ISSN 2286-4822
www.euacademic.org

EUROPEAN ACADEMIC RESEARCH
Vol. III, Issue 4/ July 2015

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

Proposed Models to Target the Poor First Year Students of Polytechnic University of the Philippines, Main Campus for SY 2014 - 2015 Using Proxy Means Test

HERO L TOLOSA
Master in Applied Statistics Student
Polytechnic University of the Philippines, Graduate School
Philippines
Dr. DIVINA GRACIA L. DEL PRADO
Adviser
Chief Labor and Employment Officer
Program Management and Technical Support Services Division
Bureau of Labor and Employment Statistics, Philippines

Abstract:

This study sought to identify the non-income variables that can predict the economic status of the first year students, SY 2014-2015 in Polytechnic University of the Philippines and determined what method is better between Ordinary Least Squares or OLS (robust standard error) regression, and Logistic regression. The researcher used the database of Polytechnic University of the Philippines as secondary data to determine the economic status of the first year students, SY 2014-2015. The researcher used OLS (robust standard errors) regression, and Logistic regression to predict the economic status of the said students. The 50 initial simulation processes are used to determine which method is better. After performing 50 simulation process for both OLS (robust standard error) and logistic regression, the researcher hereby concludes that the OLS (robust standard error) regression model is better than logistic regression as a PMT in identifying whether a student is poor or non-poor, based on the non-income variable of poverty. Majority of the first year students' fall under the non-poor category using the OLS (robust standard error) method.

Key words: poor first year students, Polytechnic University, the Philippines

1. Introduction

Poverty and inequality remain a perennial challenge in the Philippines in the wake of the current global financial crisis and increase in basic commodity prices such as food, fuel, water, shelter, and the like. The proportion of households living below the official poverty line has declined very slowly and unevenly in the past four decades, and poverty reduction has been much slower compared to neighboring countries such as the People's Republic of China (PRC), Indonesia, Thailand, and Vietnam. The growth of the economy characterized by the boom and bust cycles and current episodes of moderate economic expansion limit the impact on poverty reduction. Other reasons for the relatively moderate poverty reduction include the high rate of inequality across income brackets, regions, and sectors, and unmanaged population growth (Asian Development Bank, 2009).

Poverty incidence is the proportion of people below the poverty line to the total population. In the Philippines, poverty incidence estimated at 27.9% during the first semester of 2012. Comparing this with the 2006 and 2009 first semester figures, which was estimated at 28.8% and 28.6%, respectively. Poverty rates remains unchanged as the computed differences are not statistically significant (Albert, 2013). The source of basic data for this indicator is the Family Income and Expenditure Survey (FIES). However, starting 2012, the Annual Poverty Indicator Survey conducted with no FIES is an alternative source of data on poverty incidence.

Several reasons why poverty incidence remains unchanged were the weaknesses in the implementation of poverty alleviation programs from the government and other private agencies. Among these weaknesses were the lack of assurance that the poor will receive the benefits allotted for them, lack of coordination among social protection agencies, and no established unified criteria for identifying the poor in the country (DSWD-NCR, 2010).

One of the projects created by the government to address poverty in the country is the National Household Targeting System for Poverty Reduction (NHTS-PR) under the Department of Social Welfare and Development. NHTS-PR is the system or database that identifies who and where the poor are in the country. The database also used to determine the profile of the poor household and provide social protection programs such as education.

Poverty is one of the major reason that can affect a child's development and educational outcomes that begin in their earliest years of life, both directly and indirectly through mediated, moderated, and transactional processes. Poverty and poor education strengthen each other in a vicious cycle; but perhaps, working on one may alleviate the other. Education provides an opportunity to break and escape the cycle of poverty (Philippine Basic Education-Blogspot, 2012).

Poor students were privileged to enrol in the Polytechnic University of the Philippines because of its lowest tuition (12 pesos per unit) and other fees among the schools in the country. However, despite the low fees, still, there were students who were not able to finish their college education because of the incapacity of the parents to sustain the educational requirements of their children like tuition fees, allowances, and other financial needs. Hence, the researcher was motivated to develop a strategy to target students who need financial and educational assistance through PMT.

Currently, the University has a database of all first year students that are enrolled every year. This database consists of socio-economic profile of the students such as their Family Background, Family Economic Status and Student Educational

Data. This database serves as a record and monitor tool to the number of students in the University. The researcher utilized the database to develop a model that can determine the economic status of the students in the University.

Proxy Means Test (PMT), which based on national household survey, uses multivariate regression to correlate certain proxies such as assets and household characteristics with poverty and income. Given that household income in developing countries is often difficult and expensive to measure accurately, the methodology relies on household assets and other indicators or proxies to estimate household welfare. They include demographic characteristics (age of household members and size of household), human capital characteristics (education of household head and enrolment of children in school), physical housing characteristics (type of roof or floor), durable goods (i.e. refrigerators, televisions or cars), and productive assets (land or animals). Regression runs to find the proxies that most correlate with welfare; while individual proxies may be weakly correlated with welfare, multiple proxies show stronger correlations. The PMT uses a set of proxies (usually between 10 and 30) that best explained welfare. Each proxy gives a weight based on its estimated impact on household expenditure. Enumerators visit households to see if they have the proxies use in the PMT. Then, using the agreed weights, a score calculated for each household. Households that score below the cut-off point are eligible for the social protection program considered (Australian Agency for International Development (AusAID), Canberra, September 2011).

