Applied Mathematical Sciences, Vol. 12, 2018, no. 4, 185 - 204
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2018.818

A Comparison of Univariate Probit and Logit Models Using Simulation

Abeer H. Alsoruji

Department of Statistics, Faculty of Sciences
King Abdulaziz University, Jeddah, Saudi Arabia

Sulafah Binhimd

Department of Statistics, Faculty of Sciences
King Abdulaziz University, Jeddah, Saudi Arabia

Mervat K. Abd Elaal

Department of Statistics, Faculty of Sciences
King Abdulaziz University, Jeddah, Saudi Arabia
&
Department of Statistics, Faculty of Commerce
Al-Azhar University, Cairo, Egypt

Abstract

Predictive analytics techniques are widely used in the application field, and the most common of these is fitting data with functions. The aim of function fittings is to predict the value of a response, by combing the regressors. Univariate probit and logit models are used for the same purposes when the response variable is binary. Both models used applied for the estimation of the functional relationship between response and regressors. The question of which model performs better comes to the mind. For this aim, a Monte Carlo simulation was performed to compare both the univariate probit and logit models under different conditions. In In this paper we considered the simulation of, employing latent variable approach with different sample sizes, cut points, and different correlations between response variable and regressors were taken into account. To make a comparison between univariate logit and probit models, Pearson residuals, deviations, Hosmer

and Lemesshow, area under Receiver Operating Characteristic (ROC) curve, and Pseudo-R square statistics which are used for qualitative data analysis, were calculated and the results were interpreted.

Keywords: Univariate probit model; Univariate logit model; Latent variable; Monte Carlo simulation; Goodness-of-fit statistics

1 Introduction

Predictive analytics techniques are widely used in the application field, and the most common of these is fitting data with functions. The aim of function fittings is to predict the value of a response, by combing the regressors. In many studies variables of interest are binary and the adapted techniques to deal with these case are univariate probit and logit models.

Univariate probit and logit models are members of the family of generalized linear models (GLM) and analyze the relationship between regressors and binary response variable. The formula of univariate probit model with p regressors [6] is equal:

$$\pi = \Pr(Y = 1) = \Phi\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_{\mathrm{p}} x_{\mathrm{p}}\right) \tag{1}$$

Here f is univariate Logit model [1]:

$$\pi = \Pr(Y = 1) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_p x_p)}} \tag{2}$$

where $\Phi(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)$ is the cumulative standard normal distribution function,π is the probability of an event is present that depends on p regressors, $x_1, \ldots, x_p)$, β_0 is the coefficient of the constant term, and $\beta_1, \ldots, \beta_p$ are the coefficients of the p regressors. The coefficients $\beta_0, \beta_1, \ldots, \beta_p$ in equation s(1) and (2) are estimated using MLE [5]. In this paper we considered the simulation of the latent variable approach with different sample sizes, cut points and different correlations between response variables and regressors. To compare the models, Pearson residuals, deviations, Hosmer and Lemesshow, the area under Receiver Operating Characteristic (ROC) curve and Pseudo-R square statistics (used for qualitative data analysis) were calculated, and the results interpreted.

In the next section in this paper, basic concepts of latent variable model for univariate probit and logit models. The goodness-of-fit tests which are the basis of our comparison between the two models displayed in Section 3. He simulation performed for different sample sizes, different correlations between variables, and different cut points for latent response variable and it's results in Section 4. And the conclusion in Section 5.

2 A Latent Variable Model for Univariate Probit and Logit Models

The response variable in univariate probit and logit models have only two categories. The occurrence and nonoccurrence of events are the categories in the response variables. Univariate probit and logit models assume an underlying response variable defined as which can be presented as Y^* a functional

relationship as follow:

$$Y^* = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon \,. \tag{3}$$

Here Y^* is unobserved or a latent variable ranging from $-\infty$ to ∞ that generates the observed variable Y. The larger values of Y^* are classified as $Y = 1$, while those with smaller values of Y^* values are observed as $Y = 0$. The latent variable is assumed to be linearly dependent to the observed regressors throughout the structural model in equation (3). Y^* is related to the observed binary variable Y with the equation below:

$$Y_i = \begin{cases} 1, Y_i^* > \tau \\ 0, Y_i^* < \tau \end{cases} ; \tag{4}$$

Where τ is the cut point and $i = 1,2, \ldots, n$ are independent observations obtained from n subjects [7]. If Y_i^* is greater than τ then then $Y_i = 1$,. If Y_i^* is less than τ then then $Y_i = 0$.

The latent variable model for binary outcomes is illustrated in Figure (1). Figure (1) is a figure from [7] and considered here for clarification.

