Use of Percentage Tables with Demographic Data in Log-Linear Modelling

Myoung-Jin Lee
Department of Sociology, Kookmin University
861-Chongneung-dong, Seongbuk-gu, 136-702
Seoul, Korea
mjlee@kookmin.ac.kr

1. Introduction

The traditional way to identify an association between categorical variables is to calculate percentages within categories of the independent variable and to compare these percentages across the categories of the independent variable. If the percentages differ significantly between or among the categories, a relationship is verified. One of the major disadvantages of the traditional crosstabular analysis is the difficulty of controlling other related variables (Fienberg, 1980; Knoke and Burke, 1980; Agresti, 1990).

As a response to the problems of the crosstabular analysis, loglinear models were introduced. These models are the techniques for applying regression procedures to categorical variables in contingency tables. When the more than two variables are involved, the loglinear model can be extended to the more powerful model for multivariate tables of categorical variables.

However there are several drawbacks in the process of moving the crosstabular analysis to the loglinear analysis. Even with a limited number of variables, examining the multivariate pattern among the variables becomes extremely difficult. The loglinear analysis often suffers from a lack of intuitive interpretation of its results (Kaufman and Schervish, 1986).

2. Purging Methods

In order to overcome these problems, Kaufman and Schervish (1986) and others proposed the adjusting or standardizing method to borrow strategies from the traditional crosstabular analysis as well as loglinear analysis. The method is to focus on a bivariate relationship of percentage tables after an acceptable log-linear model is identified and the effect of other variables are controlled. One may interpret the ‘true relationship’ between independent and dependent variables in the adjusted contingency tables (Clogg 1978; Clogg and Eliason 1988; Clogg et al. 1990; Liao, 1989; Xie, 1989). There are various types of the adjusting methods. Now let X, Y, and Z denote the independent, the dependent, and the control variables in a contingency table of $X \times Y \times Z$. Its cell frequency F_{ijk} can be described by the following loglinear model.

$$\log F_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{XZ} + \lambda_{ijk}^{XYZ} .$$

Clogg et al.(1990) mainly proposed the adjusting method to eliminate the effects of XZ interaction as well as XYZ when the three-way interaction is presented. Xie(1989) presented an alternative method to eliminate the effects of YZ interaction and XYZ interaction. He asserted that YZ pursing method has more desirable property. The adjusted or standardized percentages from the YZ purging method are invariant across marginal distributions of Z. He also argued that the Z variable is controlled as an exogenous variable. However the introduction of Z as a control variable does not always means to place Z as an exogenous variable on X-Y causal system. Sometimes, the X-Y causal relation may be reinforced by introducing Z as an intervening variable. It is not necessary to assume that the two pursing methods are exclusive and that one method has statistically better property. The decision to incorporate or controlling interaction effects should be guided by the result of loglinear analysis. One of main advantages of loglinear analysis provides appropriate statistical test for each parameter.

3. An Example

Throughout the example, we use an analysis of actual data on suicides as an application of the various purging method. According to statistics of the National Statistical Office (NSO, 2001), the suicide rate for Korean has greatly increased. The financial crisis left a crucial wound in Korean Society. Many people experienced unemployment, poverty, and family dismantle. A large number of literatures have reported that these experiences are related to the increase of suicide rates (Warr, 1982; Brenner & Levi, 1987; Iversen et al., 1987; Moser et al., 1984; Smith, 1988; Martkkainen, 1990; Cutler, 2000).

Tables 1-6 contain examples of percentage tables, log-linear analysis, and the uses of standardization or purging methods. The upper panel of each table (except Table 2) presents a bivariate table between (1) the year--- 1995 vs. 1998 (*X*) and (2) the sex—mean vs. men (*Y*). The lower panel of each table shows the result of adjusted frequencies, in which the original relationship between *X* and *Y* is reanalyzed within the categories of the control variable, the marital status (*Z*). As shown in Table 1, the original bivariate relationship seems to indicate the higher increase of suicides for women. On this three-variable example, one could easily use the subgroup analysis as well as the log-linear models. Table 2 shows parameters, coefficients, and standard errors under the saturated model.

Tables 3-6 present the standardized percentages based on selective percentage-standardization methods (Clogg, 1978; Clogg and Eliason, 1988; Clogg et al., 1990). Direct standardization in Table 3 refers to the method to impose the same marginal distributions of the control variable (*Z*) to each category of the independent variable (*X*). Here we choose arithmetic average of the compositional distributions across the categories as ‘a reference group’. Note that one may choose a different category to which other categories should be compared. (See, Clogg et al., 1990.) Choosing a different reference group may lead to different results of standardized percentage. This result reveals that increase difference between men and women still exists. Only difference is that there is no change of suicides between 1995 and 1998 for men.

In Table 4, adjusted frequencies are obtained after partial interaction between *X-Z* and *X-Y-Z* is removed. There is still increase difference between men and women, but the amount of suicides increase reverses its direction compared to the result from original table or adjusted table from direction standardization. In Table 5, the partial *YZ* and *XYZ* purging method is used. Its adjusted frequencies lead to similar results: higher suicides increase for men. Now each category of control variable has the same pattern of bivariate relationship. Partial *XZ*, *YZ*, and *XYZ* purging method in Table 6 makes adjustment for *XZ*, *YZ*, and *XYZ* interactions. The method adds adjustment for a two-way interaction between YZ and a three-way interaction XYZ. Note that the adjusted frequencies here depend only on marginal distributions and X-Y interaction.

As shown in Table 3-6, there are several methods to control or eliminate possible interaction to confound the interpretation of the relationship between independent and dependent variable. Furthermore adjusted frequencies depend on the choice of reference group. It is desirable to compare results from such purging methods to validate one' s interpretation.

<Tables 1-6 about here >

References

Agresti, Alan. 1990 *Categorical Data Analysis*. New York: Wiley.

Clogg, Clifford C. 1978. "Adjustment of Rates Using Multiplicative Models." *Demography* 15:523-539.

_____ and Scott R. Eliason. 1988. "A Flexible Procedure for Adjusting Rates and Proportions, Including Statistical Methods for Group Comparisons." *American Sociological Review* 53:267-283.

_____, James W. Shokey, and Scott R. Eliason. 1990. "A General Statistical Framework for Adjustment o Rates." *Sociological Methods and Research* 19:156-95.

Liao, Tim Futing. 1989. "A Flexiable Approach for the Decomposition of Rate Differences." *Demography* 26:717-26.

Kaufman, Robert L., and Paul G. Shervish. 1986. "Using Adjusted Crosstabulation to Interpret Loglinear Relations." *American Sociological Review* 51:717-33.

Knoke, David and Peter J. Burke. 1980. *Log-Linear Models*. Beverly Hills: Sage Publication.

National Statistical Office. 2001. *http://www.nso.go.kr.*

Xie, Yu, 1989. "An Alternative Purging Method: Controlling the Composition-Dependent Interaction in an Analysis of Rates." *Demography* 26:711-16.