Log-Linear Analysis (Multi-way Frequency Tables)

by Melissa Moran


Click here for to get help with your Thesis or Dissertation.

Click here for FREE Thesis and Dissertation resources (templates, samples, calculators).

General Purpose

The log-linear analysis is appropriate when the goal of research is to determine if there is a statistically significant relationship among three or more discrete variables (Tabachnick & Fidell, 2012). It is typically used if none of the variables in the analysis are considered dependent variables, but rather all variables are considered variables of interest. Statistical significance exists if the calculated $\chi^2$ value is larger than the critical value, given the degrees of freedom and alpha value. The degrees of freedom are calculated using the following equation: $(r-1)(s-1)(p-1)$, where $r =$ number of levels of variable X, $s =$ number of levels of variable Y, and $p =$ number of levels of variable Z (Tabachnick & Fidell, 2012).

Data Level

Data can contain only discrete variables. Discrete variables are finite and contain a limited number of levels. Additionally, there is no smooth transition from one category to the next. Discrete data can be dichotomous, categorical, or ordinal. Examples include gender (male vs. female), colors (red vs. green vs. blue), and places in a race (first vs. second vs. third vs. fourth).

Assumptions

The assumptions of log-linear analysis will be assessed prior to analysis. The assumptions include that data must come from random samples of a multinomial, mutually exclusive distribution, adequate sample size, and the expected frequencies should not be too small. Traditional caution is that expected cell frequencies of less than five should not compose more than 20% of the cells, and no cell should have an expected frequency of less than one (Tabachnick & Fidell, 2012). Not meeting the assumption of expected cell frequencies will not increase the likelihood of Type I error, however it is expected to greatly decrease the overall power of the analysis (Tabachnick & Fidell, 2012). All observations must be independent of one another; participants can only contribute one observation to the data (Howell, 2010).

Caution

One danger in the use of log linear analysis is that too many variables be entered into the model, causing confusion in the interpretation of the results. To minimize this possibility, enter only variables you believe are related into the model and/or collapse the levels of variables when possible.

Questions it answers
What is the relationship between gender (male vs. female), favorite color (red vs. blue vs. green), and favorite cereal (rice pops vs. corn treats)?

What is the relationship between marital status (married vs. not married), presence of children (yes vs. no), and education (four year degree vs. less than four year degree)?

References