2. Statement of the Problem

The main problem of this study is to create a model and database that can identify the economic status of the first year students of PUP, Main Campus for SY 2014-2015 using Proxy Means Test (PMT). Specifically, this study seeks the following:

1. What are non-income variables that can predict the economic status, whether poor or non-poor, of PUP first year students using logistic regression and (OLS) regression?
2. Which of the two regression methods (logistic and OLS regression) is better as PMT in targeting poor PUP first year students?
3. What is the economic status of PUP first year student for the school year 2014-2015 using PMT.?

Hypothesis

This study has the following hypothesis as basis for its decision and conclusion:

Ho: All non-income variables considered in this study are not predictors of economic status of the first year student of PUP, Main Campus for SY 2014-2015.

or

H_0:The ß_i^' s=0, i=1,2,3,..,k

where k-1 is the number of non-income variables

3. Methodology

This chapter presents the structured process of this research. This covers the procedures followed to gather, analyze, and interpret the data, which includes the method of research, population, sample size, sampling technique, data gathering procedure, and statistical treatment of data..

3.1. Method of Research Used

This study used a quantitative method as a means of analysis. According to Neuman, (2007 as cited by Bellenas, 2014) these are the straightforward sequence for quantitative research flows: conceptualization, operationalization, application of the operational definition or measurement to collect the data. He also added that quantitative researchers established numerous ways thoroughly link abstract ideas to measurement procedures that will produce precise quantitative information about empirical reality. Moreover, quantitative data collection methods are centered on the quantification of relationships between independent and dependent variables. The quantitative approach was used as it helps the researcher to prevent bias in gathering and presenting the data. Its procedures created a logical hypothesis that was objective and complete, which could only be realized by means of surpassing individual perspective.

Descriptive and Inferential Statistics are both part of quantitative methods. Descriptive statistics were used to describe the basic features of the data in a study, and provides summaries about the sample; while in inferential statistics, we were trying to reach conclusions that extend beyond the immediate data alone (Trochim, 2006). For instance, the researcher used inferential statistics to try to infer from the sample data the characterization of the population and make judgments of the probability from an observed difference between groups that were dependable one or one that might have happened by chance in this study to a more general condition.

3.2. The Data

The researcher used the available profile of students taken from Institute for Data and Statistical Analysis (IDSA) of the Polytechnic University of the Philippines, Main Campus for SY 2014-2015. The said profile consists of three different parts namely: Family Background, Family Economic Status, and Student Educational Data.

3.3. Ordinary Least Squares (OLS) Multiple Regression

Multiple regression model is needed in order to predict the values of the response variable. For the case of k independent variables$X_1, X_2, \ldots, X_k$ the mean of response variable Y $| x_1, x_2, \ldots, x_k$ is given by the multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

where: $X_i = non - income\ variables,\ i = 1,2,3,..,k$
$\beta_j = regression\ coefficient\ of\ regression.\ j = 1,2,3,..,k$
$\varepsilon = error\ term\ or\ residual$

and the estimated response is obtained from the sample regression equation

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_k x_k$$

where each regression coefficient $\beta_j$is estimated by b_j from the sample data using the method of least squares.

Estimating the Coefficients

Each observation $x_{1i}, x_{2i}, \ldots, x_{ki}$ is assumed to satisfy the following equation.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + e_i$$

or

$$y_i = \hat{y}_i + e_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki} + e_i$$

Where y_i and e_i are the observed values and random errors, respectively. As in the case of simple linear regression, it is assumed that the $e_{i's}$ are independent and identically distributed with mean 0 and common variance σ^2 . In using the concept of least squares to arrive at estimates $b_0, b_1, \ldots, b_k$ for the regression coefficients, we minimize the expression

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i} - \cdots - b_k x_{ki})^2$$

where $\sum$ represents the symbol for summation

Differentiating SSE in with respect to bj and equating to zero, we generate the set of $k+1$ normal equations for multiple linear regression.

3.3.1 Normality of error terms

Regression assumes that variables have normal distributions. Non-normally distributed error terms (highly skewed or kurtotic variables, or variables with substantial outliers) can distort relationships and significance tests.

The researcher used using the normal probability plots, or normal QQ plots for non-formal way (graphical test) in determining if the error terms are normally distributed. For formal test, Kolmogorov-Smirnov tests and Shapiro Wilk tests are used to provide inferential statistics on normality. If our computed test is NOT significant, then the data follows normal distribution, so any value above or equal 0.05 indicates normality. Another formal way of testing for normality to support the result is using the skewness and kurtosis values. It can be calculated to assess the degree to which skewness and peakedness of the distribution vary from the normal distribution.

Violations of normality often arise either because (a) the distributions of the dependent and/or independent variables are themselves significantly non-normal, and/or (b) the linearity assumption is violated. In such cases, a nonlinear transformation of variables might cure both problems. In some cases, the problem with the residual distribution is mainly due to one or two very large errors. Such values should be examined closely. If they are merely errors, or if they can be explained as unique events not likely to be repeated, then they might be removed from dataset. In some cases, however, it may be that the extreme values in the data provide the most useful information about values of some of the coefficients and/or provide the most realistic guide to the magnitudes of forecast errors (Hair, et.al, 2010).

3.3.2 Linearity of the relationship between dependent and independent variables

Standard multiple regression can only accurately estimate the relationship between dependent and independent variables if the relationships are linear in nature. If the relationship between independent variables and the dependent variable is not linear, the results of the regression analysis will underestimate the true relationship. This underestimation carries two risks: increased chance of a Type II error for that independent variable, and in the case of multiple regression, an increased risk of Type I error (overestimation).

The most common way to assess linearity is to examine scatterplots of the variables and to identify any nonlinear patterns in the data. The points should be symmetrically distributed around a diagonal.

