Cypriot Journal of Educational Sciences

Volume 15, Issue 6, (2020) 1557-1568

www.cjes.eu

Specifics of logit and probit regression in education sciences - why wouldn't we use it?

Siniša Opić, [a] [*], University of Zagreb, Faculty of Teacher Education, Zagreb, Croatia https://orcid.org/0000-0001-9800-0145

Suggested Citation:
Opić, S. (2020). Specifics of logit and probit regression in education sciences - why wouldn't we use it? *Cypriot Journal of Educational Science*. *15*(6), 1557 - 1568. https://doi.org/10.18844/cjes.v15i6.5305

Received from June 25, 2020; revised from September, 15, 2020; accepted from December 15, 2020.

Abstract

Regression is one of the dominant analysis methods used in the social sciences and educational sciences. There are different regression methods based on the type of research that is being conducted. The probit and logit regression models are regression methods which are being used recently by most researchers. However, their interpretations are not straightfoward and most researchers end up misinterepreting the results from the probit and logit regression models. This research therefore aims to examine the differences between the probit and logit models, in comparison with other linear regression models. Using a comparative research design, this study utilises resources from previous researchers, hence, the study took a form of a literature review. The results of this study is essential to educational and social sciences researchers who make use of the probit, logit and other regression methods. The research also explains why logit and probit should be used in place of other regression models.

Keywords: education sciences; Linear regression; Logit; Probit; Regression

* ADDRESS FOR CORRESPONDENCE: Siniša Opić, University of Zagreb, Faculty of Teacher Education, Zagreb, Croatia
E-mail address: sinisa.opic@ufzg.hr

1. Introduction

Regression is one of the frequently utilized (tests) in social/education sciences (Temel, 2016; Rahimi, Soltani & Ghamarnia, 2020). The goal of regression analysis is to determine the role / influence of independent predictor variables on the outcome. The role of each or multiple predictors is determined and their total impact on the dependent variable is defined by the Coefficient of determination (R^2); that is, the total change in the dependent variable is explained by the predictor variables. Probit and logit regression is one of the commonly used regression models in social and educational sciences (Arican & Karahoca, 2017; Aboagye-Attah, 2019; Jeong, Siegel, Chen & Newey, 2020), apart from the spearman correlation coefficient of regression (Apipalakul, Jaimooka & Ngang, 2017; Kalashi, Bakhshalipour, Azizi & Sareshkeh, 2020). For instance, Kovacova & Kliestik, (2017) used probit and logit regression to analyze the models of predicting bankruptcy in Slovakia. Aboagye-Attah, (2019) also used binomial probit model to analyze the relationship between socioeconomic factors and poverty in Ghana.

Jeong et al. (2020) also conducted a research on the interaction between probit, logit and other non-linear regression models. In the research, they studied strategic management in terms of corporate performance and strategizing, from the perspective of probit, logit and nonlinear regression models and how the method of analyzing organization strategy is influenced by the regression method applied. A similar research was conducted by Delle Site, Kilani, Gatta, Marcucci & de Palma, (2019). Dell Site et al. (2019) however examined the best, worst, and best-worst choices in the application of probit and logit regression models.

Several researchers have used probit and logit regression models in their studies to understand the relationship of their variables (Ruspriyanty & Sofro, 2018; Kovacova & Kliestik, 2017; Jeong, Siegel, Chen & Newey, 2020). The probit and logit regression models, when used together in a research, helps the researcher to grasps an understanding of the phenomenon under study, from two different angles, thereby making the results more valid (Dagnew & Mekonnen, 2020). However, most researchers have failed to understand the difference between probit and logit models, their interpretation and how they differ from other regression models. This research therefore aims to improve knowledge of existing literature, by examining the differences between logit and probit regression models and why they should be used. To guide the research, the following research questions were set:

RQ1: What is the difference between Probit and Logit regression models

RQ2: Why should we use Probit and Logit regression models

2. Methodology

This research used a comparative research design to arrive at its findings and conclusion. The research method was a literature review. The data for this research consisted of previous research from renowned researchers in educational sciences as well as other field of studies.

The research also made use of established probit regression model formulars as well as established logit regression model formulars, which ensures validity and reliability of the study. A thorough examination is made on the different types of models that were considered for this study, with an example from playing piano, to give a better understanding of how the regression model works. The research compared logit and probit regression models to other linear regression models to give a better understanding of how reading results of probit and logit models differ from other regression models. This research went further to explain the differences between probit and logit regression models afterwards.

3. Findings from literature review

3.1 *Introduction to Regression*

Regression measures the relationship between two variables, based on the mean variable (Jeong et al., 2020). Although the coefficient of determination is a key statistic for interpreting regression model, there are still some limitations. For example, the magnitude of R^2 does not measure the magnitude of the slope of the regression line, which means that a higher value of R^2 does not imply a larger angle (steeper slope of the regression line initiates a better predictive role, which is the suitability of the regression model). Also, in nonlinear regression R^2 will often be high (if there is a high correlation between x and y), although it is not a linear dependence (Knobloch et al., 2018). This is a common omission among researchers because even before checking the suitability of regression using insight into linear relationship between variables x and y (regression line), R^2 is interpreted as a major factor in the suitability of the regression model. Moreover, since linear dependence is a rare case in pedagogical research, interpretations of R^2 as a level of predictive role initiate erroneous conclusions if the linear regression procedure is based on a connection that is in fact nonlinear. Likewise, the introduction of a new variable (or more) into the regression model increases R^2, although in reality the suitability of the regression model is not better (Opić & Galešev, 2013).

