The Homogeneity Test of Multiply Matched Case-Control Studies with Covariates

Minjung Kwak
Department of Computer Science and Statistics, Pyongtaek University
111 Yongyi-Dong, Pyongtaek, Korea
mjkwak@ptuniv.ac.kr

1. Introduction

In matched case-control study, the conditional likelihood analyses based on the linear logistic equation enable one to model the effects of covariates and facilitate the examination of a possible trend in risk with increasing exposure, while preserving the original matching. It assumes that the odds ratios are constant across the strata. However, apparent heterogeneity of the odds ratio is present when some covariates are involved with matching variables. Based on random effect model, we propose a score test of homogeneity which allows to adjust covariates.

2. The Conditional logistic model

Let y_{ij} be denote a binary outcome variable for the jth of N_i subjects in the ith of K matched sets. Let $\mathbf{x}_{ij}$ be a p vector of concomitant explanatory variables with $\mathbf{x}_{ij} = \{x_{ij}^{(1)}, x_{ij}^{(2)}, \cdots, x_{ij}^{(p)}\}$. Suppose the observations in the ith matched set are arranged so that the first n_i covariate vectors correspond to events and the remainder do not. Then the conditional probability of the observed outcome, given the number of events n_i is

$$L_i(\beta) = \frac{\exp(\sum_{j=1}^{n_i} \mathbf{x}_{ij}'\beta)}{\sum_{\mathbf{l}\in R(n_i,m_i)} \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}'\beta)} \ ,$$

where $\mathbf{l}$ range over the set $R(n_i, m_i)$ of partitions of the integers $\{1, 2, \ldots, N_i\}$ into two subsets of size n_i and $m_i = N_i - n_i$. We construct the random effect conditional logistic model to adjust the heterogeneity of the odds ratio for the covariate which is involved with matching variables. In the situation that the number of strata K is large and the size of strata is small, random effect models are preferred. Consider the random effect conditional logistic model which is related with vth covariate as follows;

$$f_i(\beta, \theta) = \frac{\exp(\sum_{j=1}^{n_i} \mathbf{x}_{ij}'\beta + \theta_i x_{ij}^{(v)})}{\sum_{\mathbf{l}\in R(n_i,m_i)} \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}'\beta + \theta_i x_{il_j}^{(v)})}$$

$$l_i(\beta, \theta) = \int f_i(\beta, \theta) dF.$$

The θ_i are regarded as iid from some unknown distribution F with mean zero and variance σ^2 and the null hypothesis of homogeneity could be formulated as $\sigma^2 = 0$.

3. The Score test of homogeneity

The score statistic for testing the hypothesis of homogeneity is

$$U = \sum_{i=1}^{K} \frac{\partial^2 \log l_i}{\partial \theta^2}(\hat{\beta}, 0) = \frac{1}{2} \sum_{i=1}^{K} \left\{ \left[\frac{\partial \log f_i}{\partial \theta_i}(\hat{\beta}) \right]^2 + \frac{\partial^2 \log f_i}{\partial \theta_i^2}(\hat{\beta}) \right\}$$

where $\hat{\beta}$ are the maximum likelihood estimator of β under the null hypothesis. We obtain

$$\frac{\partial \log f_i}{\partial \theta_i} = \left\{ \sum_{j=1}^{n_i} x_{ij}^{(v)} - \sum_{l} A_{il_j}^{(v)} \right\} \quad \text{and} \quad \frac{\partial^2 \log f_i}{\partial \theta_i^2} = -\sum_{l} \left\{ C_{il_j}^{(v)} - A_{il_j}^{2(v)} \right\}$$

where

$$A_{il_j}^{(v)} = \frac{x_{il_j}^{(v)} \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}' \beta)}{\sum_l \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}' \beta)} \quad \text{and} \quad C_{il_j}^{(v)} = \frac{x_{il_j}^{2(v)} \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}' \beta)}{\sum_l \exp(\sum_{j=1}^{n_i} \mathbf{x}_{il_j}' \beta)}$$

The variance of U under the null hypothesis is asymptotically

$$I = I_{\theta\theta} - I_{\theta\beta} I_{\beta\beta}^{-1} I_{\theta\beta}'$$

where

$$I_{\theta\theta} = \sum_{i=1}^{K} \left[\frac{\partial \log l_i}{\partial \theta} \right]^2 = \frac{1}{4} \sum_{i=1}^{K} \left\{ \left(x_{ij}^{(v)} - \sum_{l} A_{il_j}^{(v)} \right)^2 - \sum_{l} \left(C_{il_j}^{(v)} - A_{il_j}^{2(v)} \right) \right\}^2$$

$$I_{\beta\beta} = \sum_{i=1}^{K} \left[\left(\frac{\partial \log l_i}{\partial \beta} \right) \left(\frac{\partial \log l_i}{\partial \beta} \right)' \right] = \sum_{i=1}^{K} \left[\left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} - \sum_{l} A_{il_j} \right) \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} - \sum_{l} A_{il_j} \right)' \right]$$

$$I_{\theta\beta} = \sum_{i=1}^{K} \left[\left(\frac{\partial \log l_i}{\partial \theta} \right) \left(\frac{\partial \log l_i}{\partial \beta} \right) \right] = \frac{1}{2} \sum_{i=1}^{K} \left[\left\{ \left(x_{ij}^{(v)} - \sum_{l} A_{il_j}^{(v)} \right)^2 - \sum_{l} \left(C_{il_j}^{(v)} - A_{il_j}^{2(v)} \right) \right\} \left(\sum_{j=1}^{n_i} \mathbf{x}_{ij} - \sum_{l} A_{il_j} \right)' \right]$$

Then the score statistics $S = U/\sqrt{I}$ has asymptotically normal distribution.

REFERENCES

Commenges, D., Letenneur, L., Jacqmin, H., Moreau, T. and Dartigues. J. (1994) Test of homogeneity of binary data with explanatory variables. *Biometrics*, **50**, 613-620.

Commenges, D. and Anderson, P. (1995) Score test of homogeneity for survival data. *Lifetime Data Analysis*, **1**, 145-156.

Gray, R. (1995) Tests for Variation over groups in survival data. *JASA*, **90**, 198-203.

Liang, K. (1987) A locally most powerful test for homogeneity with many strata. *Biometrika*, **74**, 259-264.

Liang, K. and Self, S. (1985) Tests for homogeneity of odds ratio when the data are sparse. *Biometrika*, **72**, 353-358.