It is important that the nonlinear aspects of the relationship be accounted for in order to best assess the relationship between variables. If a nonlinear relationship is detected, the most direct approach is to transform one or both variables to achieve linearity. An alternative to data transformation is the creation of new variables to represent the nonlinear portion of the relationship (Hair, et.al, 2010).

3.3.3 Homoscedasticity

The assumption of homoscedasticity (or constant variance) is central to linear regression models. Homoscedasticity describes a situation in which the error term, (that is, the "noise" or random disturbance in the relationship between the independent variables and the dependent variable), is the same across all values of the independent variables (Hair, et.al, 2010).

A graphical test is done in checking homoscedasticity is by examining the scatterplot of the residuals against the predicted values. The presence of pattern or trend in the scatterplot suggests for presence of heteroscedasticity. A formal test is used is the Breusch-Pagan test, and Cook-Weisberg test.

3.3.4 Independence of the errors terms

The error term in a regression equation represents the effect of the variables that were omitted from the equation. This assumption states that an error from one observation (e_j) is independent of the error from another observation (e_j).

The researcher looked at the plots of the residuals versus independent variables, or plots of residuals versus row number in situations where the rows have been sorted or grouped in some way that depends (only) on the values of the independent variables. The residuals should be randomly and symmetrically distributed around zero under all conditions, and in particular there should be no correlation between consecutive errors no matter how the rows are sorted, as long as it is on some criterion that does not involve the dependent variable. If this is not true, it could be due to a violation of the linearity assumption or due to bias that is explainable by omitted variables (Nau, 2014).

3.3.5 Non-Multicollinearity

Multicollinearity occurs when the independent variables are correlated. If collinearity exists, the variance of the parameter estimates are all inflated. Variance inflation factor (VIF) is a common measure for detecting multicollinearity. Another is Tolerance (T) which measures the influence of one independent variable on all other independent variables. Tolerance is obtained as $T = 1 - R^2$ where R^2 is the coefficient of determination. With VIF > 10 or T < 0.1, there is an indication of multicollinearity (Hair, et.al, 2010).

3.3.6 Obtaining Robust Variance Estimates

Estimates of variance refer to estimated standard errors or, more completely, the estimated variance–covariance matrix of the estimators of which the standard errors are a subset, being the square root of the diagonal elements. Call this matrix the variance. All estimation commands produce an estimate of variance and, using that, produce confidence intervals and significance tests. In addition to the conventional estimator of variance, another estimator has been called by various names because it has been derived independently in different ways by different authors. Two popular names associated with the calculation are Huber and White, but it is also known as the sandwich estimator of variance (because of how the calculation formula physically appears) and the robust estimator of variance (because of claims made about it). Also, this estimator has an independent and long tradition in the survey literature.

The robust estimate of variance is

$$\hat{v} = \hat{V}\left(\sum_{j=1}^{N} u_j' u_j\right)\hat{V}$$

where $\hat{V} = \left(-\frac{\partial^2 lnL}{\partial \beta^2}\right)^{-1}$ is the conventional estimator of variance. Consider likelihood functions that are additive in the observations

$$lnL = \sum_{j=1}^{N} lnL_j$$

then $u_j = \partial lnL_j / \partial \beta$. In general, function L_j is a function of x_j and ß, $L_j(\beta; x_j)$. For many likelihood functions, however, it is only the linear form x_j ß that enters the function. In those cases,

$$\frac{\partial lnL_j(x_j\beta)}{\partial \beta} = \frac{\partial lnL_j(x_j\beta)}{\partial (x_j\beta)} \frac{\partial (x_j\beta)}{\partial \beta} = \frac{\partial lnL_j(x_j\beta)}{\partial (x_j\beta)} x_j$$

by writing $u_j = \frac{\partial lnL_j(x_j\beta)}{\partial (x_j\beta)}$, this becomes simply $u_j x_j$. Thus, the formula for the robust estimate of variance can be rewritten as

$$\hat{v} = \hat{V}\left(\sum_{j=1}^{N} u_j^2 x_j' x_j\right)\hat{V}$$

We refer to u_j as the equation-level score (in the singular), and it is u_j that is returned when you use predict with the score option. u_j is like a residual in that

1. $\sum_j u_j = 0$ and
2. correction of u_j and x_j, calculated over j = 1,…,N, is 0.

In fact, for linear regression, u_j is the residual, normalized,

$$\frac{\partial lnL_j}{\partial(x_j\beta)} = \frac{\partial}{\partial(x_j\beta)} lnf\left\{\frac{(y_i - x_j\beta)}{\sigma}\right\}$$
$$= \frac{(y_i - x_j\beta)}{\sigma}$$

where $f()$is the standard normal density.

Hence, the name the robust estimate of variance is associated with authors, Huber (1967), and White (1980, 1982) (who developed it independently), although many others have extended its development, including Gail, Tan, and Piantadosi (1988), Kent (1982), Royall (1986), and Lin and Wei (1989). In the survey literature, this same estimator has been developed, Kish and Frankel (1974), Fuller (1975), and Binder (1983).

3.4. Logistic Regression

Logistic regression is a specialized form of regression formulated to predict and explain a binary categorical variable or an ordinal variable rather than a metric dependent measure. It is also known as the logit analysis. The model yields regression-like coefficients that indicate the relative impact of each predictor variable. Logistic regression is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model. Logistic regression is applicable to a broader range of research situations than discriminant analysis.

