

A Note on Dichotomization of Continuous Response Variable in Presence of Measuring Errors

Yue Shentu and Minge Xie

Rutgers University

SUMMARY

The purpose of this note is to raise awareness of the complexity of the practice involving dichotomization. It is well known that the regular regression models are effective tools for analyzing Gaussian type response variables, and researchers are often told that it is a “bad idea” to use dichotomization if continuous measurements are available. We demonstrate through special cases, however, that there is the other side of the story if the response variable is contaminated. Although dichotomization causes loss of information, it can also reduce input of contamination. If the reduction of contamination input outweighs the loss of information, analysis based on dichotomization can sometimes provide better results. We derive formulas of bias and variance for binary regression estimators under a contamination model of unknown additive errors, and compare them with both the least squares and robust M estimators from the corresponding linear regression analysis using continuous responses. As a case study, we study extensively the case in which the observed response is contaminated by an error with a mean and a variance proportional to the mean and variance of the uncontaminated true response. Conditions under which dichotomization is preferred are obtained. A simulation study based on a real data setting is provided, which supports the theoretical developments.

KEY WORDS: Linear regression model; Logistic regression; Measurement error in responses; Parameter Estimation; Quantal regression; Robust method

Running Title: Dichotomization of Response Variable in Regressions