Regression (ordinal) is defined by the equation:

$$\mathbf{y=a+bx}$$

where **y** is the dependent variable, **a** is the intercept (y; value of Y when X = 0), **b** is the slope of the regression line and **x** is the independent variable. In cases where there are several predictors $\mathbf{Y=a+b_1x_{1+}b_2x_2 +\dots b_nx_n}$ (**a** is often denoted by $\mathbf{b_0}$)

The goal is to determine the regression line, that is, the line that represents new estimated values, the line which minimizes residuals (errors), sum of squares. This process of minimizing squares of distance between the observed values and a line is called Ordinal least squares (OLS). Thus, regression is an especially useful technique by which we determine new estimated values based on observed values, that is, using the regression equation for each value of x we can calculate the value of y (outcome variable), and therefore predict the value of the dependent variable. In this hypothetical model of connection between the number of hours of piano practice and errors during playing, a regression equation (figure) is shown. The determined regression line that minimized the sum of squares (residuals) shows new estimated values; for each predictor value, the number of hours of piano practice, estimated value of errors during playing is calculated. It is evident that the observed values and estimated values differ.

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	18,372	3,175		5,786	,004
	number of hours of piano practice	-2,007	,500	-,895	-4,014	,016

a. Dependent Variable: errors during playing piano

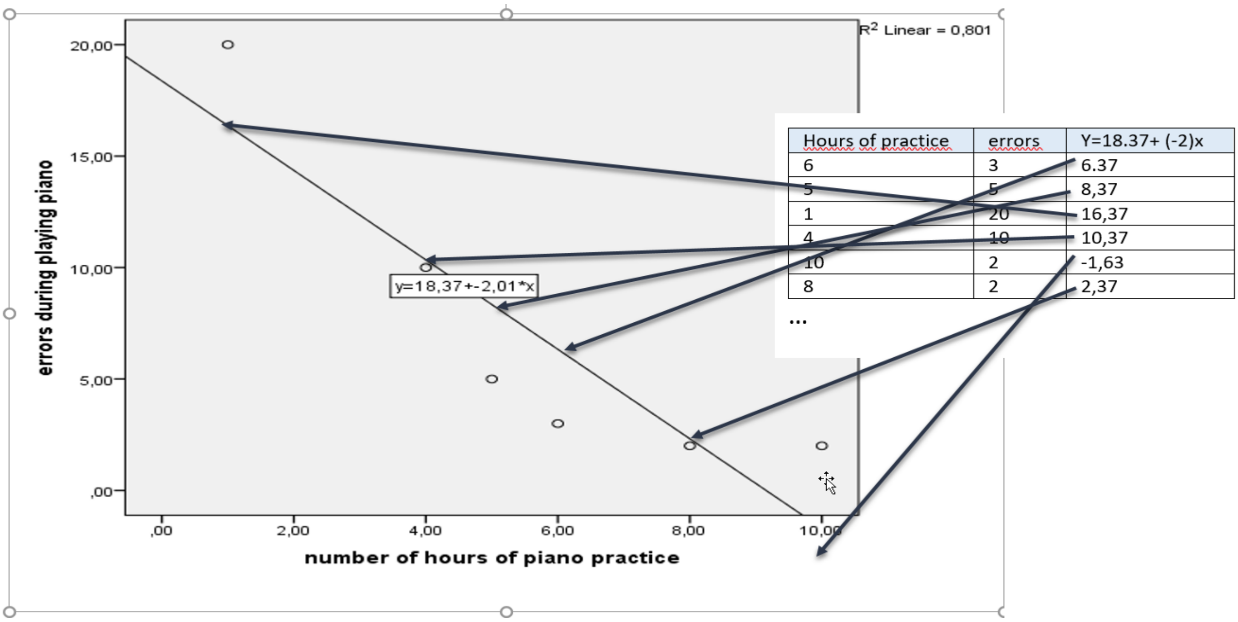

Figure 1- linear regression (OLS)

There is a visible deviation of empirical number of errors with respect to new estimated values obtained by the regression model. Using the regression equation, we can predict any value of y in the model based on x (predictors).

All the above implies that there is only one dependent variable that is a continuous variable (on at least ordinal scale). However, in case the dependent variable is dichotomous (yes-no) then we cannot apply regression analysis based on OLS, but instead approach logit and / or probit regression.

