Generalized marginal homogeneity model and its relation to marginal equimoments for square contingency tables with ordered categories

Kouji Tahata · Sadao Tomizawa

Received: 26 February 2008 / Revised: 14 August 2008 / Accepted: 16 September 2008 / Published online: 29 October 2008 © Springer-Verlag 2008

Abstract For square contingency tables with ordered categories, Tomizawa (Calcutta Stat Assoc Bull 43:123–125, 1993a; Sankhyā Ser B 60:293–300, 1998) gave theorems that the marginal homogeneity (MH) model is equivalent to certain two or three models holding simultaneously. This paper proposes a generalized MH model, which describes a structure of the odds that an observation will fall in row category $i$ or below and column category $i+1$ or above, instead of in column category $i$ or below and row category $i+1$ or above. In addition, this paper gives the theorems that the MH model is equivalent to the generalized MH model and some models holding simultaneously whose each indicates: (1) the equality of $m$-order moment of row and column variables, (2) the equality of skewness of them and (3) the equality of kurtosis of them. When the MH model fits the data poorly, these may be useful for seeing the reason for the poor fit; for instance, the poor fit of the MH model is caused by the poor fit of the equality of row and column means rather than the generalized MH model. Examples are given.

Keywords Generalized marginal homogeneity · Kurtosis · Marginal homogeneity · Moment · Ordered category · Skewness · Square contingency table

Mathematics Subject Classification (2000) 62H17

JEL Classification C39
1 Introduction

Consider a square contingency table with the same ordinal row and column classifications. For example, consider the data in Table 1 in which each observation is a pairing of father’s occupational status with his son’s occupational status (see Sect. 4). For such data, generally, many observations tend to fall in (or near) the main diagonal cells, and thus the independence between the row and column classifications is unlikely to hold. One of our interests is whether or not there is a structure of symmetry (rather than independence) in the table. For example, we are interested in (1) the structure of symmetry of cell probabilities which indicates that the probability that a father’s status is $i$ and his son’s status is $j$ ($\neq i$) equals the probability that the father’s status is $j$ and his son’s status is $i$, and also (2) the structure of marginal homogeneity (MH) which indicates that the probability that a father’s status is $i$ equals the probability that his son’s status is $i$.

On the analysis of symmetry, the symmetry model is described by, e.g., Bowker (1948), Caussinus (1965), and Bishop et al. (1975, p. 282). The MH model is described by, e.g., Stuart (1955), Bishop et al. (1975, p. 293), and Agresti (1984, p. 207). Note that the symmetry model implies the MH model.

When the symmetry and MH models do not hold, many statisticians may be interested in applying various asymmetry models; for example, McCullagh (1978), Goodman (1979a), Agresti (1984, p. 205), Tomizawa et al. (1998), Miyamoto et al. (2005), and Yamamoto et al. (2007). Especially, Tomizawa (1993b, 1995) considered the extended marginal homogeneity models (see Sect. 2). These models indicate that the log-odds of cumulative probabilities are expressed as a constant plus a term that is linear in category levels. If these models do not hold, we are interested in applying a more generalized MH model. In addition, we are then interested in a comparison of the two marginal distributions of categorical row and column variables, say $X_1$ and $X_2$, respectively (e.g., for the data in Table 1, in a comparison of the marginal distributions of a father’s status $X_1$ and his son’s status $X_2$), and in whether $X_1$ tends to be stochastically less than $X_2$ or vice versa. Thus we propose here a more generalized MH model which indicates that the log-odds of cumulative probabilities are expressed as polynomial functions of category levels. The polynomial function in this model would be useful for seeing in more details the relationships between the cumulative marginal probabilities of $X_1$ and $X_2$.

Miyamoto et al. (2005) described the relation between the MH model and the marginal cumulative logistic (ML) model [or the conditional ML (CML) model]. See Agresti (1984, p. 205) and “Appendix 1” for these models. Tomizawa (1993a, 1998) considered decompositions for the MH model into two or three models. These show, for instance, that the MH model holds if and only if the ML model and the equality of row and column means hold simultaneously (see “Appendix 1”).

When these decomposed models fit the data poorly, we are further interested in a generalization of the decompositions for the MH model such that the MH model is equivalent to more (than three) models holding simultaneously. It may be useful for analyzing the structure of the MH model in more details for the data, and for seeing in more details the reason for the poor fit when the MH model fits the data poorly.