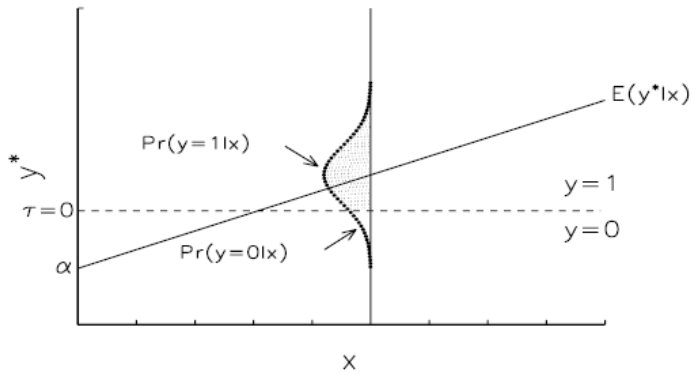


Figure 1 Relationship between latent variable Y^* and $Pr(Y = 1)$.

For a given value of x and $\tau = 0$;

$$\Pr(Y = 1|x) = \Pr(Y^* > 0|x)\,, \tag{5}$$

Substituting the structural model in equation (3) with one regressor and rearranging equation (5)

$$\Pr(Y = 1|x) = \Pr(\varepsilon > -[\beta_0 + \beta_1 x]|x)\ . \tag{6}$$

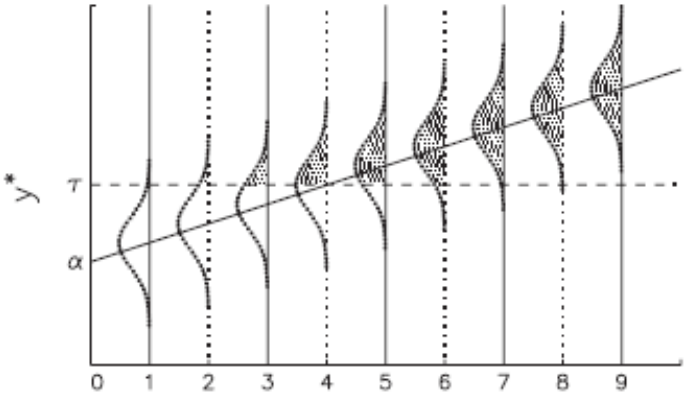
This equation shows that the probability depends on the distribution of the structural error ε. Two distributions of the structural ε are commonly assumed, both with an assumed mean of 0. First, the structural error ε is assumed to be distributed normally with $V(\varepsilon) = 1$. This leads to the univariate probit model:

$$P(Y = 1|x) = \Phi(\beta_0 + \beta_1 x)$$

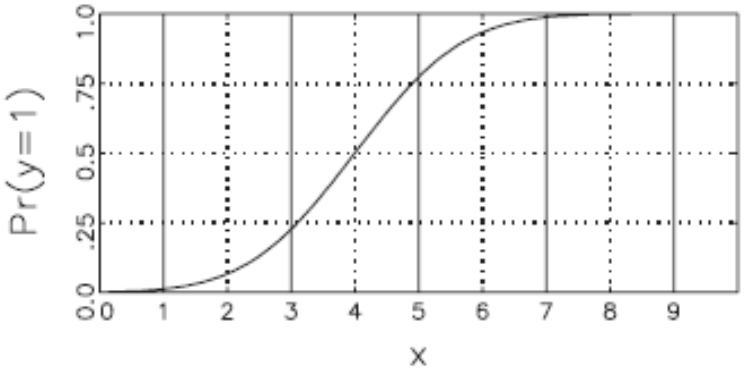
Alternatively, the structural error ε is assumed to be distributed logistically with $V(\varepsilon) = \frac{\pi^2}{3}$, leading to the univariate logit model:

$$P(Y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

In general for both models, the probability of the event occurring is the cumulative density function (CDF) of the structural error ε evaluated at given values of the regressors. F is the CDF of standard normal distribution for the probit model and the CDF of standard logistic distribution for the logit model. The relationship between the linear latent variable model and the resulting nonlinear probability model is shown in Figure (4.2) [Scott Long (2001)] for a model with a single regressor [2], which is shown in [7].



(a)



(b)

Figure 2 Relationship between the linear model $Y^* = \beta_0 + \beta_1 x + \varepsilon$ and the nonlinear probability model $\Pr(Y = 1|x) = F(\beta_0 + \beta_1 x)$: (a) plot of Y^*, (b) plot of $\Pr(Y = 1|x)$.

