A comparison of logit and probit models for a binary response variable via a new way of data generalization

Özge Akkuş[1], Atilla Göktaş[1] and Selen Çakmakyapan[2]

[1]*Muğla University, Turkey*
[2]*Istanbul Medeniyet University, Turkey*

Abstract

Logit and probit models are two members of generalized linear models family that are widely used especially when the dependent variable is observed to be binary. The properties that make a difference for these two models for the same data set are resulted from the assumptions they use and their mathematical functions. There is no study specifying a certain judgement on the preference of these models to make a decision which model is better in what condition. In this study, a new data generalization technique has been proposed for the simulation study conducted to make a comparison of the model fits to binary logit and probit models for the generated data set under certain conditions to reach an end to which condition is better.

In the process of the simulation study, a dependent and explanatory variables are generated from multivariate normal distribution which is very much different from the ordinary generating procedure. As is already known, this procedure uses the information of the interested model itself. Hence the generation of this type would always be in favor of the interested model not the alternative and there would be no sense to make a comparison from such data generalization. In the proposed generating process since the generated dependent variable is always continuous, it should be classified as binary to make the dataset usable for logit and probit models. After fitting logit and probit models to the generated data sets, goodness-of-fit-test results related to both models, residuals, deviances and some pseudo R^2's used for binary dependent variables have been obtained to make significant comparisons. These procedures have been performed for two different cut points used to classify response variables, three different relationship levels among variables (high, medium, none) and five different sample sizes. For each cut point, relationship level and sample size the simulation has been replicated for a thousand time. Since the obtained estimated probabilities from both models are considerably close, it is found that there has been no statisticaly significant difference among most pseudo R^2's. However, when the residuals are taken into account, probit model has a priority to be used for a sample size that is less than 200, whereas the Logit model is superior for a sample size that is greater than 200. Another remarkable finding is that the different cut-off levels and relationship have not any effect on the choice of the model.

Keywords

Binary logit, Binary probit, Pseudo R-square, Deviance.

References

[1] Agresti, A. (2002). *Catagorical Data Analysis* (Second Edition). Wiley, New Jersey.

[2] Aldrich, J.H. and F.D. Nelson (1984). *Linear Probability, Logit, and Probit Models.* (pp. 397–402). Sage Publications, London.

[3] Anderson-Sprecher, R. (1994). Model comparisons and R squares. *Amer. Statist. 48*, 113–117.

[4] Cameron, A.C. and A.G. Windmeijer (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econometrics 77*, 329–342.

[5] Cox, D.R. and N. Wermuth (1992). A comment on the coefficient of determination for binary responses. *Amer. Statist. 46*, 1–4.

[6] Hosmer, D.W., T. Hosmer, S. Le Cessie, and S. Lemeshow (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Stat. Med. 16*, 965–980.

[7] Uçar, Ö. (2004). Nitel Verilerin Analizinde Lojit ve Probit Modeller, *Yükseklisans Tezi, Hacettepe Üniversitesi*, Ankara-2004.

[8] Veall, M.R. and K.F. Zimmermann (1994). Evaluating pseudo-R^2's for binary probit models. *Quality & Quantity 28*, 151–164.

[9] Veall, M.R. and KF. Zimmermann (1996). Pseudo-R^2's measures for some common limited dependent variable models. *Sunderforschungsbereich 386*, 1–34.

[10] Windmeijer, F.A.G. (1995). Goodness of fit measures in binary choice models. *Econometric Rev. 14*, 101–116.

[11] Winkelmann, R. and S. Boes (2006). *Analysis of Microdata.* Springer, Berlin.

[12] Zelner, B.A. (2008). Using simulation to interpret and present logit and probit results. *Working paper.*