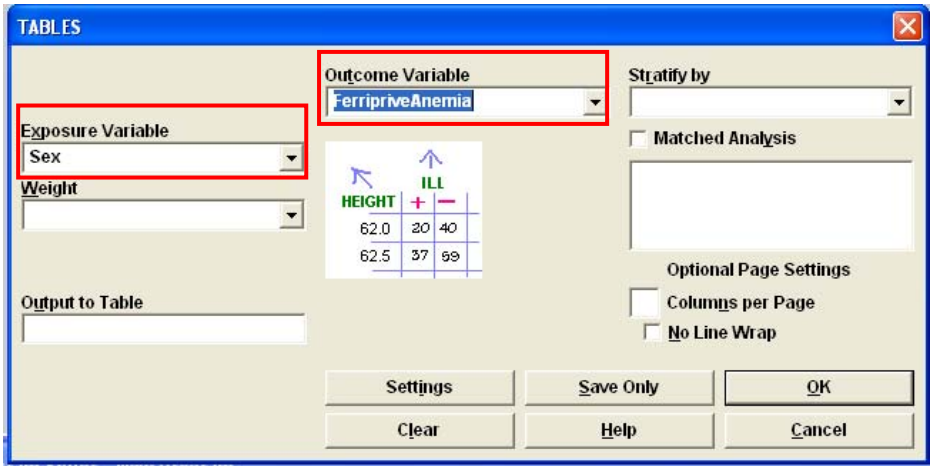



b. Confirm with OK. An message as in the picture bellow will appear:




c. Verify if the open data are the correct ones. From **Statistics** command activates with double click the **List** option. Chose all (*) and validate with OK. It is correct if you have the HT and Diabetes status for all patients (as yes/no variables).

- Use **Table** function to answer to this question. Exposure variables is Sex and outcome variable is FerripriveAnemia




The following results will be obtained:

**Single Table Analysis**

| | Point Estimate | 95% Confidence Interval Lower | Upper | |
|---|---|---|---|---|
| PARAMETERS: Odds-based | | | | |
| Odds Ratio (cross product) | 0.5586 | 0.2374 | 1.3143 | (T) |
| Odds Ratio (MLE) | 0.5616 | 0.2340 | 1.3302 | (M) |
| | | 0.2163 | 1.4324 | (F) |
| PARAMETERS: Risk-based | | | | |
| Risk Ratio (RR) | 0.6527 | 0.3483 | 1.2229 | (T) |
| Risk Difference (RD%) | -11.3416 | -28.0155 | 5.3323 | (T) |

(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)

| STATISTICAL TESTS | Chi-square | 1-tailed p | 2-tailed p |
|---|---|---|---|
| Chi-square - uncorrected | 1.8005 | | 0.1796580164 |
| Chi-square - Mantel-Haenszel | 1.7841 | | 0.1816487535 |
| Chi-square - corrected (Yates) | 1.2636 | | 0.2609656846 |
| Mid-p exact | | 0.0951260507 | |
| Fisher exact | | 0.1306007799 | |

Since the question is about risk factors, we will look and interpret Risk Ratio or Odds Ratio (depending on research methodology – 4th year of study) as parameters and associated confidence interval.
Since the values (both in this case OR = Odds Ratio and RR = Risk Ratio) are not higher than 1 the sex could not be considered a risk factor for FerripriveAmenia.
We also have the result of Chi-square test (are were obtained more easiest compared to Microsoft Excel).