Multiple Binomial Regression Models of Learning Style Preferences of Students of Sidhu School, Wilkes University

Adekanmbi, D. B[1]

[1] Department of Statistics, Ladoke Akintola University of Technology, Nigeria

Correspondence: Adekanmbi, D.B, Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Nigeria.

Abstract

The interest of this study is to explore the relationship between a dichotomous response, *learning style preferences* by university students of Sidhu School, Wilkes University, as a function of the following predictors: *Gender, Age, employment status, cumulative grade point assessment (GPA)* and *level of study,* as in usual generalized linear model. The response variable is the students' preference for either Behaviorist or Humanist learning style. Four different binomial regression models were fitted to the data. Model A is a logit regression model that fits all the predictors, Model B is a probit model that fits all the predictors, Model C is a logit model with an effect modifier, while Model D is a probit model also with an effect modifier. Models A and B appeared to have performed poorly in fitting the data. Models C and D fit the data well as confirmed by the non-significant chi-square lack of fit with p-values 0.1409 and 0.1408 respectively. Among the four models considered for fitting the data, Model D, the probit model with effect modifier fit best. There was a marginal difference in the measure of goodness-of-fit for models C and D. Since probit model usually do not lend itself to ease of interpretation, model C was focused on for interpretation of results. The four variables that made significant contributions to model C were gender, age, employment status and the interaction variable. Academic performance of the students measured by their GPA and the level of study of the students were not significant predictors of the learning style preference by the students. The results of Model C revealed that the likelihood that a student prefers Behaviorist learning style is negatively related to his or her gender, age, employment status, GPA and level of study. However, the likelihood is positively related to the interaction term: Gender* Age. The result also showed that every one year increase in age of the students leads to decrease in the log-odds of preference for Behaviorist learning style. Also the odds of an MBA student preference for Behaviorist learning style are 1.1925 times greater than the odds of an undergraduate student. The association between gender and age was significant, so that gender modifies the association between age and preference. The interaction term showed that both the male and female odds ratio indicate an increase of odds of Behaviorist learning style, with increasing age of students, but the rate of increase is greater for male students. Plots of residuals and other diagnostic procedures conducted further confirmed that models A and B did not yield good fit, while both models C and D though identified an outlier which was not influential, but the functional forms of the models appeared suitable and seemed to fit the data well, and were therefore considered adequate. The residual mean deviance of model C was slightly above 1 which an indication of a slight overdispersion problem in the model. Important issues arising from the study were also discussed.

Keywords: dichotomous response variable, Akaike Information Criterion (AIC), Logit transformation, Probit transformation, Effect modifier, QQ plot, Cook's statistics

1. Introduction

Learning style is the accustomed pattern used to acquire information, concepts and skills (Swisher, K., 1990). Learning styles is a term used to describe how learners gather, interpret, organize, draw conclusions about, and store information for further use (Coffield, F., Moseley, D. & et al., 2004). Cultural influences have a great impact on the ways learners consume information. Due to differences in cultural effects, background and individual nature, each learner has a unique preference to how he or she prefers to learn in a classroom environment. Understanding the preferred learning style is important for any university teacher who wishes to ensure that a maximum learning efficacy takes place when teaching. Since students have differences in the way they learn, it is important to understand the learning preferences of individual student.

There are over seventy different learning styles schemes which have been popularized by professional development for teachers and educators (Coffield, F., Moseley, D. & et al., 2004; Pashler, H., McDaniel, M. & et al., 2008). The fundamental idea behind learning styles is that each learner has a specific learning style referred to as preference, and individuals learn best when information is presented in the preferred style. Two of the groups of learning theories, namely: Behaviorism and Humanism will be focused on in this study. Behaviorist belief is based on the fact that learning is a change in observable behavior and happens when communication occurs between two events, which are a stimulus and a response (Hauser, L., 2005 ; Saif, A. A., 2009). Learners with behaviorist reinforcement should be conditional for good performance (Basavanthappa, B. T., 2003). Behaviorists therefore insist on practice and repetition in learning, believing that the final behavior would make it to repeat again. Behaviorist training method disadvantage is that the method depends on the students' motivation, so that enough attention is not paid to strengthen their intrinsic motivation. Learning in Behaviorist method is in form of a response to stimulus and less attention given to the process of thinking and critical thinking in students (Quinn, F. M., 2007).

Humanist learning method respects the feelings and experiences of learners. The main aim of education is to reach the final limit of the individual growth potential (Quinn, F. M., 2007). Humanists know human have freedom of choice and not under the influence of unconscious forces, so that people are free to choose their targets. Humanist learning method tries to get around individual limitations in order to develop their mental capabilities (Sobhaninejad, M., 2005). Humanism view human as a free person who tries to maximize its capabilities. In this learning method, the teacher gives the students priority to learn by their curiosity, so that the students have freedom to learn as they wish to learn. Nevertheless there are problems to this approach of learning. Humanist learning approach requires a teacher who has self-discovery and sincerity to be honest with the students in sharing expectations and feelings with the students, and also share the students' feelings and their desires. On the part of the students, humanist learning approach may cause annoyance and dissatisfaction for them, since it could be difficult for them to compare the learning method to the traditional methods of teaching (McKenna, G., 1995). Adult students in university participate in setting the training, and not to conform to it.