The logistic regression model

The relationship between the dependent variable and the independent variables is given by the equation:

$$\mathrm{Logit(Y)} = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3+\dots+}\, \beta_k X_k$$

where $Y = 1\ for\ poor\ and\ Y = 0\ for\ non-poor$

$\beta_o, \beta_1, \dots, \beta_k$ -regression coefficient parameters

$X_1, X_2, \dots, X_k$ -independent variables

Then, the logistic regression model equation is transformed as:

$$P(Y=1) = \frac{\exp[Logit(Y)]}{1 + exp[Logit(Y)]}$$

where can be used to predict the probability of being classified as poor.

Logistic regression functions are used for describing the nature of the relation between mean response and one (or more) prediction purposes; there is a need to estimate the parameters of the logistic response function. The method of maximum likelihood is used to estimate these parameters. This method is well suited to deal with the problem associated with the response of being binary, when the logistic regression model contains only qualitative variables; it is often referred to as a log linear model (Neter, 1990).

Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable. In this way, logistic regression estimates the probability of a certain event to occur. Thus, it calculates changes in the log odd of the dependent, not changes in the dependent itself. (Rencher, 1985).

Johnson, 2005 states that the difference between logistic regression and discriminant analysis is that logistic regression does not make any assumptions regarding the distribution of the independent variables. Therefore, it is preferred than discriminant analysis when the independent variables are a combination of categorical and continuous because in such cases the multivariate normality assumption is clearly violated. On the other hand, when the multivariate normality is not violated, then discriminant analysis is preferred because is computationally more efficient than logistic regression analysis.

3.4.1. Psuedo R^2

A value of overall model fit that can be calculated for logistic regression, comparable to the R^2 measure used in multiple regression.

The coefficient of determination denoted R^2 and pronounced R squared, indicates how well data points fit a line or curve. It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypothesis, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model.

In logistic regression, there is no true R^2 value. However, because deviance can be thought of as a measure of how poorly the model fits (i.e., lack of fit between observed and predicted values), an analogy can be made to sum of squares residual in ordinary least squares. The proportion of unaccounted for variance that is reduced by adding variables to the model is the same as the proportion of variance accounted for, or R^2

$$R^2{}_{LOGIT} = \frac{-2LL_{null} - (-2LL_k)}{-2LL_{null}}$$

where the null model is the logistic model with just the constant and the k model contains all the predictors in the model (Hosmer & Lemeshow, 2000).

3.4.2 Classification Functions

Classification of observations is done by first estimating the probabilities. These probabilities can be used to classify observations into two groups. Classification of observations in two groups will be based on a cut off value for predicted probabilities, which is usually assumed to be 0.5. All the observations whose predicted probabilities are greater than or equal to 0.5 will be classified as "poor," and those whose values are less than 0.5 will be classified as "non-poor".

3.5 Selection of Method as Proxy Means Test

The researcher used a combination in backward stepwise methods for OLS (robust standard error) and logistic regressions to identify significant indicator that can determine the economic status of the first year students of PUP, main campus for SY 2014-2015. If the least-significant term is not significant remove it and re-estimate, otherwise, stop.

To determine which statistical method is better, the researcher used simulation process using 50 simulation datasets. Each dataset consists of 354 students sample obtain randomly from population of 4,431 students. The researcher performed OLS (robust standard error) and logistic regression for each 50 dataset. Correct classification rates for both OLS (robust standard error) and logistic regression for each 50 simulation datasets were obtained. The higher the correct classification rate, the better the regression.

4. Results and Discussion

This chapter presents the results of summary of data and hypothesis testing. The discussion covers the following: 1)

Identify the non-income variables that can predict the economic status, whether poor or non-poor, of the first year students using logistic regression and ordinary least squares (OLS) regression; 2) Determine which of the two regression methods, logistic regression and OLS regression, is better as PMT in targeting poor PUP first year students SY 2014-2015 3) Predicting the economic status of PUP first year student for the school year 2014-2015 using best model.

4.1.1. Test for Multicollinearity

Before performing OLS (robust standard error) and logistic regression, multicollinearity between predictors was cross-examined. Multicollinearity among predictors can lead to biased estimates and inflated standard errors. Thus, multicollinearity was checked using the regress and vif command in STATA. The table below shows variance inflation factor (VIF) for each predictor.

The VIF is the number of times that variance of corresponding parameter was increased due to multicollinearity as compared to as it would be if there is no multicollinearity. Values of VIF exceeding 10 are often regarding as indicating multicollinearity

Table 1 shows the result of the analysis for multicollinearity test. Since the computed VIF has greater than 10 specifically in estimated family expense, it means that there is influence of one independent variable on other independent variables. Among these independent variables, the researcher will run the stepwise backward method to eliminate other variables that are not significant in predicting the economic status of the students in PUP, main campus for SY 2014-2015.

Table 1. First Run for Multicollinearity Test in Data Set 1

Variable	VIF	1/VIF
aircon	2.6	0.4
amfmradio	1.5	0.7
audiosystem	1.7	0.6
bike	2.2	0.4
business	1.8	0.6
cabletvconn	1.9	0.5
carvan	2.4	0.4
cellphone	1.9	0.5
coffeemaker	1.6	0.6
commissions	1.6	0.6
commprop	1.6	0.6
desktoppc	1.7	0.6
digitalcam	1.9	0.5
estimated family expense (7000 and below)	33.7	0.0
Variable	VIF	1/VIF
estimated family expense (7001 - 14000)	24.2	0.0
estimated family expense (14001 - 21000)	27.3	0.0
estimated family expense (21001 - 28000)	4.4	0.2
estimated family expense (28001 - 35000)	8.8	0.1
estimated family expense (35001 - 42000)	3.7	0.3
estimated family expense (49000 plus)	2.3	0.4
earninginvestment	1.3	0.8
elecfan	2.0	0.5
elecwaterdispenser	1.8	0.6
elecwaterpump	1.5	0.7
employedsiblings	2.8	0.4
famowncar	1.6	0.6
farm	1.5	0.7
farmland	1.4	0.7
ferryboat	1.4	0.7
flatiron	1.5	0.7
fsize	4.0	0.3
houselot	2.0	0.5
internet	2.7	0.4
isfatheralive	1.4	0.7
jeep	1.2	0.8
laptop	2.1	0.5
lrtmrt	1.4	0.7
microwave	1.8	0.6
motorboat	1.3	0.8
motorcycle	1.6	0.6
mp3mp4ipod	2.0	0.5
oventoaster	1.9	0.5
owncar	2.5	0.4
ownertype	1.4	0.7