Figure (2)) (a) shows the structural error distribution for nine values of *x*, which labeled 1 , . The area where $Y^* > \tau$ corresponds to $\Pr(Y = 1|x)$ and has been shaded. Figure (2) (b) plots $\Pr(Y = 1|x)$ corresponding to the shaded regions in (a). As moving from 1 to 2, only a portion of the thin tail crosses the cut point in (a), resulting in a small change in $\Pr(Y = 1|x)$ in (b). As moving from 2, to to 4, thicker regions of the structural error distribution slide over the cut point and the increase in $\Pr(Y = 1|x)$ becomes larger. The resulting curve is the well-known S-curve associated with the binary response model.

3 Goodness-of-fit Measures

The goodness of fit measures of a statistical model describe how well it fits a set of data. There are several methods for assessing the fit of an estimated univariate probit and logit model in equation (1) and (2), and show how effectively the model describes the response variable. Some of these methods are goodness-of-fit such as Pearson chi-square, deviance, Hosmer and Lemeshow, Pseudo-R square, and area under ROC curve.

3.1. Pearson Chi-square χ^2 Goodness of Fit Test

Pearson's Chi-square test works well when the regressors (covariates) are categorical. When one or more regressors are continuous, the disadvantages of Pearson chi-square test provide incorrect p-values. Pearson Chi-Square χ^2goodness-of-fit test obtained by comparing the overall difference between the observed and fitted values.

First, define items that needed to describe this test: suppose the fitted model has p regressors (covariates) $\boldsymbol{X} = (x_1, \dots, x_p)$. A single covariate pattern is defined as a single set of values of the regressors (covariates) used in the model. For example, if two regressors (covariates), gender (Female and Male) and class (smoking, non-smoking) are used in a data set, then the maximum number of distinct covariate patterns is four. There are basically two types of covariate patterns. Type one pattern, there are no tied regressors (covariates) which indicates that each subject has a unique set of regressor values, and the number of covariate patterns J is equal to the number of subjects, i.e., $J \approx n$). This is very common, when only continuous regressors are involved in the univariate probit and logit models and this is the case of our simulation in this paper. For the type two pattern, some subjects have tied regressors (covariates), that is, they have the same regressor values, making the number of covariate patterns less than the number of subjects, i.e., $J < n$). In the type two pattern, let the total number of success $(y = 1)$ be n_1 and the total number of failures $(y = 0)$ be n_0 , and it follows that $n_1 + n_0 = n$. Suppose m_j subjects in the j^{th} covariate pattern has the same regressor values, then, $\sum_{j=1}^{J} m_j = n$. Let y_{j1} denote the number of successes in the j^{th} group with j^{th} covariate pattern. It follows that $\sum_{j=1}^{J} y_{j1} = n_1$. Similarly, let y_{j0} denote the number of failures observed by subjects in the j^{th} group with j^{th} covariate pattern. It follows that $\sum_{j=1}^{J} y_{j0} = n_0$. Under type two covariate pattern, the binary response variable can be represented by a J by 2 frequency table. The two columns of the table corresponding to the response variable, $y = 1,0$ and the J rows corresponding to J possible covariate patterns [8].

The likelihood function and the log-likelihood function can be written respectively as below for type two covariate pattern

$$L = \prod_{j=1}^{J} \binom{m_j}{y_{j1}} (\pi_j)^{y_{j1}} (1-\pi_j)^{m_j - y_{j1}}. \quad (7)$$

$$LogL = \sum_{j=1}^{J} \{log \binom{m_j}{y_{j1}} + y_{j1} log\pi_j + (m_j - y_{j1})\log(1-\pi_j)\} \quad (8)$$

Let $\hat{\pi}_j$ be the maximum likelihood estimate of π_j associated with the j^{th} covariate pattern, then the expected number of success observed by the subject in the j^{th} group with j^{th} covariate pattern is

$$\hat{y}_{j1} = m_j \hat{\pi}_j \quad (9)$$

So The Pearson Chi-square test statistic is

$$\chi^2 = \sum_{j=1}^{J} pr(y_{j1}, \hat{y}_{j1})^2 \quad (10)$$

Where $pr(y_{j1}, \hat{y}_{j1})$ is the **Pearson residual** defined as:

$$pr(y_{j1}, \hat{y}_{j1}) = \frac{(y_{j1} - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1-\hat{\pi}_j)}} \quad (11)$$

Hence, Pearson $\chi 2$ statistic is summing the square of all observation Pearson residuals. The Pearson Chi-square test statistic in equation (10) follows a Chi-Square distribution with degrees of freedom degrees of freedom $J -$ number of the parameters in the model.