Logit regression

Logit regression (more precisely, logistic regression) is often used in medicine with the aim of, for example, testing a certain drug (predictor) for its effectiveness (outcome variable - 1 / effective; 0 / ineffective), and also in natural sciences (Sufahani & Jun 2020). However, it is very rarely used in education sciences, especially in pedagogy, although there are justifications for its application (Joyami & Salmani, 2019; Andegiorgis, 2020; Niu, 2020). The fact is that very often in the research design we have only dependent variables that are categorical / dichotomous, which makes it impossible to use tests for complex relationships between variables (usually only chi square is used when the dependent variable is categorical). But logistic regression in this case allows us to investigate the predictive role even though the dependent variable is categorical (Kim, Song, Kim, Lee & Cheon, 2018; Ozvurmaz, 2016).

Logit regression is actually a nonlinear regression (Kim et al., 2018). Logit regression (logistic) can be divided into two basic groups: binary logit in the case when the outcome variable has only two categories and multinomial logit when the dependent variable has more than two categories. Logit regression is not based on residuals (OLS), a regression line that minimizes the sum of squares, nor is R^2 calculated; instead, it is based on maximum likelihood. The maximum likelihood in the normal distribution is the position µ where the center of normal distribution is, while the maximum likelihood of σ is the standard deviation of the normal distribution (curve width). The log-likelihood is based on summing the probabilities associated with the predicted and actual outcomes (Said, Salman & Elnazer, 2019), that is, the goal is to find the best linear combination of predictors to maximize the likelihood of obtaining the observed outcome frequencies (ibidem, 440). Said et al. (2019) indicates that log likelihood is analogous to the residual sum of squares in multiple regression due to the fact that it is an indicator of how much unexplained variance remains after the model is set (this is evident from LL values which if high indicate a poorly fitting model). LL is usually multiplied by -2 because -2LL has a chi square distribution that is more suitable for comparison (χ^2= -2 (Co$_{nstant}$ –LL$_{full\ model}$).

The essence of computing Logit (and also Probit) involves understanding statistical procedures; **odds, probability and log odds**. Odds is the ratio between something happening to something not happening (p / 1-p) and probabilty is the ratio of something happening (event occurs) to everything that could happen (occurs). For example if a p=0,1 odds=0,111; p=0,2 odds=0,25; p=0,3, odds=0,4286; p=0,9 odds=9….)

Log odds is simply natural log of odds. Systematic view of equibalance between probabilities, odds and log odds is presented in table 1 (Menard, 2010,15).

Table 1 - equilalance between probabilities, odds and log odds (Menard, 2010,15)

Definition: **Probability that $(Y=1)=\Pi(Y=1)=P(Y=1)$**	Probabilities to Odds: $\Omega(Y=1)=P(Y=1)/[1-P(Y=1)]$	Probabilities to Logits: $\text{logit}(Y=1)=\ln[P(Y=1)/[1-P(Y=1)]$
Odds to Probabilities: $P(Y=1)=\Omega(Y=1)/[1+\Omega(Y=1)]$	**Definition:** **Odds that $(Y=1)=\Omega(Y=1)=\text{Odds}(Y=1)$**	Odds to Logits: $\text{logit}(Y=1)=\ln[\Omega(Y=1)]$
Logits to Probabilities: $P(Y=1)=e^{\text{logit}(Y=1)}/[1+e^{\text{logit}(Y=1)}]$	Logits to Odds: $\Omega(Y=1)=e^{\text{logit}(Y=1)}$	**Definition:** **Logit that $(Y=1)=\Lambda(Y=1)=\text{logit}(Y=1)$**

Main differences between OLS and ML are shown in Table 2 (Menard, 2010,342)

Table 2 – diffreneces between OLS and ML (Menard, 2010,342)

	Ordinary Least Squares/Linear Regression	*Maximum Likelihood/Logistic Regression*
What the model is trying to predict	*Exact observed value of dependent variable*	*Probability of obtaining observed value of dependent variable*
Criterion for selecting model parameters $(a, b_1, b_2, \ldots, b_K)$	Select parameters to minimize the sum of squared errors, $\Sigma(e_j)^2 = \Sigma(Y_j - \hat{Y}_j)^2 =$ $\Sigma[Y_j - (a + b_1X_{1j} + \cdots + b_KX_{Kj})]^2$	Select parameters to maximize the likelihood of obtaining the observed values of the dependent variable, given the independent variables: (Equation here in bivariate form) $l(Y) = \prod P(Y_j \mid X_j) = \prod[P(Y_j = 1 \mid X_j)]^{[P(Y_j=1)]}[1 - P(Y_j = 1 \mid X_j)]^{[1-P(Y_j=1)]}$ $= \prod[e^{a+bX_j}/(1 + e^{a+bX_j})]^{[P(Y_j=1)]}[1/(1 + e^{a+bX_j})]^{[1-P(Y_j=1)]}$ Equivalently, select parameters to maximize the natural logarithm of the likelihood, $\ln[l(Y)]$: $L(Y) = \ln[l(Y)] = \Sigma[P(Y_j = 1 \mid X_j)$ $= \Sigma\{P(Y_j = 1)\ln[e^{a+bX_j}/(1 + e^{a+bX_j})] + [1 - P(Y_j = 1)\ln[1/(1 + e^{a+bX_j})]\}$
Process of estimating parameters	Solve directly for parameters using calculus and linear algebra	Solve for parameters iteratively, using calculus, linear algebra, and numerical algorithms that propose, test, and modify tentative solutions until the solutions converge
Form of the function	Linear	Logistic
Properties of coefficients (if assumptions are correct)	Best linear unbiased estimates for $a, b_1, b_2, \ldots, b_K$	Asymptotically efficient, unbiased estimates for $a, b_1, b_2, \ldots, b_K$