ownmotorcycle	1.4	0.7
passengerjeep	1.4	0.7
pension	1.3	0.7
pianoelectrickeyboard	1.6	0.6
playstationwiixbox	2.4	0.4
profession	1.3	0.8
realestate	1.3	0.8
refrigerator	1.9	0.5
remitancesabroad	1.7	0.6
reshouseonly	1.5	0.7
ricecooker	1.6	0.6
Variable	VIF	1/VIF
stove	1.6	0.6
studyingsiblings	3.0	0.3
tellandline	1.9	0.5
tricycle	1.2	0.8
tricycleped	1.4	0.7
trolleyrai~d	1.2	0.8
turbobroiler	1.5	0.7
tvset	2.0	0.5
vacuum	1.6	0.6
vcddvd	1.7	0.6
walk	1.3	0.8
washingmachine	1.9	0.5
waterheater	1.5	0.7
workwhilestudy	1.3	0.8
workwhilestudyyes	1.2	0.8
Mean VIF	3.0	

Table 2 shows the result of the analysis for multicollinearity test after the stepwise backward method conducted. Since the computed VIF is less than 10, it means that there is no influence of one independent variable on all other independent variables, or no multicollinearity.

Table 2. Second Run for Multicollinearity Test in Data Set 1

Variable	VIF	1/VIF
aircon	2.5	0.4
amfmradio	1.4	0.7
audiosystem	1.7	0.6
bike	2.2	0.5
business	1.8	0.6
cabletvconn	1.8	0.5
carvan	2.3	0.4
cellphone	1.9	0.5

Variable	VIF	1/VIF
coffeemaker	1.6	0.6
commissions	1.5	0.7
commprop	1.5	0.7
desktoppc	1.7	0.6
digitalcam	1.8	0.6
earninginvestment	1.3	0.8
elecfan	1.9	0.5
elecwaterdispenser	1.7	0.6
elecwaterpump	1.5	0.7
employedsiblings	2.7	0.4
famowncar	1.6	0.6
farm	1.5	0.7
farmland	1.4	0.7
ferryboat	1.3	0.8
flatiron	1.5	0.7
fsize	3.9	0.3
houselot	2.0	0.5
internet	2.6	0.4
isfatheralive	1.4	0.7
jeep	1.2	0.8
laptop	2.0	0.5
lrtmrt	1.4	0.7
microwave	1.7	0.6
motorboat	1.3	0.8
motorcycle	1.5	0.7
mp3mp4ipod	1.9	0.5
oventoaster	1.8	0.6
owncar	2.3	0.4
ownertype	1.4	0.7
ownmotorcycle	1.3	0.7
passengerjeep	1.4	0.7
pension	1.3	0.8
pianoelectrickeyboard	1.5	0.7
playstationwiixbox	2.3	0.4
profession	1.2	0.8
realestate	1.3	0.8
refrigerator	1.8	0.5
remitancesabroad	1.5	0.7
reshouseonly	1.4	0.7
reslotonly	1.3	0.8
ricecooker	1.6	0.6
salarieswages	1.9	0.5
stove	1.6	0.6
studyingsiblings	2.9	0.3
tellandline	1.9	0.5

tricycle	1.2	0.8
tricycleped	1.3	0.8
trolleyrailroad	1.1	0.9
turbobroiler	1.4	0.7
tvset	1.9	0.5
vacuum	1.6	0.6
Variable	VIF	1/VIF
vcddvd	1.7	0.6
walk	1.3	0.8
washingmachine	1.8	0.6
waterheater	1.5	0.7
workwhilestudy	1.3	0.8
workwhilestudyyes	1.2	0.8
Mean VIF	1.7	

4.1.2. Ordinary Least Squares (robust standard error) Regression

Table 3 shows the result from OLS (robust standard error) regression. This includes the coefficient, standard error t-statistic and p-value. These were the following independent variables that has a significant linear association to the dependent variable, which is per capita income of the family.