Logistic regression and probit regression are special cases of generalized linear models, (GLMs), which extend ordinary linear regression (OLS) to situations where the response variable is binomial (Efron, B. & Hastie, T., 2016). The use of binomial regression is increasing from social sciences, epidemiological to educational measurement (Austin, J. T., Yaffee, R. A. & eta al., 1992; Cabrera, A. F., 1994; Chuang, H. L., 1997; Peng, C.Y., Manz, B. D. & et al., 2001; Tolman, R. M. & Weisz, A., 1995). Logistic regression is a specialized technique analysis of proportion data, commonly used tool in applied statistics to analyze data with dichotomous or polytomous response variable (Agresti, A., 1996; Efron, B. & Hastie, T., 2016; Krzanowski, W. J., 1998; McCullagh, P. & Nelder, J. A., 1998). Binomial regression model is suitable for describing and testing hypotheses about relationships between a categorical response variable and one or more categorical or continuous predictors. The response variable to predict could be binary or dichotomous categories, such as "alive or dead" "survive or does not survive" "present or absent", "win or loose" and so on(Krzanowski, W. J., 1998; Lawal, B., 2003). In essence, binomial regression is a method of fitting curve when the response variable consists of proportions or probabilities, or binary coded.

In classical linear model, transformation of the response could be considered as need arises. This is usually impractical for a generalized linear model (GLM), since it would change the assumed distribution of the response variable(Faraway, J. J., 2006). The link function is a fundamental assumption of any GLM. The link function in GLM is used to link the linear predictor to the mean of the response. The three common choices of link function for a binary regression model are: Logit, Probit and Complementary log-log(Faraway, J. J., 2006; Krzanowski, W. J., 1998).

In this study, the research hypothesis is to investigate the likelihood that a student in Sidhu School of Wilkes University Business School will prefer a learning style preference is related to his or her gender, age, level of study (undergraduate or postgraduate), academic performance (GPA) and employment status. The response variable is *Pref* (learning style preference) which is coded as 1=Behaviorist (if a student prefers Behaviorist learning style), 0=Humanist (if a student prefers Humanist learning style). The *Gender* is coded as 0= Male, 1= Female. *Level* of study is coded as 0= Undergraduate, 1= MBA, while the academic performance of the students, *GPA* which is a measure of students performance is coded as 1= 2.6-3.0; 2=3.1-3.5, 3=3.6-4.0. The employment status of the students *(Empstatus)* is coded as 1=unemployed, 2= employed for less than 10 hours per week, 3=employed for more than 35 hours per week. *Age* of the students is the only continuous variable in the data.

This study is aimed at investigating the factors responsible for the choice of learning style preference by Wilkes University students and to describe differences between students that prefer the Behaviorist versus Humanist learning style. The remainder of this study is divided into five sections. Section two focuses on the description of the learning style preferences data, employed in this study. A full discussion of the methodology of binomial regression model is given in section three, while sections 3.1, 3.2.1, 3.2.2, and 3.4 focus on effect modifier, goodness of fit, diagnostics of binomial

regression model and model selection criteria respectively. Results of the analysis and model selection are discussed in sections 4.0 and 4.1 respectively. Discussion on some critical issues in the study is given in section 5.

2. Data

The data employed in this study is a sample of 91 students in Sidhu School of Wilkes University Business School in United States of America. Undergraduate and postgraduate academic programmes are offered in the school. The two learning styles used in the school are Behaviorist learning style and Humanist learning style. 39 of the students preferred Behaviorist learning style, while the remaining 52 preferred Humanist learning style. 45 of the students in the sample are undergraduate students, while 46 of them are postgraduate students. The minimum age of the students in the sample is 18, while the maximum age is 57. 51 of the students are males and 40 of them are females.

3. Methodology of Binomial Regression Model

Given a binary response variable Y, to model the conditional probability $P(Y = 1 | X = x)$ as a function of x, then the standard way of modeling such data is to use binomial regression model, which is a special case of GLM. For binary data, appropriate probability model is the Bernoulli distribution. Generally binomial regression is suitable for describing the relationships between a dichotomous response variable and one or more categorical or continuous predictor variables (Demaris, A.,1992; Peng, C. Y., Manz, B. D. & et al., 2001). The mean is restricted to lie between 0 and 1, and the variance of response variable is a function of the mean (Hosmer, D. W., Jr. & Lemeshow, S., 2000; Krzanowski, W. J., 1998). The two transformations that will be considered in this study are logit and probit. For logistic model, the underlying mathematical concept of is logit which is the logarithm of an odds ratio. Logistic regression model has three components which are:

(1) Random component: The response variable is binary, so that $Y_i = 1 \text{ or } 0$. The interest is in probability that, $Y_i = 1$, with $\pi(x_i)$. The distribution is therefore Binomial.

(2) Systematic component: The linear predictor is

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_j x_{ji} \tag{1}$$

The value of the coefficient β determines the direction of the relationship between X and the logit of Y. The predictor variables may be continous or discrete or both.