3.2 Deviance D Goodness of Fit Test

The deviance goodness of fit test is based on a likelihood ratio test of reduced model $E(y_j) = \hat{\pi}_j$ against the full model $E(y_j) = \pi_j \quad j = 1,2,\dots,J$ where π_j are parameters, $j = 1,2,\dots,J$. To carry out the likelihood ratio test, we must obtain the values of maximized likelihoods for the full and reduced models, namely $L(F)$ and $L(R)$). $L(R)$ is obtained by fitting the reduced model, and the maximum likelihood estimates of the J parameters in the full model are given by the sample proportions $\pi_j = p_j = \frac{y_{j1}}{m_j} \quad j = 1,2,\dots J$. The likelihood ratio test statistic is given by;

$$D = \sum_{j=1}^{J} dr(y_{j1}, \hat{y}_{j1})^2 \quad (12)$$

Where $dr(y_{j1}, \hat{y}_{j1})$ is the **deviance residual** for the j^{th} covariate pattern which is define as;

$$dr(y_{j1}, \hat{y}_{j1}) = \pm\left\{2\left[y_{j1}\log\left(\frac{y_{j1}}{m_j\hat{\pi}_j}\right) + (m_j - y_{j1})\log\left(\frac{(m_j - y_{j1})}{m_j(1-\hat{\pi}_j)}\right)\right]\right\}^{1/2} \quad (13)$$

The sign of the deviance residual is the same as that of $(y_{j1} - m_j\hat{\pi}_j)$Devi. The statistic in Equation (12) follows a chi-square distribution with degree of freedom $J - \text{number of the parameters in the model}$. Like the Prearson Chi-square test the p-value in Deviance test are not correcting under type one covariate pattern for which $J \approx n$.

3.3 The Hosmer and Lemeshow Test

Many goodness-of-fit tests of univariate probit and logit models are developed to proposed test statistics dealing with the situation in which both discrete and continuous regressors are involved in these models. The widely known test which group the subjects based on the estimated probabilities of success $\hat{C}$. The test statistic, $\hat{C}$, is calculated based on the percentiles of estimated probabilities. The test statistic $\hat{C}$ method is included in several major statistical packages. The tests proposed by [5] do not require the number of covariate patterns less less than the total number of the subjects ($J \approx n$).

In this method, the subjects are grouped into g groups with each group containing $n/10$ subjects. The number of groups g is about 10, the first group contains $\grave{n_1} = n/10$ subjects having the smallest estimated probabilities obtained from the fitted assumed model. The second group contains $\grave{n_2} = n/10$ subjects having the second smallest estimated probabilities, and so on. Let $\bar{\pi}_k$ be the average estimated probability, then the expected number of success observed by the subject in the k^{th} group with j^{th} is

$$\hat{y}_{j1} = \grave{n_k}\bar{\pi}_k \quad (14)$$

where $\sum_{k=1}^{g}\grave{n_k} = n$, and $\grave{n_k}$ $k = 1,2,\ldots g$, is the total subjects in the k^{th} group. let o_k be the number of subjects with $y = 1$ in the k^{th} group. Hence a g by 2 frequency table with the two columns of the table corresponding to the two values

of the response variable, $y = 1,0$ and the g rows corresponding to the g groups. So, the formula of Hosmer and Lemeshow test statistic $\hat{C}$ is

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{15}$$

The test statistic $\hat{C}$ is approximately distributed as a Chi-square distribution with $g - 2$ degrees of freedom (by simulated study). Small values (with large p-value closer to 1) indicate a good fit to the data, therefore, good overall model fit. Large values (with $p < 0.5$) indicate a poor fit to the data [4].

3.4 Pseudo-*R* Square

In linear regression using ordinary least squares, R^2 represents the proportion of variance explained by the model. This measure provides a simple and clear interpretation, takes values between 0 and 1, and becomes larger as the model fits better. Using univariate probit and logit models, an equivalent statistic does not exist, and therefore several pseudo-R^2 statistics have been developed.

McFadden's R-square

McFadden's R-square comparing a model without any predictor to a model including all predictors. It is defined as:

$$R^2 = 1 - \frac{LL_p}{LL_0} \tag{16}$$

LL_0 is the log likelihood of null model (contains intercept only) and LL_p is the log likelihood of given model with p regressors [3]. In many software, there are two modified versions of this basic idea, one developed by Cox and Snell and the other developed by Nagelkerke.

Cox and Snell R-square

Cox and Snell R-square is expressed as:

$$R^2 = 1 - \left[\frac{L_0}{L_p}\right]^{\frac{2}{n}}. \tag{17}$$

L_0 is the likelihood of null model (contains intercept only) and L_p is the likelihood of given model with p regressors. Because this R^2 value cannot reach 1, Nagelkerke modified it. The correction increases the Cox and Snell version to make 1 a possible value for R^2 [10].

Nagelkerke R-square

Nagelkerke R-square allowed the Cox and Snell version to have a value of 1 for

R^2, which is defined as:

$$R^2 = \frac{1-\left[\frac{L_0}{L_p}\right]^{\frac{2}{n}}}{1-[L_0]^{\frac{2}{n}}} \quad (18)$$

Where L_0 is the likelihood of null model (contains intercept only) and L_p is the likelihood of given model with p regressors.