In addition, this table shows the effect or impact of the non-income variables to predict the dependent variable, which is per capita income. For instance, the bike or daily mode of transportation (bike) has positive beta coefficients. This means that for every unit change in daily mode of transportation (bike), the dependent variable per capita income predicts to increase by 7904.6 “Ceteris Paribus (all other things being equal)”. The regression coefficient in carvan or family owned car/van is 15316.4. This means that for every unit change in family owned car/van, the dependent variable per capita income predicts to increase by 15316.4 “Ceteris Paribus (all other things being equal)”. The regression coefficient in famowncar or daily mode of transportation (family-owned car) is 45448.4. This means that for every unit change in daily mode of transportation (family-owned car), the dependent variable per capita income predicts to increase by 45448.4 “Ceteris Paribus (all other things being equal)". The regression coefficient in ferryboat or daily mode of transportation (ferry boat) is 24250.2. This means that for every unit change in daily mode of transportation (ferry boat), the dependent variable per capita income predicts to increase by 24250.2 "Ceteris Paribus (all other things being equal)". The regression coefficient in fsize or family size is -4018.8. This means that for every unit change in family size, the dependent variable per capita income predicts to decrease by -4018.8 "Ceteris Paribus (all other things being equal)". The regression coefficient in internet or family owned internet connection is 10099.3. This means that for every unit change in family owned internet connection, the dependent variable per capita income predicts to increase by 10099.3 "Ceteris Paribus (all other things being equal)". The regression coefficient in own car or daily mode of transportation (own car) is 76392.6. This means that for every unit change in daily mode of transportation (own car), the dependent variable per capita income predicts to increase by 76392.6 "Ceteris Paribus (all other things being equal)". The regression coefficient in playstationwiixbox or family owned playstation/wii/xbox is 19164.1. This means that for every unit change in family owned playstation/wii/xbox, the dependent variable per capita income predicts to increase by 19164.1 "Ceteris Paribus (all other things being equal)". The regression coefficient in remittances abroad or source of income (remittances from abroad) is 10848.8. This means that for every unit change in source of income (remittances from abroad), the dependent variable per capita income predicts to increase by 10848.8 "Ceteris Paribus (all other things being equal)". The regression coefficient in rest house only or family owned residential house only is -6580.8. This means that for every unit change in family owned residential house only, the dependent variable per capita income predicts to decrease by -6580.8 "Ceteris Paribus (all other things being equal)". The regression coefficient in salaries / wages or source of income (salaries/wages) is 7496.3. This means that for every unit change in source of income (salaries/wages), the dependent variable per capita income predicts to increase by 7496.3 "Ceteris Paribus (all other things being equal)". The regression coefficient in washing machine or family owned washing machine is 9177.5. This means that for every unit change in family owned washing machine, the dependent variable per capita income predicts to increase by 9177.5 "Ceteris Paribus (all other things being equal)".

Table 3. Estimates of the Parameters of the Final OLS Regression Model (Robust Standard Error)

Linear regression		Number of obs = 354		
		F(9, 341) = .		
		Prob > F = .		
		R-squared = 0.4561		
		Root MSE = 20874		
pcapita_OLS	Coef.	Robust Std. Err.	t	P>t
_cons	39673.6	5296.6	7.5	0.000
bike	7904.6	2967.2	2.7	0.008
carvan	15316.4	5197.6	3.0	0.003
famowncar	45448.4	6088.9	7.5	0.000
ferryboat	24250.2	6980.4	3.5	0.001
fsize	-4018.8	807.7	-5.0	0.000
internet	10099.3	2849.8	3.5	0.000
owncar	76392.6	5490.9	13.9	0.000
playstationwiixbox	19164.1	6263.3	3.1	0.002
remitancesabroad	10848.8	3942.0	2.8	0.006
reshouseonly	-6580.8	2362.7	-2.8	0.006
salarieswages	7496.3	2428.3	3.1	0.002
washingmachine	9177.5	2428.7	3.8	0.000

where: _cons = Constant; carvan = Family owned Car/Van; famowncar = Daily Mode of Transportation (family-owned car); ferryboat = Daily Mode of Transportation (ferry boat); fsize = Family Size; internet = Family owned internet connection; owncar = Daily Mode of Transportation (own car); playstationwiixbox = Family owned playstation/wii/xbox; remitancesabroad = Source of Income (remitances from abroad); reshouseonly = Family owned Residential house only; salarieswages = Source of Income (salaries/wages); washingmachine = Family owned washing machine

The OLS (robust standard error) Regression Model

The relationship between the dependent variable and the independent variables is given by the equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12}$$

Hence,

$$Y = 39673.6 + 7904.6\text{bike} + 15316.4 carvan + 45448.4 famowncar + 24250.2 ferryboat - 4018.8 fsize + 10099.3 internet + 76392.6\text{ owncar} + 19164.1\text{ playstationwiixbox} + 10848.8\text{ remitancesabroad} - 6580.8\text{ reshouseonly} + 7496.3\text{ salarieswages} + 9177.5\text{ washingmachine}$$

Overall Model Evaluation

The most commonly used measure of predictive accuracy for the regression model is the coefficient of determination (R^2). Calculated as the squared correlation between the actual and predicted value of the dependent variable, it represents the combined effects of the entire variate (one or more independent variables plus the intercept) in predicting the dependent variable. It ranges from 0.0 (no prediction) to 1.0 (perfect prediction). The computed R^2=0.456, which means that 45.6% of the variation in dependent variable, which is the per capita income, explains by the independent variables/non-income variables.

4.1.3 Logistic Regression

Table 4 shows the result of the logistic regression. This includes the coefficient, standard error, t-statistic and p-value. These are the following independent variables that have a significant linear association to the dependent variable, which is economic status of the student.