OLS estimates for the linear regression coefficient are identical to the estimates one would obtain using maximum likelihood estimation (Elisaon, 1993:13-18 as cited by Österlund, 2020).

The logit regression formula is:

Logit (odds)= $\mathbf{ln}\left(\frac{p(y=1)}{1-p(y=1)}\right)$ $=\mathbf{a+\beta_1x_1+\ \beta_2x_2....\ \beta_nx_n}$ that is, ln(p)-ln(1-p) =logit (p).

Therefore $\frac{p(y=1)}{1-p(y=1)}$ **= exp (a+** $\mathbf{\beta_1x_1+\ \beta_2x_2....\ \beta_nx_n)}$

where Ln is the natural logarithm of the odds of Y (dependent variable). Since the dependent variable is binomial / dichotomous (1,0) it follows the Bernoulli distribution / (0,1); the Bernoulli distribution is a special form of Binomial distribution which is given by **P(x I,n, p,) = {n! / [x! (n - x)!]} * $\mathbf{P^x}$ * (1 - P)** $^{n-x}$

In logistic regression, we estimate the probability for each linear combination of independent variables.

Accordingly, the Probit regression formula: probit (y)= Φ^{-1}[P(Y=1)]; that is,

$$\Phi(X) = (1/\sqrt{2\pi}) \int_{-\infty}^{X} e^{-z^2/2}\, dz.$$

,where Φ is Cumulative Distribution Function (CDF) of the standard normal distribution.

The probit function is the inverse of the cumulative distribution function of the standard normal distribution (y)= Φ^{-1} that is, probit^{-1}(x) = Φ(x).

In the previous example of linear regression (OLS; figure 1) we had a linear regression model with a dependent variable being the number of errors during piano playing while the predictor variable is practice hours. But very often in pedagogy we use categorical variables (dichotomous, trichotomous or polytomous). In the following example, we have the independent variable number of hours of piano practice while the dependent variable is categorically dichotomous; concentration during playing (1,0). The question is if we can predict whether the musician is concentrated or not given the hours of practice. The obtained value of the model (with 7 iterations) is -2LL = 75.393, Cox & Snell R Square = 0.622 and Nagelkerke R Square = 0.832. The value of Nagelkerke R Square is similar to R^2 (coefficient of determination) in linear regression; the higher the value, the higher the variance explained by the model, and the model fits the data better. When Cox & Snell R Square value is 1, it means that the model is "perfect" which is only possible to expect in theory. It is based on a comparison of the LL of the new model with the LL of the original (baseline); $\boldsymbol{R^2_{C\&S} = 1 - (L_0 / L_M)^{2/n}}$. But there is still a dilemma as to which value of R^2 in logit regression is the most reliable. The values Cox & Snell R Square and Nagelkerke R Square are just some of the most commonly used although Mittlbock and Schemper (1996 as cited by Mood, 2017) reviewed 12 different measures. R Square, which is quite easy to calculate, as $\boldsymbol{R^2_L}$**= -2LL (model)/-2LL (original)**, is also used. But we can call all these values pseudo R^2 and one should be careful in interpretation. Thus, for example, McFadden`s $\rho^2$ tends to be smaller than $R^2$ in linear regression and the range of values 0.2-0.4 should be considered satisfactory (Paetz, Hein, Kurz & Steiner, 2020); McFadden`s ρ^2 is calculated as **1- LL (model)/LL (base model)**. In this example, both values are relatively high, which indicates that the model fits the data well.

Figure 2 shows a graph of the relationship between respondents and log probability, which is an insight into our model (but not with predicted probabilities, yet).