3.5 Discrimination with ROC Curve

Before we start talking about ROC curve we have to mention a classification table and discussed. The classification table is a method to evaluate the predictive accuracy of the univariate probit and logit models. Table (1) is a 2×2 classification table of the predicted values for response variable (at a user defined cut-off value) from the model which can takes two possible values $\hat{y} = 0$ and $\hat{y} = 1)$ versus the observed value of response variable $(y = 0$ or $1)$. For example, if a cut-off value is 0.5, all predicted values above 0.5 can be classified as $\hat{y} = 1$, and all below 0.5 classified as $\hat{y} = 0$. Then a two-by-two table of data can be constructed with binary observed response variable, and binary predicted response variable.

Table 1 Sample Classification Table

Observed	**Predicted**	
	1	**0**
1	a	b
0	c	d

a, b, c and d are number of observations in the corresponding cells. If the univariate probit and logit models have a good fit, then expect to see many counts in the a and d cells, and few in the *b* and *c* cells.

Consider sensitivity = $P(\hat{y} = 1|y = 1) = {}^{a}/_{(a+b)}$ and specificity = $P(\hat{y} = 0|y = 0) = {}^{d}/_{(c+d)}$. Higher sensitivity and specificity indicate a better fit of the model [9]. Using different cut of points and calculation the classification tables, sensitivity and specificity of the model for each cut off point to choose the best cut off point for the purposes of classification. One might select cut off point that maximizes both sensitivity and specificity.

Extending the above two-by-two table idea, rather than selecting a single cut off point, the full range of cut off points from 0 to 1 can be examined. For each possible cut off point, a two-by-two classification table can be formed. Plotting the pairs of sensitivity and one minus specificity on a scatter plot provides an ROC (Receiver Operating Characteristic) curve. The area under this curve (AUC) provides an overall measure of fit of the model. The AUC varies from 0.5 (no predictive ability) to 1.0 (perfect predictive ability) [11].

4 Simulation Study

The main aim of this study is to determine whether there exists a difference between univariate probit and logit models in fitting under certain conditions that are different sample sizes, different coefficients correlations between variables and different cut points for latent response variable. In simulation, latent response variable in equation (3) with $p = 3$ is continuous and affected by three regressors coming from multivariate standard normal distribution with means are zero and the variances are one. Also, we consider three different variance-covariance matrices for multivariate standard normal distribution in data generating process. These matrices were chosen arbitrarily that they were positive definitive and correlations between regressors were zero. Special covariance values were chosen to create different correlation between response and regressors [2]. Covariances between variables means that correlations between them because the variables have been generated from multivariate standard normal distributions. The three variance-covariance matrices are:

$$\sum_{high} = \begin{bmatrix} 1 & 0.4 & 0.5 & -0.7 \\ 0.4 & 1 & 0 & 0 \\ 0.5 & 0 & 1 & 0 \\ -0.7 & 0 & 0 & 1 \end{bmatrix}$$

$$\sum_{low} = \begin{bmatrix} 1 & 0.4 & 0.2 & -0.3 \\ 0.4 & 1 & 0 & 0 \\ 0.2 & 0 & 1 & 0 \\ -0.3 & 0 & 0 & 1 \end{bmatrix}$$

$$\sum_{no} = \begin{bmatrix} 1 & 0.01 & 0.1 & -0.1 \\ 0.01 & 1 & 0 & 0 \\ 0.1 & 0 & 1 & 0 \\ -0.1 & 0 & 0 & 1 \end{bmatrix}$$

To examine the effect of sample size in model selection, five different sample sizes were considered: 1000, 500, 200, 100 and 70. For each of the matrices and sample sizes, the number of simulation was 1,000 times which was found to be sufficient. After data generation, the latent response variable transformed to a binary case for two different cut points: 0 and 0.25. These two cut points are z score in standard normal distribution table corresponds to event probability.

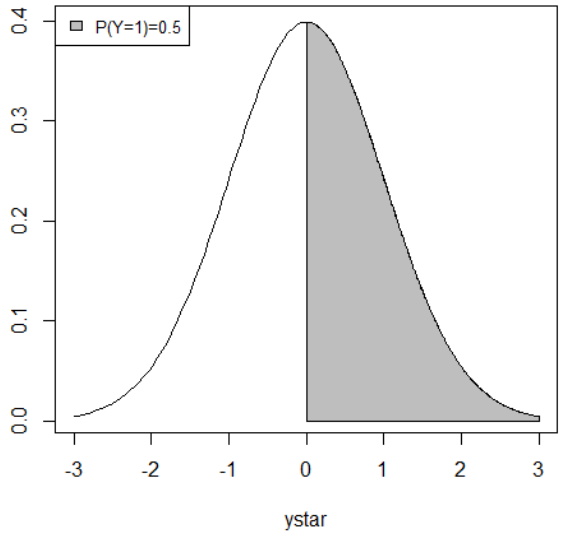


Figure 3 the cut point for $P(Y = 1) = 0.5$ is zero.