(3) Link Function: Logit is the natural logarithm transformation of odds of Y, while the odds are the ratios that of probabilities (π) of Y happening. The logit is the linear part of logistic model (Hosmer, D. W., Jr. & Lemeshow, S., 2000). The logit parameter η is given as:

$$\eta = \text{logit}\,(\pi(x)) = \ln\left(\frac{\pi}{1-\pi}\right) \tag{2}$$

Consider k independent random variables $Y_1, Y_2, \ldots, Y_k$, and each following binomial distribution, and corresponding to the number of successes in k different subgroups. Given that the probability of success $P(Y = \pi_i)$ is constant within each group, and n_i is the corresponding number of trials in each subgroup, then counts y_i are modeled as independent binomials, so that:

$$y_i \sim \text{Bin}(n_i, \pi_i) \qquad \text{for } y_i = 0,1,2,\ldots n_i \text{ and } i = 1,2,\ldots k \tag{3}$$

The mean function $\hat{\pi}_i$ thus transforms a real value into a value between 0 and 1, and a possible link function is the logit transforms (Marin, J. M. & Robert, C. P., 2014). Hence, logistic regression is linear regression on the logit transform of the response variable.

In order to predict probabilities that individual fall into one of two categories of a dichotomous response variable as a function of some set of predictor variables, then the logit of $\pi(x)$ is therefore:

$$\eta = \text{logit}\,(\pi(x)) = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_j x_{ji} \tag{4}$$

where

π is the probability of the event.

β_0 is the Y intercept.

βs are regression coefficients.

Xs are a set of predictors.

Then, the specific form of the logistic model is:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_j x_{ji})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_j x_{ji})} \tag{5}$$

The logarithm of odds changes linearly with x; but is not intuitively easy to interpret. The values of the coefficient of parameters βs determine the direction of the relationship between the Xs and the logit of Y. Another link function for binomial regression is probit. In order to ensure that π is between 0 and 1, a positive monotone function that maps the linear predictor, into the unit interval is:

$$\eta = \beta_0 + \beta X_i \tag{6}$$

$$\pi_i = \Phi(\eta_i) = \Phi(\beta_0 + \beta X_i) \tag{7}$$

where

$\Phi(.)$: standard normal cumulative distribution function.

βs : are parameters to be estimated.

The probit model can also be generalized to *k* explanatory variables, such that

$$\begin{aligned}\pi_i = \Phi(\eta_i) &= \Phi(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}) \\ &= \Phi(X_i'\beta)\end{aligned} \tag{8}$$

In probit regression, the errors are assumed to have a standard normal distribution. The choice of link function is made based on assumptions derived from physical knowledge or simple convenience (Austin, J. T., Yaffee, R. A. & et al., 1992; Faraway, J. J., 2006). The choice of link function could be set by the characteristics of the response or ease of interpretation. In order to ensure that π lie in the interval (0,1), the probability scale must be transformed from the range (0,1) to $(-\infty, +\infty)$. Then we can formulate a linear model for the transformed variable which will ensure that the fitted probabilities lie between 0 and 1, when we back-transform. For logistic model for binary data, the estimated coefficients of predictors can be used to estimate the odds ratio, but does not work for probit model for binary data. No simple interpretation exists for other link function such as probit(Demaris, A., 1992). The method of maximum likelihood (ML) estimation is usually used to estimate the vector of parameters $\beta = (\alpha, \beta_1, \beta_2 \ldots \beta_j)$ of the model.

3.1. Effect Modifier

Effect modification occurs when a factor regarded as effect modifier, modifies the causal relationship between risk factor and the response variable (Hosmer, D. W., Jr. & Lemeshow, S., 2000; Woodward, M., 2004). It occurs when the magnitude of the effect of the primary exposure on a response variable differs depending on the level of a third variable. It could be tested whether a variable is an effect modifier by including an interaction term between the variable and the risk factor of interest into the model. If the interaction term is both meaningful and statistically significant, then the variable is said to be an effect modifier (Collet, D., 2003;Hosmer, D. W., Jr. & Lemeshow, S., 2000). In this study, the relationship between the response variable, learning style preference (*Pref*) and *Age* is different among males and females. It is therefore suspected that there could be an interaction between gender and age. It is necessary to determine if gender modifies the effect of *Age* on *Pref*. In this case, one cannot come up with a single estimate for the effect of *Age*, but need to fit separate models for the two categories of gender of the students; male and female. The odds ratio for age cannot be estimated without specifying the gender at which the comparison is being made. *Age* is quantitative while *Gender* is categorical, so that the interaction term could be represented by the set of variables defined by the product of the quantitative variable and each of the dummy variables for the categorical variable.

3.2 Inference for Binomial Regression Model

Goodness-of-fit statistics could be used to assess the fit of an estimated logistic model, to determine how effectively the proposed model describes the response variable, and thereby assess the goodness of fit of the model (Hosmer, D. W., Jr. & Lemeshow, S., 2000). Some of the measures of goodness-of-fit are:

(1) Wald statistic
   For large samples

$$z = \frac{\hat{\beta}}{ASE} \tag{9}$$

Z is approximately N(0,1) when $H_0 : \beta = 0$ is true.

To determine the statistical significance of individual regression coefficient, the Wald Chi-square statistic can be used. For 1-tailed test, we could refer to standard normal distribution that is the Wald's statistic (w_j)

$$w_j = \left(\frac{\hat{\beta}}{ASE}\right)^2 \tag{10}$$

The Wald statistic is approximately chi-square distributed with df = 1, if the null hypothesis is true. For large samples, the distribution of parameter estimates is approximately normal. A large sample $(1-\alpha)100\%$ confidence interval for β is:

$$\hat{\beta} \pm z_{\alpha/2}(ASE) \tag{11}$$

where

α is the significance level.

ASE is the asymptotic standard error.

