





-  **Current Issue**
-  **Browse Issues**
-  **Search**
-  **About this Journal**
-  **Instruction to Authors**
-  **Online Submission**
-  **Subscription**
-  **Contact Us**
-  **RSS Feed**

Acta Medica Iranica  
2009;47(4) : 41-45

### Original Article

#### Using Latent Variables to Eliminate Multicollinearity Effect in A Logistic Regression on Risk Factors for Breast Cancer

Pourhosseingholi MA: Biostatistician, Shaheed Beheshti University of Medical Sciences,  
Mehrabi Y: Associate professor of Medical Statistics, Shaheed Beheshti University of Medical Sciences,  
Alavi-Majd H: Assistant professor of Biostatistics, Shaheed Beheshti University of Medical Sciences,  
Yavari P: Associate professor of Epidemiology, Shaheed Beheshti University of Medical Sciences,

#### Corresponding Author:

Mehrabi Y: Associate professor of Medical Statistics, Shaheed Beheshti University of Medical Sciences,  
E-mail: [mehrabi@sbmu.ac.ir](mailto:mehrabi@sbmu.ac.ir) ; [ymehrab@gmail.com](mailto:ymehrab@gmail.com)

Received: January 16,2006  
Accept : April 23,2006  
Available online: January 5,2009

#### Abstract:

**Background and Objectives:** Logistic regression is one of the most widely used generalized linear models for analysis of the relationships between one or more explanatory variables and a categorical response. Strong correlations among explanatory variables (multicollinearity) reduce the efficiency of model to a considerable degree. In this study we used latent variables to reduce the effects of multicollinearity in the analysis of a case-control study.

**Methods:** Our data came from a case-control study in which 300 women with breast cancer were compared to 300 controls. Five highly correlated quantitative variables were selected to assess the effect of multicollinearity. First, an ordinary logistic regression model was fitted to the data. Then, to remove the effect of multicollinearity, two latent variables were generated using factor analysis and principal components analysis methods. Parameters of logistic regression were estimated using these latent as explanatory variables. We used the estimated standard errors of the parameters to compare the efficiency of models.

**Results:** The logistic regression based on five primary variables produced unusual odds ratio estimates for age at first pregnancy (OR=67960, 95%CI: 10184-453503) and for total length of breast feeding (OR=0). On the other hand, the parameters estimated for logistic regression on latent variables generated by both factor analysis and principal components analysis were statistically significant (P<0.003). The standard errors were smaller than with ordinary logistic regression on original variables. The factors and components generated by the two methods explained at least 85% of the total variance.

**Conclusions:** This research showed that the standard errors of the estimated parameters in logistic regression based on latent variables were considerably smaller than that of model for original variables. Therefore models including latent variables could be more efficient when there is multicollinearity among the risk factors for breast cancer.

#### Keywords:

[Multicollinearity](#) . [Latent Variables](#) . [Factor Analysis](#) . [Principal Components Analysis](#) . [Logistic Regression](#) . [Breast Cancer](#)

TUMS ID: 12526

Full Text HTML  Full Text PDF  306 kB

top ▲