

THE PEARSON CHI-SQUARE TEST IN ELECTRON SPREADSHEETS

Ghotme B., Arsenyev A.

The National University of Pharmacy, Kharkiv, Ukraine

nevs@nuph.edu.ua

The Pearson *Chi*-square is one of the most common test for significance of the relationship between categorical variables. Usually it realized in special statistical computer programs (e.g. Statistica, IBM SPSS Statistics and others). But all these programs are payware and that is why they often are not public domain software. At the same time there are some simple freeware office applications (spreadsheets), which have intrinsic means for calculation value of the *Chi*-square and clearly recognize relations between categorical variables.

Present work contains sketchy description of calculations value of the *Chi*-square using LibreOffice Calc. After entering data in cells of a spreadsheet, the algorithm calculates number of degrees of freedom and expected frequencies for the two-way table (i.e., frequencies that we would expect if there was no relationship between the variables). Then it evaluates the value of the *Chi*-square. The only assumption underlying the use of the *Chi*-square (other than random selection of the sample) is that the expected frequencies are not very small. The reason is that the *Chi*-square inherently tests the underlying probabilities in each cell; and when the expected cell frequencies fall, for example, below 5, those probabilities cannot be estimated with sufficient precision. The approximation of the *Chi*-square statistic in small 2 x 2 tables can be improved by reducing the absolute value of differences between expected and observed frequencies by 0.5 before squaring (Yates' correction). This correction, which makes the estimation more conservative, is usually applied when the table contains only small observed frequencies, so that some expected frequencies become less than 10. In these cases was used formula for Yates correction. All these actions carry into effect automatically with the help of constructed conditional statement.

On the final stage the value of the calculated *Chi*-square should be compared with the critical estimation which is derived by build-in statistical function of LibreOffice Calc.

The obtained result demonstrated good agreement with conclusion which was obtained by professional statistical programs.