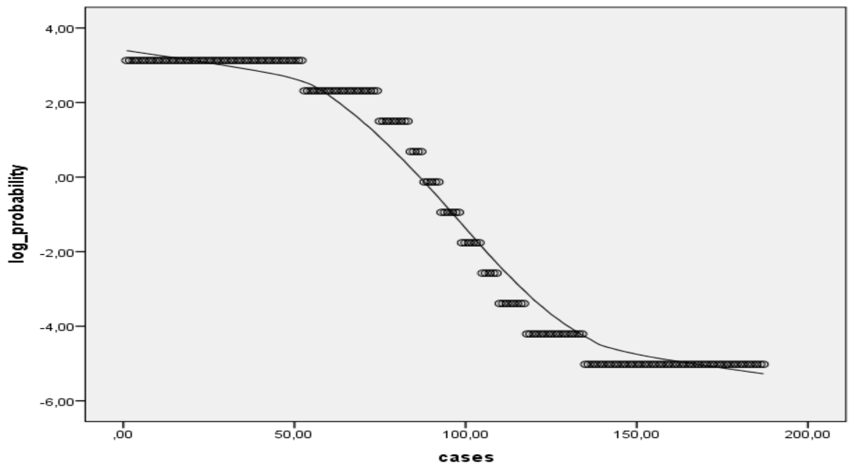

Figure 2 - log probabilities

As can be seen in Figure 2, it is an "S" shape curve (sigmoid), which is a prerequisite for a "good" model. If we approached the same example with linear regression analysis, we would get a direction that does not present us with the accurate value of probabilities of the outcome of the dependent variable; concentrated (yes, no).

In addition to the pseudo-R square value; Cox & Snell R Square and Nagelkerke R Square; other values are key to understanding the model. Wald statistic shows whether the value of the b coefficient of the predictor differs from 1; if it is statistically significantly different then we assume that it statistically significantly predicts the value of the dependent variable (1,0) (**Wald=b/SEb**). Because it is based on standard error it is identical to t value in linear regression. In this example B = 0.815; S.E. = 0.114; df = 1; sig. 0.000. **Exp (B) value,** exponentiation of the B coefficient is an odds ratio and is a simpler parameter to interpret than Beta (Exp (B) = 2.259). Odds ratio can be displayed as:

$$\text{Odds ratio} = \frac{\left[\dfrac{P(Y=1)}{1-P(Y=1)}\right]_X}{\left[\dfrac{P(Y=1)}{1-P(Y=1)}\right]_{X-1}}$$

where **p(Y=1) =**$\frac{1}{1+e^{-xb}}$

If the odds ratio (Exp (B)) is greater than 1 it initiates that the odds of getting a 1 on the dependent variable grows when the independent variable / predictor grows, while when the value is less than 1 the odds of getting a 1 on the dependent variable decreases when the independent variable grows (adapted according to Weisburd & Britt, 2007). In the given example Exp (B) = 2.259 tells us that as the odds of predictor grow so will the odds of dependent variable (concentration).

Since statistical values have shown the suitability of the predictor model; and we also got a curve in approximately "S" shape the most important part is to show predicted probabilities (Figure 3)

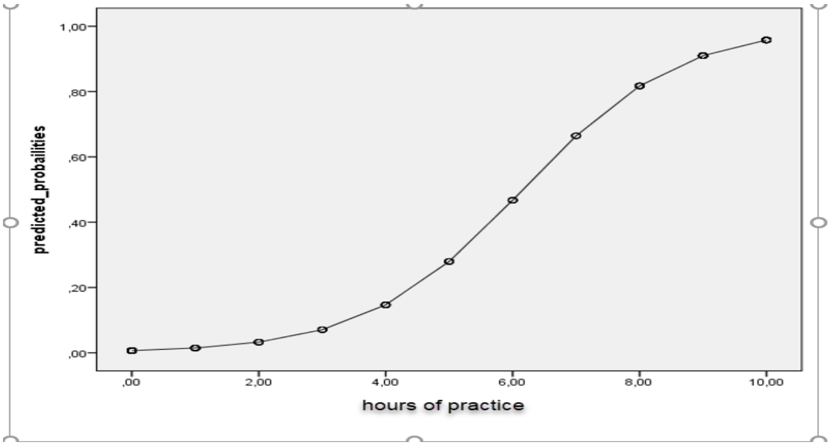

Figure 3 - predicted probabilities

In Figure 3 we have one of the main indicators of the model, and that is predicted probabilities. The curve is not exactly perfectly sigmoid ("S" shape), but the shape characteristic of logit regression is retained. As can be seen from the scatter plot, the probability of being concentrated increases over 6 hours of practice (more than 50%). Thus, for example, those who practice for 7 hours have a 70% probability of feeling concentrated, while those who practice for 4 hours do not feel concentrated (only 15%). Thus, the number of hours of playing (practicing) is a predictor of whether they will (1) or will not (0) feel concentrated. If the curve were perfectly sigmoid, then the beginning and end of the curve would be flat, which would mean the lowest and highest probabilities for different values of the predictor variable, and there would also be a sharp increase in probability. The listed probabilities were calculated using syntax in SPSS, which makes it difficult for beginners to use (present) exact probabilities for each value of the independent variable. In addition, graphical representation of observed groups and predicted probabilities provides an insight into the affiliation of groups of respondents into predicted probabilities. (figure 4)

```
              Observed Groups and Predicted Probabilities

      80 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R     60 +                                                                                                    +
E        I1                                                                                                   I
Q        I1                                                                                               2   I
U        I1                                                                                               2   I
E     40 +1                                                                                               2   +
N        I1                                                                                               2   I
C        I1                                                                                               2   I
Y        I1                                                                                               2   I
      20 +1                                                                                          2    2   +
         I11                                                                                         2    2   I
         I11 2                                                                              2        2    2   I
         I11 1   1      1            1                  1                   2               1        2    2   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        ,1        ,2        ,3        ,4        ,5        ,6        ,7        ,8        ,9         1
  Group:  111111111111111111111111111111111111111111111111112222222222222222222222222222222222222222222222222222

          Predicted Probability is of Membership for 2,00
          The Cut Value is ,50
          Symbols: 1 -  not concentrated
                   2 -  concentrated
          Each Symbol Represents 5 Cases.
```