Moreover, this table shows the effect or impact of the non-income variables to predict the dependent variable, which is economic status. For instance, the internet or family owned internet connection has a negative beta coefficient. This means that for every unit change in the family owned internet connection, the dependent variable economic status-logged odds predicts to decrease by 1.2, "Ceteris Paribus" (all other things being equal). For isfatheralive or father (alive/deceased), the regression coefficient is -1.0. This means that for every unit change in the father (alive/deceased), the dependent variable economic status-logged odds predicts to decrease by 1.0 "Ceteris Paribus" (all other things being equal). For remittances abroad or source of income (remittances from abroad), the regression coefficient is -1.8. This means that for every unit change in the source of income (remittances from abroad), the dependent variable economic status-logged odds was predicted to decrease by 1.8 "Ceteris Paribus" (all other things being equal). For ricecooker variable or family owned rice cooker, the regression coefficient is -1.0. This means that for every unit change in the family owned rice cooker, the dependent variable economic status-logged odds predicts to decrease by 1.0 "Ceteris Paribus" (all other things being equal). For salaries wages or source of income (salaries/wages), the regression coefficient is -1.1. This means that for every unit change in the source of income (salaries/wages), the dependent variable economic status-logged odds predicts to decrease by 1.1 "Ceteris Paribus" (all other things being equal). For vcddvd or family owned vcd/dvd, the regression coefficient is -0.6. This means that for every unit change in the family owned vcd/dvd, the dependent variable economic status-logged odds predicts to decrease by -0.6 "Ceteris Paribus" (all other things being equal). For washing machine or family owned washing machine, the regression coefficient is -0.7. This means that for every unit change in the family owned washing machine, the dependent variable economic status-logged odds predicts to decrease by 0.7 "Ceteris Paribus" (all other things being equal). For waterheater variable or family owned water heater, the regression coefficient is 0.9. This means that for every unit change in the family owned water heater, the dependent variable economic status-logged odds predicts to increase by 0.9 "Ceteris Paribus" (all other things being equal).

Table 4. Estimates of the Parameters of the Final Logistic Regression Model

Logistic regression		Number of obs = 338		
		LR chi2(8) = 89.53		
		Prob > chi2 = 0.0000		
Log likelihood = -181.79101		Pseudo R2 = 0.1976		
poor_logit	Coef.	Std. Err.	z	P>z
_cons	3.1	0.6	5.0	0.000
internet	-1.2	0.3	-3.7	0.000
isfatheralive	-1.0	0.5	-2.2	0.028
remitancesabroad	-1.8	0.6	-3.0	0.002
ricecooker	-1.0	0.3	-3.6	0.000
salarieswages	-1.1	0.3	-3.6	0.000
vcddvd	-0.6	0.2	-2.4	0.016
washingmachine	-0.7	0.3	-2.6	0.008
waterheater	0.9	0.4	2.1	0.037

where : internet = Family owned internet connection; isfatheralive = Father (alive/deceased); remitancesabroad = Source of Income (remitances from abroad); ricecooker = Family owned rice cooker; salarieswages = Source of Income (salaries/wages); vcddvd = Family owned vcd/dvd; washingmachine = Family owned washing machine; waterheater = Family owned water heater

The Logistic Regression Model

The relationship between the dependent variable and the independent variables is given by the equation:

$$\text{Logit(Y)} = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3+}\ \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8$$

Hence,

$$\text{Logit (Y)} = 3.1\ - 1.2\text{internet}\ - 1.0\text{isfatheralive} - 1.8\text{remitancesabroad} - 1.0\text{ricecooker} - 1.1\text{salarieswages}\ - 0.6\text{vcddvd}\ - 0.7\text{washingmachine} + 0.9\text{waterheater}$$

The logistic regression model equation can be transformed into:

$$P(Y=1) = \frac{\exp[\text{Logit}(Y)]}{1+\exp[\text{Logit}(Y)]}$$

$$= \frac{\exp\left[\begin{array}{c}3.1\ -1.2\text{internet}\ -1.0\text{isfatheralive} - 1.8\text{remitancesabroad} - \\ 1.0\text{ricecooker} - 1.1\ \text{salarieswages}\ -0.6\ \text{vcddvd}\ -0.7\ \text{washingmachine} + \\ 0.9\ \text{waterheater}\end{array}\right]}{1+\ \exp\left[\begin{array}{c}3.1\ -1.2\text{internet}\ -1.0\text{isfatheralive} - 1.8\text{remitancesabroad} - \\ 1.0\text{ricecooker} - 1.1\ \text{salarieswages}\ -0.6\ \text{vcddvd}\ -0.7\ \text{washingmachine} + \\ 0.9\ \text{waterheater}\end{array}\right]}$$

4.2 Correct Classification Rate

Table 5 shows the summary result of simulation process from 50 dataset using OLS (robust standard error) and logistic regression. On the average, 96.8% of the total correct specification for both poor and non-poor using OLS (robust standard error), while 73.5% for logistic regression. These results indicate that there is a sufficient evidence that the OLS (robust standard error) is better than logistic regression.

Economic Status	**OLS (robust standard error) regression**			**Logistic regression**		
Method	**Non-Poor**	**Poor**	**Total**	**Non-Poor**	**Poor**	**Total**
Data Set No.1	98.3	0.0	97.7	82.4	51.9	70.9
Data Set No.2	95.1	75.0	94.9	72.7	65.2	69.8
Data Set No.3	98.0	60.0	97.5	89.3	52.1	76.6
Data Set No.4	97.4	40.0	96.6	75.9	68.3	73.2
Data Set No.5	98.0	70.0	97.2	74.5	67.4	71.8
Data Set No.6	98.3	50.0	97.2	78.6	64.8	73.7
Data Set No.7	98.3	0.0	97.2	86.5	56.4	76.6
Data Set No.8	100.0	50.0	99.2	82.0	53.8	71.5
Data Set No.9	98.3	100.0	98.3	78.0	51.9	68.4
Data Set No.10	96.9	33.3	96.3	78.6	64.2	73.2
Data Set No.11	100.0	42.9	98.9	87.0	40.3	70.6
Data Set No.12	97.1	0.0	95.8	87.5	50.0	76.3
Data Set No.13	95.4	66.7	94.9	80.6	56.1	71.5
Data Set No.14	96.8	50.0	95.8	76.2	55.7	68.6
Data Set No.15	97.7	83.3	97.5	87.7	55.6	76.3
Data Set No.16	98.6	0.0	97.7	84.7	52.9	74.0
Data Set No.17	99.7	50.0	98.6	82.7	55.0	72.6
Data Set No.18	96.6	100.0	96.6	87.1	59.8	77.7
Data Set No.19	96.3	60.0	95.8	84.1	62.2	76.3
Data Set No.20	96.0	66.7	95.5	88.4	42.1	72.6
Data Set No.21	94.8	40.0	94.1	83.6	66.4	78.0