So, response variable gets value: $Y_i = \begin{cases} 1 & Y_i^* > 0 \\ 0 & Y_i^* \leq 0 \end{cases}$ for $P(Y = 1) = 0.5$, see figure (3).

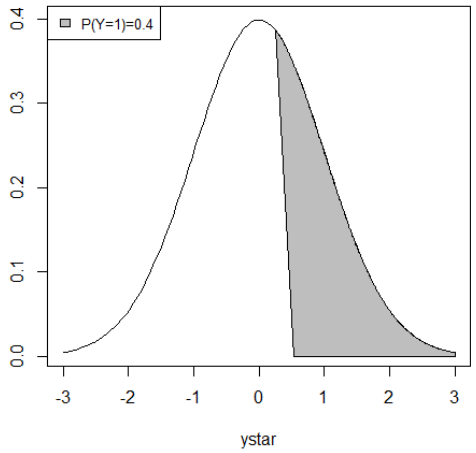


Figure 4 the cut point for $P(Y = 1) = 0.4$ is 0.25.

response variable gets value: $Y_i = \begin{cases} 1 & Y_i^* > 0.25 \\ 0 & Y_i^* \leq 0.25 \end{cases}$ for $P(Y = 1) = 0.4$, see figure (4)

In simulation study, 30 different data generation were performed and generated a total of 30,000 data. In the next step, parameter and probability estimations were obtained using both univariate probit and logit models with MLE approach. Then, goodness of fit statistics and their means on 1,000 replication were calculated.
Table (2) and (4) present Pearson, deviance, and Hosmer and Lemeshow statistics. In the tables *L* denotes univariate logit model, *P* denotes univariate probit model and N denotes sample size. According to Pearson and Hosmer and Lemeshow statistics in Table (3.2) and (3.4), univariate logit model is better than the univariate probit model in "high" case, for 200, 500 and 1,000 sample sizes, which is wrote in bold black. This is because statistic mean values from the univariate logit model are significantly smaller than the values from the univariate probit. In "low" and "no" cases, the mean of statistics in univariate logit model is very close to the mean of statistics in univariate probit model and hence both models fit the data set identically so there is no preference. When response and regressors are uncorrelated, used models are expected to give inaccurate results so goodness of fit measure values for those models should be bad. In low and no cases, this is true. In low and no cases, this is true. Since there is not difference between Table (2) and Table (4) in interpretation thus cut points do not affect model

selection. In the rest of the other measure (AUC and Pseudo-R square) in Table (3) and (5), there are not significantly different between univariate probit and logit models. So, Pearson and Hosmer and Lemeshow statistics were considered more appropriate for comparison univariate probit and logit models.

Table 2 Comparing between logit and probit model by Pearson residuals, Deviance, and Hosmer Lemeshow statistic for cut point = 0

Σ	N	L_pearson residuals	P_pearson residuals	L_ Deviance	P_ Deviance	L_HL	P_HL
High	**70**	36.76149	**33.15507**	27.70749	27.53299	3.988159	**3.572430**
	100	56.87683	**52.62601**	40.97290	40.75278	4.146386	**3.878234**
	200	**127.81244**	130.27128	87.31464	87.03228	**4.886725**	5.364130
	500	**319.50521**	462.00325	223.94021	223.22678	**5.317973**	7.055363
	1000	**635.12650**	772.71111	450.31252	448.93898	**5.725871**	6.579886
Low	**70**	69.25624	68.86719	78.84989	78.81512	8.200096	8.193086
	100	99.15019	99.02480	114.24074	114.19892	8.193125	8.220276
	200	198.77726	199.33853	232.73153	232.67934	8.143497	8.114471
	500	497.36930	499.15366	587.29632	587.20244	8.253728	8.216445
	1000	995.11937	999.12730	1177.65173	1177.48210	8.100131	8.014304
No	**70**	70.00456	69.95863	91.96579	91.96436	8.236264	8.264419
	100	99.94329	99.92927	133.26438	133.25923	8.277938	8.207141
	200	199.98571	199.98107	270.55491	270.55339	8.023801	8.006905
	500	499.98375	499.98683	682.57044	682.56874	8.042246	8.062053
	1000	999.99316	999.99486	1369.37435	1369.37372	8.088584	8.071789

Table 3 Comparing between logit and probit model by area under ROC curve and Pseudo-R Square for cut point = 0

M	N	L_AUC	P_AUC	L_ McFadden	P_ McFadden	L_ Cox&Snell	P_ Cox&Snell	L_ Negelkerke	P_ Negelkerke
High	**70**	0.9718240	0.9717568	0.71172339	0.71353739	0.62031004	0.62127543	0.83088190	0.83217435
	100	0.9717039	0.9716409	0.70231041	0.70390872	0.61789647	0.61874927	0.82659119	0.82773178
	200	0.9693842	0.9693405	0.68396396	0.68498598	0.61024758	0.61081025	0.81499205	0.81574352
	500	0.9686178	0.9686019	0.67647439	0.67750495	0.60758952	0.60815655	0.81063809	0.81139457
	1000	0.9685808	0.9685726	0.67494401	0.67593548	0.60723106	0.60777379	0.80989863	0.81062250
Low	**70**	0.7636957	0.7631849	0.17950998	0.17987006	0.21415445	0.21453497	0.28684644	0.28735495
	100	0.7591984	0.7588759	0.17001171	0.17031403	0.20560711	0.20594048	0.27504887	0.27549383
	200	0.7532736	0.7530998	0.15736911	0.15755793	0.19396049	0.19417128	0.25907888	0.25936035
	500	0.7506179	0.7505414	0.15151275	0.15164840	0.18860486	0.18875534	0.25163826	0.25183904
	1000	0.7499802	0.7499425	0.14990623	0.15002866	0.18724732	0.18738481	0.24974410	0.24992747
No	**70**	0.6223829	0.6220093	0.04308940	0.04310480	0.05643909	0.05645927	0.07558100	0.07560837
	100	0.6054816	0.6052945	0.03147233	0.03150968	0.04183090	0.04187760	0.05595715	0.05601960
	200	0.5880763	0.5879865	0.02069464	0.02070020	0.02797878	0.02798591	0.03736604	0.03737559
	500	0.5734949	0.5734664	0.01379210	0.01379455	0.01884483	0.01884814	0.02514326	0.02514769
	1000	0.5686334	0.5686207	0.01147954	0.01148000	0.01574808	0.01574870	0.02100463	0.02100545

Table 4 Comparing between logit and probit model by Pearson residuals, Deviance, and Hosmer Lemeshow statistic for cut point = 0.25

M	N	L_pearson residuals	P_pearson residuals	L_ Deviance	P_ Deviance	L_HL	P_HL
High	**70**	34.892431	**31.47006**	26.77874	26.58990	3.901033	**3.444007**
	100	55.50360	**50.46109**	39.80654	39.52430	4.306746	**3.959275**
	200	**121.3708**	123.88557	83.77797	83.46914	**4.720483**	4.934612
	500	**311.6097**	376.14851	218.08867	217.38232	**5.253441**	7.457209
	1000	**613.1337**	735.29885	437.63500	436.11728	**5.568973**	6.273796
Low	**70**	68.79674	68.39074	76.47525	76.42832	8.195851	8.159588
	100	99.35165	99.26839	111.04797	111.01867	8.233724	8.304696
	200	198.0020	198.64822	225.87236	225.79687	7.903869	8.043156
	500	496.1151	498.56616	570.30887	570.18804	8.066401	8.047429
	1000	992.6650	998.78935	1142.62201	1142.38617	8.013722	7.961701
No	**70**	70.02500	69.94279	89.17352	89.16991	8.115662	8.229075
	100	99.94429	99.91853	129.42152	129.41650	8.188187	8.326184
	200	200.0043	199.99300	262.64461	262.64473	8.208051	8.175270
	500	499.9935	500.00031	663.09945	663.09877	7.949987	7.907068
	1000	1000.013	1000.01915	1329.94787	1329.94980	8.009890	7.976142

Table 5 Comparing between logit and probit model by area under ROC curve and Pseudo-R Square for cut point = 0.25

