Logistic regression models in small area investigations

Krystyna Pruska

Chair of Statistical Methods
University of Lodz
Poland
kpruska@uni.lodz.pl

Different qualitative variables are considered in statistical investigations. Some of them have only two variants of values (categories). We can assume that these variables have Bernoulli distribution. Logistic regression models are applied to the analysis of such variables. It means these models are used to the analysis of binary data.

If we consider a population analysed with respect to Bernoulli variable and some auxiliary variables then we can construct logistic regression models for this population.

In this paper we consider a population divided into M small areas: $A_1,...,A_M$. We assume that Y is an investigated variable and x_i is a vector of auxiliary variables for i-th small area. Moreover, a distribution of Y is given by the function:

$$P(Y=1|A_i) = \theta_i \text{ and } P(Y=0|A_i) = 1 - \theta_i \text{ for } i = 1,...,M, \quad (1)$$

where θ_i is unknown.

We construct the following logistic regression model:

$$L^{-1}(p_i) = x_i'\alpha + \varepsilon_i \quad \text{for } i = 1,...,M, \quad (2)$$

where

$$L^{-1}(p_i) = \ln \frac{p_i}{1-p_i} \quad (3)$$

and p_i is an estimator of parameter θ_i, α is the model parameter, ε_i is a random error, $E(\varepsilon_i) = 0$.

We draw m small areas from M small areas. We determine the values of estimators p_i for drawn areas on the basis of small area sample ($i = 1,...,m$).We estimate the parameters of model (2) on the basis of data for drawn areas. Next we determine the estimates of θ_i for undrawn areas on the basis of model (2).

We can consider different sampling methods and different estimators of θ_i for drawn small areas.

In this paper an estimation of variance of estimator p_i ($i = 1,...,M$) is considered, too. The simulation experiments are conducted for this purpose. The estimator p_i is determined on the basis of small area sample for drawn m areas and on the basis of model (2) for other areas.