(2) Likelihood ratio test statistic
   Is a more powerful alternative to Wald test. The test statistic is

$$\text{test statistc} = -2(L_0 - L_1) \tag{12}$$

where

L_0 : is the log of the maximum likelihood for the model, $\text{logit}(\pi(x)) = \alpha$

L_1 : is the log of the maximum likelihood for the model, $\text{logit}(\pi(x)) = \alpha + \beta x$

The likelihood ratio test statistic is approximately chi-square distributed with df = 1, if the null is true.

3.3 Model Checking for Multiple Binomial Regression Model

Having completed the model building stage, it is necessary to know how effectively a multiple binomial regression has described the response variable, to assess the fit of the model. This involves diagnostics of the model.

3.3.1 Diagnostic of Logistic Regression Model

It is necessary to confirm if the proposed model approximates the true relationship between $\pi(x)$ and x; and that the model is a good one. In logistic regression, influential observations and goodness of fit can be assessed by residual statistics (Lawal, B., 2003). Residuals are essential to explore the adequacy of generalized linear model (Faraway, J. J., 2006). In fact, a logistic model should not be accepted without conducting diagnostic tests for lack of fit and influential observations (Collet, D., 2003). There are two types of diagnostic methods. Some diagnostic methods are designed to detect outliers, while other methods are designed to check the assumptions of the model (Faraway, J. J., 2006; Hosmer, D. W., Jr. & Lemeshow, S., 2000). Many of the diagnostics and plots are based on residuals.

1. Pearson residual is defined as:

$$e_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}} \tag{13}$$

where

y_i : observed number of events.

$\hat{\pi}_i$: estimated probability of x_i .

n_i : number of observations with predictor variable equal to x_i .

$n_i\pi_i$: estimated number of events.

The measure provides a single number that summarizes the agreement between observed and fitted values (Hosmer, D. W., Jr. & Lemeshow, S., 2000). When the model holds, the Pearson residuals are approximately normal with mean 0 and variance less than 1.

2. Deviance residual is a measure of goodness of fit, depends on $D(y_i, \hat{\pi}_i)$, is defined as:

$$D(y_i, \hat{\pi}_i) = \pm \left\{ 2\left[y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\ln\left(\frac{(n_i - y_i)}{n_i(1-\hat{\pi}_i)}\right) \right] \right\}^{1/2} \tag{14}$$

where

y_i : the observed value for the ith observation.

$\hat{\pi}_i$: estimated logistic probability for the ith case.

D has an asymptotic chi-squared distribution on $n - p$ degree of freedom under the null hypothesis that the data could have been generated by logistic model (Efron, B. & Hastie, T., 2016). The deviance can be used to test whether the model is an adequate fit (Faraway, J. J., 2006). The larger the deviance, the poorer the fit of a GLM model. It is possible to determine whether a deviance is large or small by computing the p-value. Also, the null hypothesis can be rejected, if the deviance is far in excess of the degrees of freedom. Residual deviance is the deviance for the current model, while the null deviance is the deviance for a model with just an intercept term with no predictors (Faraway, J. J., 2006).

Either of these two residuals could be plotted against the observation number, to check for the linear predictor. A systematic pattern in this plot indicates that the model in not good. Also, either of the residuals could be plotted against the linear predictor, which should show hyperbolic lines of residuals clustered around 0 and 1. Influential observations can be identified by standardized residuals greater than +2 or smaller than -2. Systematic patterns can be checked by plotting the residuals (Woodward, M., 2004). Graphical methods of residual analysis include plot of the jackknife deviance residuals against the fitted values, QQ plot of the residuals to check the normality assumption on the errors made for a linear model, plot of Cook statistics against the standardized leverages and the Cook statistic plotted against case number which can be used to indentify observations that are influential.

3.4 Model Selection Criterion

A popular criterion for choosing among competing statistical models is the Akaike Information Criterion (AIC) defined as:

$$AIC = -2L + 2p \tag{15}$$

where

p: the number of parameters of a model.

L: maximum log-likelihood

The measure considers the distance between a "true" model and a model fit to the data. Akaike suggested that the preferred model would be the one with a minimum AIC value (Lawal, B., 2003). A caution that is warranted when using AIC to compare models is that the same data should be fit by models that are being compared (Agresti, A., 1996).

3.5 Overdispersion in Binomial Regression

Overdispersion is one of the problems that often arise with modeling binary data, where data involving proportions tend to be more variable than the underlying binomial distributions can accommodate (Krzanowski, W. J. (1998; Lawal, B., 2003). Binomial distribution could seem suitable for a response variable y from theoretical point of view, but the relationship between mean and variance might not be consistent with the distributional assumptions. If the variance is larger than it should be, there is an evidence that the data exhibit overdispersion.

Overdispersion can be caused by several factors. It could be due to systematic misspecification of a model where some interaction term or other important variables have been omitted from the model. The problem could also arise when the observations of the response variable are correlated, so that the independent assumption is violated.