Data Set No.22	97.1	0.0	96.0	85.1	44.4	70.6
Data Set No.23	96.0	50.0	95.2	88.6	39.8	72.3
Data Set No.24	98.0	83.3	97.7	83.5	54.7	74.0
Data Set No.25	99.4	50.0	98.9	77.9	62.6	72.6
Data Set No.26	97.4	57.1	96.6	84.3	55.0	73.4
Data Set No.27	98.3	50.0	97.5	83.8	50.0	71.8
Data Set No.28	95.4	100.0	95.5	82.6	59.7	74.6
Data Set No.29	98.0	100.0	98.0	86.8	61.1	77.7
Data Set No.30	99.7	66.7	99.4	86.9	51.5	76.6
Data Set No.31	98.8	69.2	97.7	83.4	51.3	72.6
Data Set No.32	99.4	90.0	99.2	84.0	52.0	72.9
Data Set No.33	96.8	80.0	96.6	87.0	44.4	74.0
Economic Status	**OLS (robust standard error) regression**			**Logistic regression**		
Method	**Non-Poor**	**Poor**	**Total**	**Non-Poor**	**Poor**	**Total**
Data Set No.34	96.8	60.0	96.3	91.2	37.5	71.8
Data Set No.35	97.4	60.0	96.9	84.9	52.6	74.3
Data Set No.36	98.6	66.7	98.3	84.5	53.1	73.2
Data Set No.37	98.6	25.0	97.7	80.0	62.6	73.2
Data Set No.38	97.7	37.5	96.3	88.8	45.6	74.9
Data Set No.39	94.8	83.3	94.6	88.3	44.7	74.3
Data Set No.40	96.0	33.3	95.5	85.7	53.2	75.7
Data Set No.41	98.6	33.3	98.0	84.5	50.8	72.3
Data Set No.42	95.4	71.4	94.9	80.7	54.2	70.9
Data Set No.43	96.0	87.5	95.8	85.1	53.2	72.6
Data Set No.44	97.7	20.0	96.6	83.8	50.9	73.2
Data Set No.45	93.2	100.0	93.2	88.1	51.7	76.0
Data Set No.46	98.3	75.0	98.0	88.4	51.6	75.7
Data Set No.47	97.4	66.7	96.9	79.9	58.6	71.5
Data Set No.48	95.7	40.0	94.9	85.6	51.7	74.3
Data Set No.49	99.4	0.0	97.2	85.8	41.1	72.3
Data Set No.50	96.3	40.0	95.5	81.5	59.0	73.7
Average	**97.4**	**54.7**	**96.8**	**83.8**	**53.8**	**73.5**

4.3 Predicted Economics Status of PUP First Year Students Using Final Model

Table 6 shows the predicted economic status using the OLS (robust standard error) of the First Year Students in PUP, main campus SY 2014-2015. Majority of the first year students fall under the non-poor category, which is 97.7%, while 2.3% falls to poor category.

Table 6. Predicted Economic Status of First Year Students in PUP, Main Campus SY 2014-2015 using Final OLS (robust standard error)

Economic Status	Frequency	Percentage
Non-Poor	4,330	97.7
Poor	101	2.3
Total	4,431	100.0

5. Conclusion

After performing 50 simulation process for both OLS (robust standard error) and logistic regression, the researcher hereby concludes that the OLS (robust standard error) regression model is better than logistic regression as a PMT in identifying whether a student is poor or non-poor, based on the non-income variable of poverty.

Majority of the first year students' fall under the non-poor category using the OLS (robust standard error) method.

6. Recommendation

Based on the foregoing findings of the study and derived conclusions, the following recommendations are formulated:
1. The government should adopt and compare the results of PMT in identifying who and where the poor are in the country. This process may be able to help ensure the poor really receive or enjoy the benefits from different poverty alleviation programs.
2. Identifying the poor is a very crucial part for the proper implementation of programs for poverty reduction. The researcher does not directly suggest to use this model in identifying the poor students of the University but it can be use as a baseline or pattern in identifying or targeting poor students. It is recommended that other statistical models be considered and compared with what other frameworks exist

today as to which is best in terms of identifying the economic status of the students.

3. The researcher suggests that improving the quality of data in gathering information about the profile of the students, and enhancing some questions in the survey could further enrich the credibility of the said study. This process makes the results more valid and accurately identify or target the poor students in the University.

4. Further research on the same topic in other locales should be conducted in order to verify, amplify, or negate the findings of the study.

5. The researcher used the uniform random selection of income in the income category P7,001 - P14,000, where the poverty threshold falls as of 2012. It is advice to use another selection method, which is exponential distribution in selecting random numbers for logistic regression to determine if the results are the same.

REFERENCES

Books

Cochran, W. G. (1963). *Sampling Techniques, 2nd Ed.*, New York: John Wiley and Sons, Inc

Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". J Roy Stat Soc B 20: 215–242.

Draper, N.R.; Smith, H. (1998) *Applied Regression Analysis*, