

## The Chi-square ( $X^2$ ) statistic

The chi-square statistic used in the analysis of cross-tabulations (or “contingency tables”) is

$$X^2 = \sum_{i=1}^r \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where  $O_{ij}$  = the observed count (or frequency) in the cell in  $i$ -th row and  $j$ -th column of the table,

$E_{ij}$  = the expected count (or frequency) in the cell in the  $i$ -th row and  $j$ -th column of the table

$r$  = the number of rows in the table

$k$  = the number of columns in the table

The expected frequency of observations in cell  $ij$  of the table is the product of the row total ( $R_i$ ) and column total ( $C_j$ ) divided by the total number of observations,  $N$ , or

$$E_{ij} = \frac{R_i C_j}{N}.$$

The expected frequency  $E_{ij}$  is the thus number of observations that one would expect to find in cell  $ij$  of the table, if the frequency of observations in each cell was indifferent to which particular row or column the cell was in, or in other words, if there was no pattern within the table. As the differences between the observed and expected frequencies or counts gets larger (i.e. as the pattern within the table becomes more evident, the value of  $X^2$  becomes larger.