The dispersion parameter can be estimated using

$$\hat{\varphi} = \frac{\chi^2}{n-p} = \frac{\sum_i (y_i - \hat{\mu}_i)^2 / \hat{\mu}_i}{n-p}$$

(16)

Also, the residual mean deviance (RMD) which is the ratio of the residual deviance to the corresponding degrees of freedom is an estimate of $\hat{\varphi}$ (Faraway, J. J., 2006).

$$\hat{\varphi} = \frac{\text{Deviance}}{n-p}$$

(17)

$\hat{\varphi}$ could take one of the three possible values leading to overdispersion, underdispersion and no dispersion respectively (Lawal, B., 2003).

$$\hat{\varphi} = \begin{cases} >1 & \text{overdispersion} \\ <1 & \text{underdispersion} \\ 1 & \text{no dispersion} \end{cases}$$

4. Results of the Multiple Binomial Regression Analysis

The binomial regression models were fitted to the data to determine the functional relationship between the likelihood that a student will prefer either of the two learning style preferences: Behaviorist and humanist, based on and his or her Gender, Age, Employment status, academic performance (GPA), and level of study.
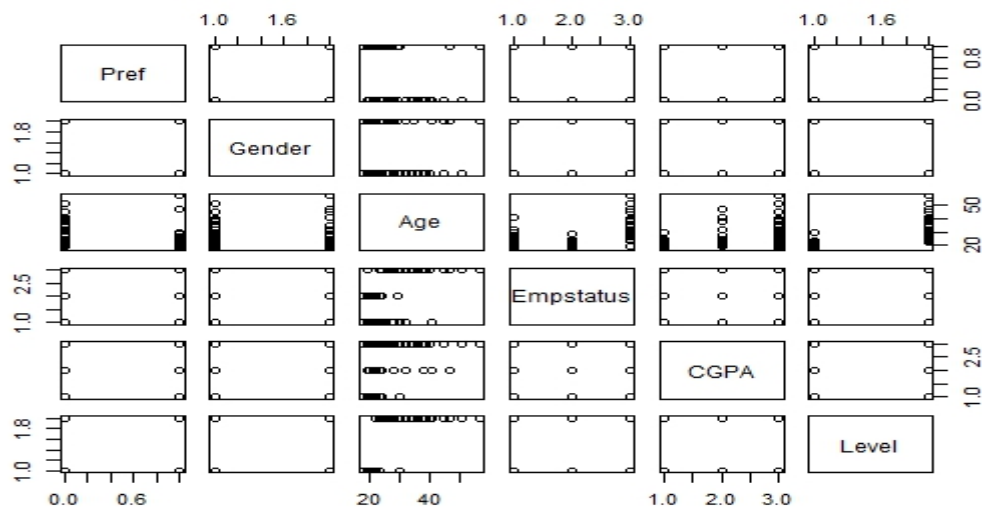

Figure 1: Matrix scatter plot of the response and predictor variables for the Sidhu School, Wilkes University students.

The matrix scatter plot in figure 1 is used to provide an impression of the nature of relationship between the response variable, learning style preference (*Pref*), and the predictors, which are Gender of the students (*Gender*), Age of the students, (*Age*), Employment status of the students (*Empstatus*), Academic performance of the students (GPA), and level of study (undergraduate and MBA). The response variable is whether a student prefers Behaviorist or Humanist learning style. Figure 1, depicts the dichotomous nature of the response variable *Pref* so that all points fall on either Behaviorist (y=1) and Humanist (y=0), and also valuable in gaining an understanding of the data, but does not show a clear picture of the nature of the relationship between *Pref* and the predictor variables.

The interpretation of the fitted binomial regression models involves drawing practical inferences from the estimated coefficients of the model. Table 1 presents the results of fitting logit and probit regression models separately to the data without interaction, and also the two models with effect modifiers or interaction. The estimated coefficients for each predictor in the models, with their corresponding standard errors, z-values and p-values are provided in table 1, for the four models considered. In addition to the deviances of the models, the p-values of the deviances are also given in the table. It should be noted that the regression coefficients for the reference levels of the categorical variables in the models are suppressed.

Model A and B are the logit and probit models respectively that fit all the predictors together. According to models A and B, the log of odds that a student selects a particular learning style preference is negatively related to his or her gender, age, employment status and level of study. The residual deviance of model A is 107.07 on 83 d.f yields p-value of 0.0389 while the residual deviance of model B is 107.92 on 83 d.f yields a p-value of 0.0344, which are indications that models A and B are not adequate. The residual deviances of models A and B are also far in excess of their respective degrees of freedom, a further confirmation that the models do not fit.

Models C is the logit with effect modifier, while model D is the probit with effect modifier. Each of models C and D separately adds the interaction between age and gender to models A and B that contain all the predictors. The models C and D are therefore improvements over models A and B. The fitted models C and D based on logit regression and probit regression respectively are:

$$\eta = \text{logit}\left(\pi(x)\right) = 7.7908 - 0.2653\text{Age} - 7.3403\text{Gend er}^{(1)} - 1.5531\text{Emps tatus}^{(1)}. - 0.4127\text{Emps tatus}^{(2)}.$$
$$- 0.8051\text{GPA}^{(1)} - 0.5366\text{GPA}^{(2)} - 0.1770\text{Level}^{(1)} + 0.2713\text{Age} * \text{Gender}^{(1)}$$

$$\pi_i = \Phi(\eta_i) = \Phi\left(\begin{array}{l} 4.7051 - 0.1605\text{Age} - 4.4206\text{Gend er}^{(1)} - 0.9164\text{Emps tatus}^{(1)}. - 0.2398\text{Emps tatus}^{(2)}. \\ - 0.4903\text{GPA}^{(1)} - 0.3119\text{GPA}^{(2)} - 0.1163\text{Level}^{(1)} + 0.1635\text{Age} * \text{Gender}^{(1)} \end{array}\right)$$

The log of odds of model D, which is the probit model with interaction, is not intuitively easy to interpret. Since model C, the logit model with interaction is equally good and lends itself to ease of interpretation, we focus on model C. According to model C, the log of odds of a student preferring a learning style is negatively related to all the predictors but is positively related to the interaction term: *Gender*Age* (p<0.05). In fact the interaction variable is quite significant (p<0.0001). This implies that *Gender* modifies the effect that age has on the choice of learning style by the students. The two models with effect modifier also show that gender, age, employment status and the interaction variable are significant predictors of learning style preferences by the students (p <0, 0.001, 0.1 and 0.05 respectively). Academic performance of the students measured by the GPA and the level of study of the students are not significant predictors. The intercept for the models are significant, suggesting that intercept should be included in the models.

The coefficient for age $\beta_1 = -0.2653$, implies that every one year increase in age of the students, leads to decrease in the log-odds of preference for Behaviorist learning style by 0.2653, when other variables are controlled. The coefficient for age could also be interpreted that for every one year increase in age, the odds of preferring Behaviorist learning style decrease by $e^{-0.2653} = 0.7669$, with other predictors held fixed. The dummy variables in model C include Gender, Employment status (*Empstatus*), GPA, and Level. The estimates for the dummy variables show that there is a decreased preference for Behaviorist learning style for students employed for more than 10 hours and those employed for less than 35 hours relative to students that are unemployed, also for students with GPA 3.1-3.5 and 3.6-4.0 relative to students with GPA 2.6-3.0; and for MBA students relative to the undergraduates. The odds of preference for Behaviorist learning style among undergraduate students are exp(-0.1769) = 0.8379 times higher as compared to MBA students, when other variables are controlled. All things being equal, the odds of an MBA student preference for Behaviorist learning style is therefore $1/\exp(-0.1769) = 1.1935$ times higher than the odds of an undergraduate student. The odds of preference for Behaviorist learning for students employed for less than 10 hours are about exp(-1.5531) = 0.2116 times higher than the odds of the unemployed students, after adjusting for the effects of other variables. With regard to Behaviorist style, there is statistically significant difference between male and female students, (p=0.0024), after adjusting for the effect of other variables. Gender differences in preference for Behaviorist style increases with age. For the female students, a one year increase in age yields a change in odds ratio of (-0.7340-0.2653+0.2713)=-0.728, while for male students, the odds ratio is -0.2653, controlling for level of other variables. In terms of odds, for female students, the odds is exp(-0.728)=0.48 for a one year increase in age, controlling for other variables; and the odds is exp(-0.2653)=0.77 for a one year increase in age of male students, after adjusting for the effect of other variables. So that for the interaction variable, both the male and female odds ratio indicate an increase of odds of Behaviorist learning style, with increasing age of students, but the rate of increase is greater for male students.

Table 1. Binomial regression output, showing parameter estimates, associated standard errors, and inferences for the parameters in the model

Model	Parameter	Variable name	Estimate	Std. Error	Z –value	$p > \|z\|$	Null deviance (df)	Residual deviance	p-chisq(deviance)
Model A:Logit	β_0		3.3534	1.2841	2.611	0.0090*	124.29 (90)	107.07(83)	0.0389
	β_1	Age	-0.0788	0.0513	-1.534	0.1251			
	$\beta_2^{(1)}$	Gender$^{(1)}$	-0.6256	0.4860	-1.287	0.1980			
	β_3	Empstatus$^{(1)}$	-1.5599	0.6448	-2.419	0..0156*			
	$\beta_3^{(2)}$	Empstatus$^{(2)}$	-0.4666	0.7098	-0.657	0.5110			
	$\beta_4^{(1)}$	GPA$^{(1)}$	-0.6067	0.7012	-0.865	0.3870			
	$\beta_4^{(2)}$	GPA$^{(2)}$	-0.6687	0.7293	-0.917	0.3592			
	$\beta_5^{(1)}$	Level$^{(1)}$	-0.3905	0.8238	-0.474	0.6355			
Model B:Probit	β_0		1.7359	0.6978	2.488	0.0129*	124.29 (90)	107.92(83)	0.0344
	β_1	Age	-0.0341	0.0274	-1.244	0.2136			
	$\beta_2^{(1)}$	Gender$^{(1)}$	-0.3241	0.2888	-1.122	0.2618			
	$\beta_3^{(1)}$	Empstatus$^{(1)}$	-0.9540	0.3831	-2.490	0.0128*			
	$\beta_3^{(2)}$	Empstatus$^{(2)}$	-0.3205	0.4274	-0.750	0.4533			
	$\beta_4^{(1)}$	GPA$^{(1)}$	-0.3639	0.4202	-0.866	0.3865			
	$\beta_4^{(2)}$	GPA$^{(2)}$	-0.4067	0.4377	-0.929	0.3528			
	$\beta_5^{(1)}$	Level$^{(1)}$	-0.3233	0.4874	-0.663	0.5071			
Model C:Logit effect modifier	β_0		7.7908	2.3716	3.385	0.0010**	124.289 (90)	95.836(82)	0.1409

β_1	Age	-0.2653	0.0988	-2.685	0.0073**		
$\beta_2^{(1)}$	Gender[1]	-0.7340	2.4180	-3.036	0.0024**		
$\beta_3^{(1)}$	Empstatus[1]	-1.5531	0.6856	-2.265	0.0235*		
$\beta_3^{(2)}$	Empstatus[2]	-0.4127	0.7293	-0.566	0.5715		
$\beta_4^{(1)}$	GPA[1]	-0.8051	0.7651	-1.052	0.2927		
$\beta_4^{(2)}$	GPA[2]	-0.5366	0.7850	-0.684	0.4943		
$\beta_5^{(1)}$	Level[1]	-0.1770	0.8616	-0.205	0.8372		
β_6	Gender*Age	0.2713	0.0978	2.773	0.0056**		
						95.587(82)	0.1409
Model D:Probit effect modifier β_0		4.7051	1.3249	3.551	0.0004***	124.289 (90)	
β_1	Age	-0.1605	0.0558	-2.873	0.0041**		
$\beta_2^{(2)}$	Gender[1]	-4.4206	1.3746	-3.216	0.0013**		
$\beta_3^{(1)}$	Empstatus[1]	-0.9164	0.3995	-2.294	0.0218*		
$\beta_3^{(2)}$	Empstatus[2]	-0.2398	0.4394	-0.546	0.5852		
$\beta_4^{(1)}$	GPA[1]	-0.4904	0.4483	-1.094	0.2740		
$\beta_4^{(2)}$	GPA[2]	-0.3119	0.4639	-0.672	0.5014		
$\beta_5^{(1)}$	Level[1]	-0.1163	0.5125	-0.227	0.8205		
β_6	Gender*Age	0.1635	0.0553	2.956	0.0031**		

The test of model adequacy for models A and B though not reported, further confirms that the models do not fit. In order to test for the adequacy of models C and D, there is need to examine the plots of residuals and conduct other diagnostic procedures. Figures 2(a) and 2(b) show the plots of jackknife deviance residuals against linear predictor in the upper left panel of the figures, normal scores plots of standardized deviance residuals referred to as normal QQ plot in the upper right panel in the figures, plot of approximate Cook statistics against leverage/(1-leverage) in the lower left panel of the figures, and case plot of Cook statistic against case number in the lower right panel of the figures, each for models C and D. The dichotomous nature of logistic residuals in the plot of residuals against linear predictors makes it almost

impossible to discern any pattern in the plot. The normal QQ plot shows that the standardized residuals are normally distributed. An observation above the horizontal line on the plots of the cook statistics may be an outlier. The functional forms of the models appear to be suitable, and on the average, the diagnostic plots show no much reason for concern. It could therefore be concluded that models C and D each fit the data well.

Influential diagnostics was conducted to check for influential observation in the data. The *DFBeta* for the parameter estimates indicate that none of the observation when deleted from the model have any significant effect on the original parameter estimates. Also none of the standardized deviance residual is outside the interval [-2, 2].

For model C, the estimate of dispersion parameter $\hat{\phi}$ yields 1.16 which is slightly greater than 1. The value though not unusually large but suggested slight overdispersion problem in the data.
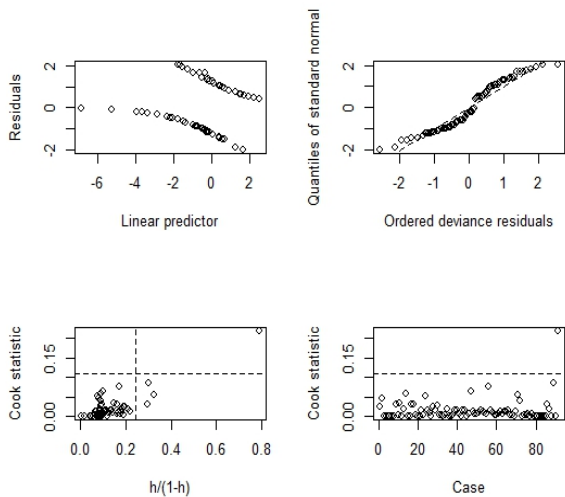

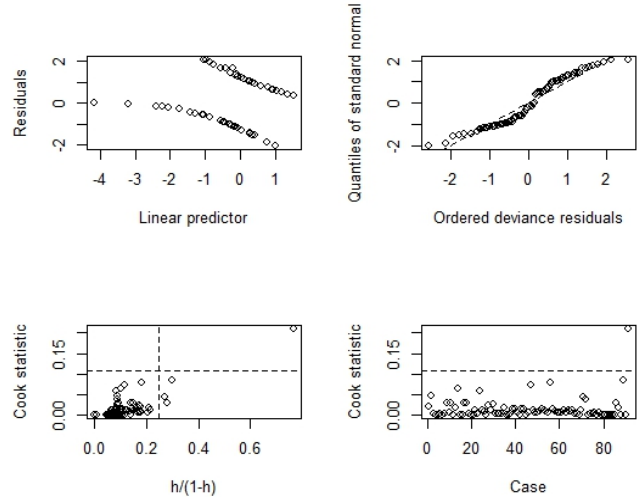Figure 2(a) Diagnostic plots of Logit model with effect modifier

Figure 2(b) Diagnostic plots of Probit model with effect modifier

4.1 Model Selection

Table 2. Model Selection for the Binomial Regression Models

Model	Link	Residual Deviance	df	AIC
A	Logit	107.07	83	123.06
B	Probit	109.92	83	123.92
C	Logit	95.836	82	113.84
D	Probit	95.587	82	113.59

The four basic models for the systematic structure of the learning style preferences data of the Sidhu School of Wilkes University students are listed in table 2. Name of the model, the residual deviances, the degrees of freedom and the AIC values of the models are given in the table. The residual deviances of 107.07 and 109.92 respectively on 83 d.f each seem to show that models A and B do not fit the data well. Introducing the interaction term into models A and B yielded models C and D respectively, and the residual deviances are now reduced to 95.84 on 82 d.f (p-value= 0.1409)for model C and 95.59 for model D also on 82 d.f with p-value=0.1409. Models C and D therefore fit the data reasonably well. If the AIC criterion is used as the criteria for model selection for the data, then the best parsimonious model is model D, with AIC value of 113.59. As pointed out earlier, model C was focused on for interpretation of result of analysis, due to its ease of interpretation.

5. Discussion

This basic aim of this analysis is to describe the way in which choice of learning style preferences by students of Sidhu School, Wilkes University, varies by age, gender, employment status, academic performance, and level of study. The purpose is to describe differences between students that prefer Behaviorist versus Humanist learning style, and to identify factors that determine the students' preference. Four separate binomial regression models proposed to analyze the data

were binomial regression model with logit link tagged model A, model with probit link tagged model B; model with logit link and interaction term tagged model C; and model with probit link and interaction term tagged model D.

Possible causes of the lack of fit found in models A and B could be that further predictors are needed to predict the response variable or an important interaction was missing, or influential outliers may be in the data. Having ruled out possibilities of influential outliers, the only possibility is the omission of important interaction term. Thus to determine factors responsible for the preferred learning style by the students at different ages correctly, the interaction of this variable must be included with their gender, because the odds ratio is not constant over gender. Worthy of note is the inclusion of the interaction variable in models A and B which resulted in models C and D, with significant reduction in deviances of 11.23 on 1 d.f (p =0.0008) and 12.33 on 1 d.f (p = 0.0004) for models C and D respectively.

Apart from one observation identified as an outlier for models C and D, the diagnostic plots showed that the models fit the data reasonably well. The value of the influence diagnostic was not larger than 1.0 so that the covariate pattern did not have an effect on the coefficients of the parameters of the models.

Acknowledgements

References

Agresti, A. (1996). *Introduction to categorical data analysis.* New-York: John Wiley & Sons.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. *Higher Education: Handbook of Theory and Research, 8,* 379-410.

Basavanthappa, B. T. (2003). *Nursing Education*. New Delphi, India: Jaypee Brothers.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: an applied perspective. *Higher Education: handbook of Theory and Research, 10*, 225-256. Cambridge: Cambridge University Press.

Chuang, H. L. (1997). High school youth's dropout and re-enrollment behavior. *Economics of Education Review, 16*(2), 171-186. https://doi.org/10.1016/S0272-7757(96)00058-1

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). Learning styles and pedagogy in post-16 learning. A systematic and critical review. London: Learning and Skills Research Centre.

Collet, D. (2003). *Modelling binary data. (2nd ed.).* London: Chapman and Hall.

Demaris, A. (1992). *Logit modeling: practical applications.* Newbury Park, CA: Sage. https://doi.org/10.4135/9781412984836

Efron, B., & Hastie, T. (2016). *Computer age statistical inference: Algorithms, evidence and data science.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316576533

Faraway, J. J. (2006). Extending the linear model with R, generalized linear, mixed effects and nonparametric regression models. New-York: Chapman & Hall/ CRC.

Hauser, L. (2005). Behaviorism. Internet encyclopedia of philosophy: A peer reviewed academic resource. Retrieved from: http://www.iep.utm.edu/behavior/

Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression (2nd ed.).* New-York: Wiley. https://doi.org/10.1002/0471722146

Krzanowski, W. J. (1998). An introduction to statistical modelling. London: Arnold.

Lawal, B. (2003). *Categorical data analysis with SAS and SPSS applications*. London: Lawrence Erlbaum Associates Publishers.

Marin, J. M., & Robert, C. P. (2014). *Bayesian essentials with R. (2nd ed.)*. New-York: Springer. https://doi.org/10.1007/978-1-4614-8687-9

McCullagh, P., & Nelder, J. A. (1998). *Generalised linear models.* London: Chapman and Hall.

McKenna, G. (1995). Learning theories made easy: Humanism. *Nurs Stand, 9*, 29–31. https://doi.org/10.7748/ns.9.31.29.s32

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 103-119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

Peng, C. Y., Manz, B. D., & Keck, J. (2001). Modelling categorical variables by logistic regression. *American Journal of Health Behaviour, 25*(3), 278-284. https://doi.org/10.5993/AJHB.25.3.15

Quinn, F. M. (2007). Principles and practice of nurse education (5th ed.). London: Nelson Thornes.

Saif, A. A. (2009). Modern educational psychology: Psychology of learning and instruction (6th ed.). Tehran: Douran Publishing.

Sobhaninejad, M. (2005). Theories of Learning and Their Applications in Process of Teaching Humanities. *Journal of Research and Planning in Higher Education, 10*, 1–9.

Swisher, K. (1990). Coperative learning and the education of American India/Alaska native students: A review of the literature and suggestions for implementation. *Journal of American India Education, 29*(2), 36-43.

Tolman, R. M., & Weisz, A. (1995). Coordinated community intervention for domestic violence: the effects of arrest and prosecution on recidivism of woman abuse perpetrators. *Crime and Delinquency, 41*(4), 481-495. https://doi.org/10.1177/0011128795041004007

Woodward, M. (2004). *Epidemiology: study design and data analysis*. Florida:Chapman and Hall/CRC.