M	N	L_AUC	P_AUC	L_ McFadden	P_ McFadden	L_ Cox&Snell	P_ Cox&Snell	L_ Negelkerke	P_ Negelkerke
High	**70**	0.9723738	0.9723394	0.71300575	0.71503142	0.610414	0.61149192	0.828955	0.83042076
	100	0.9722357	0.9721806	0.70271345	0.70480517	0.607681	0.60879300	0.823759	0.82526203
	200	0.9705603	0.9705227	0.68780482	0.68895460	0.601696	0.60232357	0.81461076	0.81546005
	500	0.9689431	0.9689347	0.67573676	0.67678653	0.596664	0.59724048	0.80685674	0.80763517
	1000	0.9690219	0.9690133	0.67471583	0.67584358	0.596389	0.59700395	0.80640791	0.80723820
Low	**70**	0.7658839	0.7654057	0.18157564	0.18208219	0.211121	0.21162113	0.28662117	0.28730136
	100	0.7605421	0.7602562	0.17015221	0.17037129	0.200792	0.20102854	0.27222031	0.27254077
	200	0.7552278	0.7550846	0.15858862	0.15886990	0.190453	0.19075566	0.25776464	0.25817306
	500	0.7520997	0.7520388	0.15204413	0.15222393	0.184336	0.18453370	0.24927656	0.24954363
	1000	0.7519711	0.7519385	0.15090101	0.15107648	0.183498	0.18369098	0.24809604	0.24835653
No	**70**	0.6259819	0.6258730	0.04408239	0.04412356	0.055968	0.05601841	0.07611535	0.07618497
	100	0.6092164	0.6090140	0.03279222	0.03282996	0.042400	0.04244605	0.05747674	0.05753883
	200	0.5903347	0.5902773	0.02134196	0.02134149	0.027991	0.02799182	0.03792384	0.03792379
	500	0.5744011	0.5743721	0.01382832	0.01382932	0.018361	0.01836273	0.02483488	0.02483664
	1000	0.5695765	0.5695627	0.01158434	0.01158291	0.015441	0.01543948	0.02087761	0.02087507

5 Conclusion

In this paper, we have compare between univariate probit and logit models which are members of the family of generalized linear models (GLM). In addition, different goodness-of-fit tests are discussed and their behavior are compared in both models through a Monte Carlo simulation under different conditions. simulation, employing latent variable approach, different sample sizes, different cut points, and different correlations between response variable and regressors were taken into account. To make a comparison between univariate logit and probit models, Pearson residuals, deviations, Hosmer and Lemesshow, area under Receiver Operating Characteristic (ROC) curve, and Pseudo-R square statistics which are used for qualitative data analysis, were calculated and the results were interpreted. In simulations, Pearson residuals and Hosmer and Lemeshow statistics were considered more appropriate for comparison univariate probit and logit model. While according to model's deviance, AUC, and pseudo R square there is no difference between the models in all conditions. The cut points did not affect statistics measure. According to model's Pearson residuals and Hosmer and Lemeshow statistics the models fit differently in high case and also for sample sizes. In high case, logit model's Pearson residuals and Hosmer and Lemeshow statistics were lower for large sample sizes so it was better model. But, when the sample sizes are small, probit model's Pearson residuals and Hosmer and Lemeshow statistics were lower so it was better model. So, the sample size is efficient to choose which model is better.

This is because of variance of probit model is one and variance of logit model is $\frac{\pi^2}{3}$, so logit model has more flat distribution.

Although the logit model has heavier tails because it is greater spread of the distribution curve. In another word, univariate logit model is better than univariate probit model in larger sample size because when the sample size increases, probability of observes in tail increases too. This is the reason for univariate logit model is better than univariate probit model in large sample sizes.

References

[1] P.D. Alisson, *Logistic Regression Using SAS: Theory and Application*, SAS Institute Inc., 2012.

[2] S. Cakmakyapan, A. Goktas, A comparison of binary logit and probit models with a simulation study, *Journal of Social and Economic Statistics*, **2** (2013), 1-17.

[3] J. M. Hilbe, *Logistic Regression Models*, Chapman & Hall/CRC Texts in Statistical Science, 2009.

[4] D. W. Hosmer, T. Hosmer, S.Le Cessie, S. Lemeshow, A comparison of goodness-of-fit tests for the logistic regression model, *Statistics in Medicine*, **16** (1997), 965-980. https://doi.org/10.1002/(sici)1097-0258(19970515)16:9<965::aid-sim509>3.0.co;2-o

[5] D. W. Hosmer and S. Lemeshow, *Special Topics*, Wiley Online Library, 2000.

[6] M. H. Kutner, C. Nachtsheim and J. Neter, *Applied Linear Regression Models*, McGraw-Hill/Irwin, 2004.

[7] J. S. Long and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*, Stata Press, 2006.

[8] P. McCullagh and J. A. Nelder, *Generalized Linear Models* (Second ed.), London, Chapman and Hall, 1989.

[9] H. Park, An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain, *Journal of Korean Academy of Nursing*, **43** (2013), 154-164. https://doi.org/10.4040/jkan.2013.43.2.154

[10] S. Sarkar, H. Midi and S. Rana, Model selection in logistic regression and performance of its predictive ability, *Australian Journal of Basic and Applied Sciences*, **4** (2010), 5813-5822.

[11] M. H. Soureshjani and A. M. Kimiagari, Calculating the best cut off point using logistic regression and neural network on credit scoring problem-A case study of a commercial bank, *African Journal of Business Management*, **7** (2013), 1414.

Received: January 23, 2018; Published: February 12, 2018