Chi-Squared Cutoffs for Statistical Significance

Abstract: Chi-squared was one of the earliest distributions to be used to determine statistical significance. The minimum value of chi-squared for statistical significance is typically presented in a table or determined using a complex formula. This paper presents a simple linear model of the minimum value of chi-squared needed for statistical significance as a function of the degrees of freedom involved.

1. *Chi-squared in the everyday media*

Chi-squared is seldom – if ever – mentioned in the everyday media. But count tables often appear in the everyday media. The question arises as to whether the association between the row and column factors is dependent. Of course a dependence in a sample could also be due to sampling error. So the question is whether a dependence in the sample is statistically significant.

It is a simple matter to determine what counts would be expected in a given cell if the row and column variables were totally independent. And with some effort, one can calculate the deviation of the actual counts from these expected values. The most common method of calculating such deviations using the square of each deviation over the expected is called chi-squared. But determining whether the sum of these chi-squared deviations is statistically significant normally requires access to a table, a calculator, a computer or the web.

2. *Exact Cutoff Values for Chi-squared*

The minimum values of chi-squared needed for statistical significance are readily generated in Excel.

Equation 1:   **=ChiSq.INV.RT (0.05, degrees of freedom)**

These exact values are shown in an Excel spreadsheet: Schield (2014a).   Noted that these minimum values are a monotonically increasing function of the degrees of freedom.

3. *Models of the Cutoff Values for Chi-Squared*

Since the exact values are a monotonically increasing function of degrees of freedom, a linear model should be investigated. Three linear models were found to have good fits.

Equation 2:   **Model 1 Chi-squared Cutoff = 2.0 DF + 2**

Equation 3:   **Model II Chi-squared Cutoff = 1.5 DF + 4**

Equation 4:   **Model III: Chi-squared Cutoff = 1.2 DF + 8**

All three are conservative. These minimum values are sufficient for that value of chi-squared to be statistically significant. They are seldom necessary however. Consult a table, a formula or a statistician for exact values.

The first linear model gives has an error of less than 10% when degrees of freedom are less than five. The second has an error less than 10% when degrees of freedom are between three and 20. The third has an error of less than 10% when degrees of freedom are between 10 and 35.

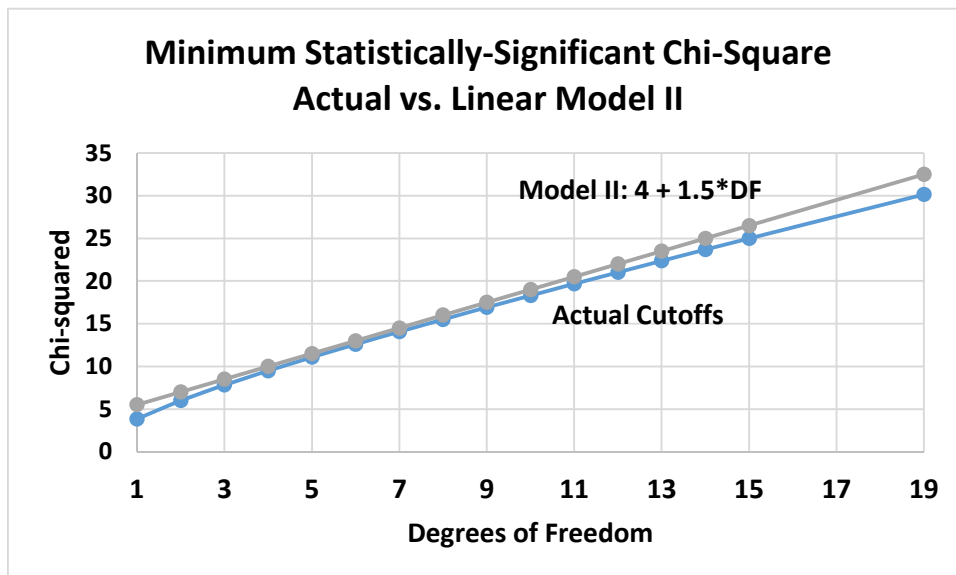The actual vs. the first model is shown in Figure 1.

Chi-Squared Cutoffs for Statistical Significance

Figure 1: Minimum Statistically-Significant Chi-Square: Actual vs. Model I

**Minimum Statistically-Significant Chi-Square
Actual vs. Linear Model I**

Model I: 2(DF+1)