Figure 4 - Observed groups and Predicted probabilities

As can be seen, the cut value is 0.5, which means that the probability of concentration during playing in the subjects is mostly grouped with a high probability of 0.9 (observed groups). According to the Cat Value and the code of the variable (1 - not concentrated) we can see groups of respondents (left end of the curve) which according to the probability are not concentrated. The more respondents are grouped at the ends of the curve we can (as is the case here) see better (occur) or worse (not occur) probability. Additional confirmation of the accuracy of the predictive model is confirmed by the ROC plot (figure 5)

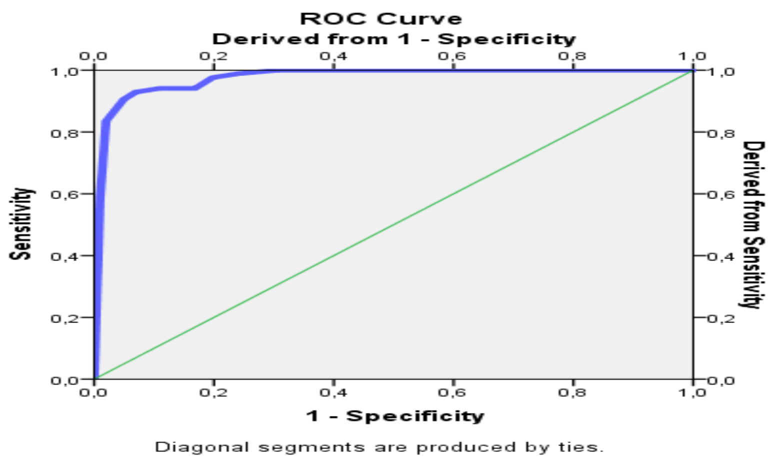

Figure 5 - ROC (AUC)

AUC (Area under curve) is 0,976 (95% IC= 0,957 to 0,995), which confirms the model to be excellent, that is, the validity of the model (true positive rates and false positive false).

Differences between logit and probit

Logit and probit are essentially similar statistical approaches and the empirical predicted probabilities obtained are similar. Chambers and Cox (1967, as cited by Sen, Al-Mofleh & Maiti, 2020) , after investigating logit and probit concluded that differences between these functions (differences in arithmetic means) are visible only when the sample of subjects is greater than 1000. The essential difference between these regression functions is in conversion to z values. Probit treats predicted probabilities as cumulative probabilities of normal distribution and converts them into z scores. Therefore, a probability of 0.025 becomes approx. -1.96 and the probability of 0.5 becomes 0.

The main difference between the logit and probit functions is in the interpretation of the odds ratio which is valid only in logistic regression (Smithson and Merkle, 2014). German (2008 as cited by Klieštik, Kočišová & Mišanková, 2015) emphasizes two basic advantages of logistic regression over probit, and these are simplicity and interpretability. In graphical representation of the logit and probit curves, the probit function curve function has slightly fatter tails (figure 6)

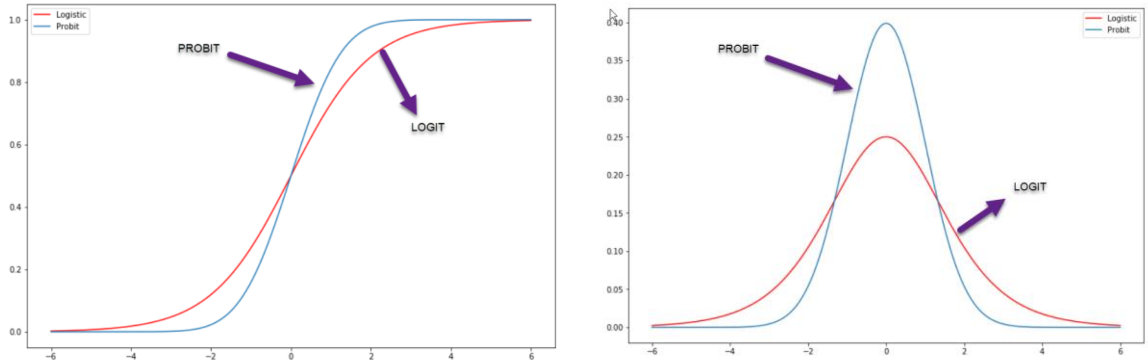

Figure 6 - logit and probit distribution

Source: Discrete Choice Models

Although assumptions for using logit and probit are not as strict as in linear regression, the following prerequisites should be met (Meyers, Gamst, Guarino, 2016): absence of perfect multicollinearity, no specification errors and the dependent variable should be on ordinal or ratio scale (although it can also be a dichotomous variable). Also, the number of respondents should be at least 30 times higher than the number of estimated parameters / variables (Mehta et al., 2020). In addition to the above prerequisites Wright (1995) adds that scores on the dependent variable must be statistically independent of each other. Unlike linear regression, it is not necessary to have normally distributed values of the dependent variable or homogenic variances of the dependent variable along the categories of predictors. Same as with Chi square test (2 * 2 contingency tables), logit regression (binary) will not give valid results if one or more cells have expected frequencies value less than 5. Additionally, if outliers are present (if the value range on the predictor variable is large) we can also get invalid results (Warner, 2013).

CONCLUSION

The nature of variables in education sciences is often dichotomous, binary, which, when it comes to dependent variables, makes it impossible to use complex statistical tests. Very often, a priori, in research design, researchers tend to make dependent variables discrete on an ordinal most often 5-degree scale. But sometimes this is not justified because the very nature of the research requires that the dependent variables are binary (yes / no, 0/1). In this case, chi square is most often resorted to, and it is based on determining the differences between observed and theoretical f. But when we want to investigate the predictive role of independent variables on the dependent (binary) then logit and probit regression is the solution. Linear regression based on ordinary least squares is in principle similar to logit and probit regression, and logit and probit regression is actually a specific form of linear regression. In fact, we could apply the regression line to minimize the sum of squares (residuals), but it would not be as suitable for interpretation due to sigmoid curve ("S" shape).

Logit and probit regression are unjustifiably rarely used in education sciences, although often the nature of the dependent variables is such that logit and probit regression are an immensely powerful statistical procedure (test) for determining the predictive role, which is the predicted probabilities for dependent variables (1,0). The reason for non-use can lie in a more complex mathematical background (log likelihood, predicted probabilities, log probabilities, odds ratio) and interpretation of statistics in the analysis.

References

Aboagye-Attah, K. (2019). *Socioeconomic correlates of poverty in Ghana using Ghana Living Standards Survey round 6 and 7* (Master's thesis, Norwegian University of Life Sciences, Ås). http://hdl.handle.net/11250/2609325

Andegiorgis, G. E. (2020). Counselling approaches used in solving students' disciplinary problems in secondary schools in Keren sub-zone, Anseba region, Eritrea. Global Journal of Psychology Research: New Trends and Issues, 10(1), 95–101. https://doi.org/10.18844/gjpr.v10i1.4263

Apipalakul, C., Jaimooka, E., & Ngang, T. K. (2017). The effect of community participation on conflict management. Global Journal of Sociology: Current Issues, 7(2), 95–103. https://doi.org/10.18844/gjs.v7i2.2394

Arican, E., & Karahoca, A. (2017). Comparing prediction algorithms in disorganized data. Global Journal of Computer Sciences: Theory and Research, 6(2), 26–35. https://doi.org/10.18844/gjcs.v6i2.1471

Chan, H. C., Chang, C. C., Chen, P. A., & Lee, J. T. (2019). Using multinomial logistic regression for prediction of soil depth in an area of complex topography in Taiwan. *Catena*, *176*, 419-429. https://doi.org/10.1016/j.catena.2019.01.030

Dagnew, A., & Mekonnen, D. (2020). Effect of using guided inquiry teaching method in improving grade eight students' concept of photosynthesis, primary school: Ethiopia. International Journal of Innovative Research in Education, 7(1), 01–15. https://doi.org/10.18844/ijire.v7i1.4655

Delle Site, P., Kilani, K., Gatta, V., Marcucci, E., & de Palma, A. (2019). Estimation of consistent Logit and Probit models using best, worst and best–worst choices. *Transportation Research Part B: Methodological*, *128*, 87-106. https://doi.org/10.1016/j.trb.2019.07.014

Discrete Choice Models¶. (n.d.). Retrieved December 13, 2020, from https://www.statsmodels.org/devel/examples/notebooks/generated/discrete_choice_example.html

Jeong, Y., Siegel, J. I., Chen, S. Y. P., & Newey, W. K. (2020). A recentering approach for interpreting interaction effects from logit, probit, and other nonlinear models. *Strategic Management Journal*, *41*(11), 2072-2091. https://doi.org/10.1002/smj.3202

Joyami, E. N., & Salmani, D. (2019). Value creation for students using the communication marketing approach at Tehran University. New Trends and Issues Proceedings on Humanities and Social Sciences, 6(7), 88–100. https://doi.org/10.18844/prosoc.v6i7.4517

Kalashi, M., Bakhshalipour, V., Azizi, B., & Sareshkeh, S. K. (2020). The effect of the application of ICT skills on the process of knowledge management components and the effectiveness of creativity indicators for the improvement of employees' performance system in the Ministry of Sports and Youth. World Journal on Educational Technology: Current Issues, 12(1), 48–61. https://doi.org/10.18844/wjet.v12i1.4382

Kim, A., Song, Y., Kim, M., Lee, K., & Cheon, J. H. (2018). Logistic regression model training based on the approximate homomorphic encryption. *BMC medical genomics*, *11*(4), 83. https://doi.org/10.1186/s12920-018-0401-7

Klieštik, T., Kočišová, K., & Mišanková, M. (2015). Logit and probit model used for prediction of financial health of company. *Procedia economics and finance*, *23*, 850-855. https://doi.org/10.1016/S2212-5671(15)00485-2

Knobloch, G., Colgan, T., Wiens, C. N., Wang, X., Schubert, T., Hernando, D., ... & Reeder, S. (2018). Relaxivity of Ferumoxytol at 1.5 T and 3.0 T. Investigative radiology, 53(5), 257. https://dx.doi.org/10.1097%2FRLI.0000000000000434

Kovacova, M., & Kliestik, T. (2017). Logit and Probit application for the prediction of bankruptcy in Slovak companies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(4), 775-791. http://dx.doi.org/10.24136/eq.v12i4.40

Mehta, R., Rice, S., Winter, S. R., Anania, E. C., Baugh, B. S., & Milner, M. N. (2020). Determining the predictors for ease of sleep while on aircraft: Regression and qualitative analyses. Journal of Air Transport Management, 83, 101756. https://doi.org/10.1016/j.jairtraman.2019.101756

Menard, S. (2010). *Logistic Regression – from Introductory to Advanced Concept and Applications.* Sage publication. https://doi.org/10.15358/0344-1369-2019-4-3

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). Applied multivariate research: Design and interpretation. Sage publications. https://us.sagepub.com/en-us/nam/applied-multivariate-research/book246895

Mood, C. (2017). Logistic regression: Uncovering unobserved heterogeneity. Working Paper. https://www.su.se/polopoly_fs/1.341160.1501927239!/menu/standard/file/Logit2%20%286%29.pdf

Niu, L. (2020). A review of the application of logistic regression in educational research: Common issues, implications, and suggestions. Educational Review, 72(1), 41-67. https://doi.org/10.1080/00131911.2018.1483892

Opić, S., & Galešev, V. (2013). Specifics of application of regression analysis in research of education (pedagogy). Pedagogy and Culture, 183-192. https://www.bib.irb.hr/595561?rad=595561

Österlund, V. (2020). Missing Data-A Gentle Introduction. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1444918&dswid=1401

Ozvurmaz, S. (2016). Identifying Ergonomic Risk Factors in Bank. New Trends and Issues Proceedings on Advances in Pure and Applied Sciences, (7), 123–129. https://doi.org/10.18844/gjpaas.v0i7.3170

Paetz, F., Hein, M., Kurz, P., & Steiner, W. (2020). Latent Class Conjoint Choice Models: A Guide for Model Selection, Estimation, Validation, and Interpretation of Results. Marketing ZFP, 41(4), 3-20. https://doi.org/10.15358/0344-1369-2019-4-3

Rahimi, A., Soltani, A., & Ghamarnia, M. (2020). Enhancing EFL teachers' intercultural sensitivity through instruction, is it really feasible? An ethnocentrism versus ethnorelativism perspective. Contemporary Educational Research Journal, 10(3), 88–96. https://doi.org/10.18844/cerj.v10i3.4782

Ruspriyanty, D. I., & Sofro, A. (2018, November). Analysis of hypertension disease using logistic and probit regression. In Journal of Physics: Conference Series (Vol. 1108, No. 1, p. 012054). IOP Publishing. https://iopscience.iop.org/article/10.1088/1742-6596/1108/1/012054/meta

Said, I., Salman, S. A., & Elnazer, A. A. (2019). Multivariate statistics and contamination factor to identify trace elements pollution in soil around Gerga City, Egypt. Bulletin of the National Research Centre, 43(1), 43. https://doi.org/10.1186/s42269-019-0081-2

Sen, S., Al-Mofleh, H., & Maiti, S. S. (2020). On discrimination between the Lindley and xgamma distributions. Annals of Data Science, 1-17. https://doi.org/10.1007/s40745-020-00243-7

Smithson, M., Merkle, E.C. (2014*). Generalized Linear Model for Categorical and Continuous Limited Dependent Variables.* Taylor and Francis group *(CRC group). https://doi.org/10.1201/b15694*

Sufahani, S. F., & Jun, T. L. (2020). A Comparison Between Probit And Logit Regression in Identifying Risk Factors of Cervical Cancer in Public Hospital. International Journal of Advanced Computer Systems and Software Engineering, 1(2), 12-19. https://journal.scientiaca.org/index.php/ijacsse/article/view/147

Temel, S. (2016). Perceived problem-solving skills: As a predictor of prospective teachers' scientific epistemological beliefs. Cypriot Journal of Educational Sciences, 11(3), 111–118. https://doi.org/10.18844/cjes.v11i3.775

Warner, R.M. (2013). *Applied Statistics – from Bivariate Through Multivariate Techniques.* Second edition. Sage. https://psycnet.apa.org/record/2007-00676-000

Weisburd, D., & Britt, C. (2014). Statistics in criminal justice. Springer. http://ir.mksu.ac.ke/handle/123456780/6022