Actual Cutoffs

Chi-squared
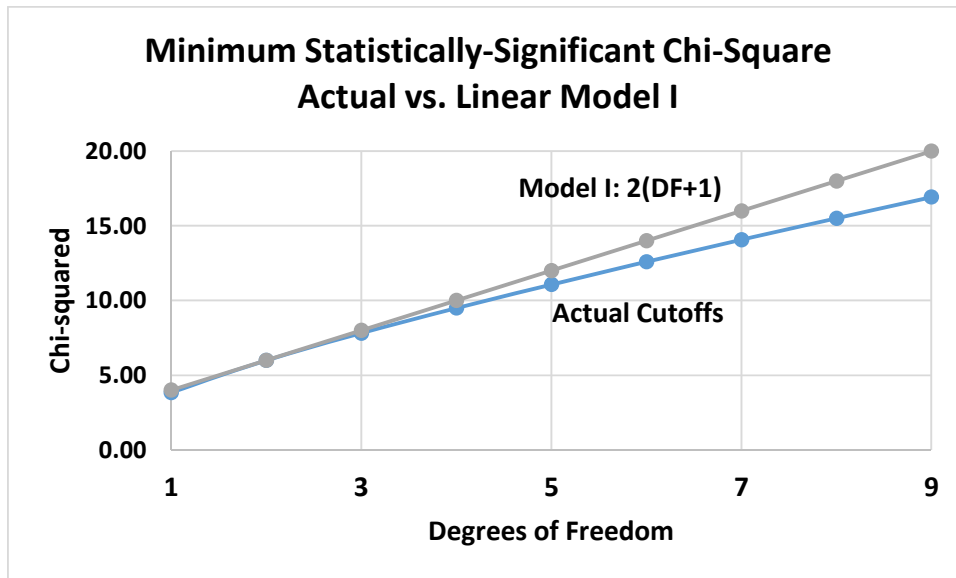
Degrees of Freedom

The actual versus the second model is shown in Figure 2.

Figure 2: Minimum Statistically-Significant Chi-Square: Actual vs. Model II

**Minimum Statistically-Significant Chi-Square
Actual vs. Linear Model II**

Model II: 4 + 1.5*DF

Actual Cutoffs

Chi-squared

Degrees of Freedom

Model two spans a wider range of the degrees of freedom than model one. But it is less accurate than model one when df = 1.

Another way to compare the actual versus the model is to plot the percentage error.

Figure 3: Minimum Statistically-Significant Chi-Square: Percentage Error of Actual vs. Model I

Minimum Statistically-Significant Chi-Square Error: Actual vs. Linear Model I

Model I: 2(DF+1)
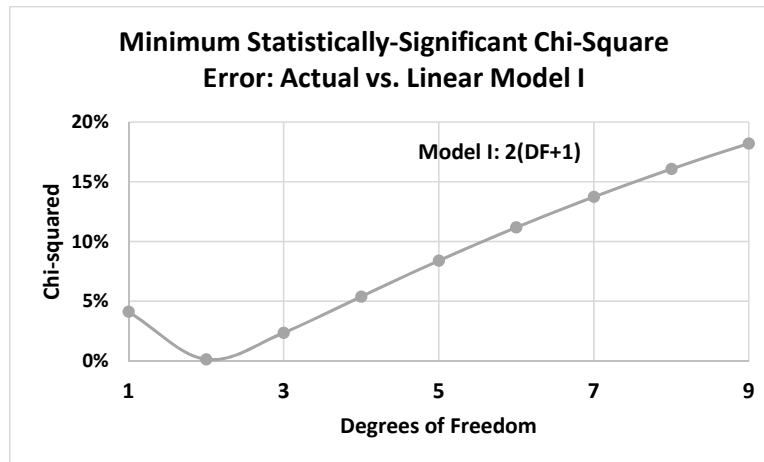
Figure 4: Minimum Statistically-Significant Chi-Square: Percentage Error of Actual vs. Model II

Minimum Statistically-Significant Chi-Square Error: Actual vs. Linear Model II

Model II: 4 + 1.5 *DF
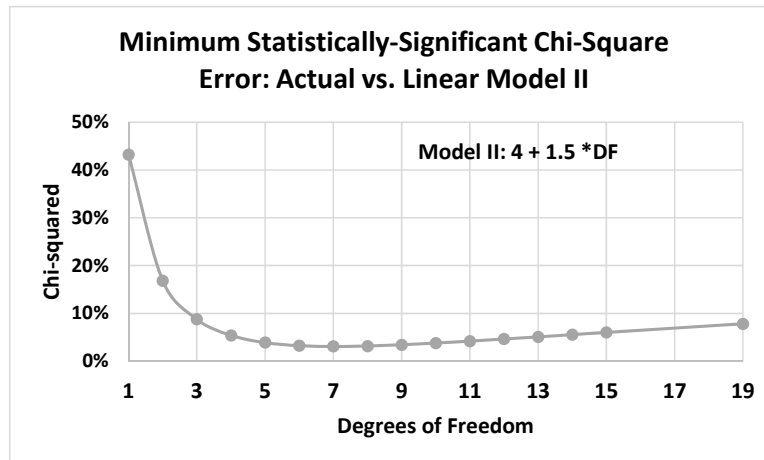
4. Conclusion

Note that the error in model I is smallest for the smallest degrees of freedom (df). For a 2x2 table, df = 1. For an RxC table, df = (R-1)*(C-1) = R*C - R – C + 1. For a 3x2 table, df = 2; for a 4x2 table, df = 3; and for a 3x3 table, df = 4. Since tables having less than 10 cells are quite common, df is commonly less than five. In those cases, this simple model should be most useful.

Since degrees of freedom are generally small (less than 10), model I appears to be the most useful of the three. Since it is simplest, it is also the most memorable of the three. This shortcut should be useful when introducing the idea of chi-squared in introductory statistics or statistical literacy for managers.

5. Bibliography:

Schield, Milo (2014). Statistically-Significance Chi-Squared Model
    www.StatLit.org/xls/2014-Schield-SS-Shortcut-Chi-Squared.xlsx

Schield, Milo (2014b). Chi-Squared Cutoffs for Statistical Significance.
    www.StatLit.org/xls/2014-Schield-SS-Shortcut-Chi-Squared.pdf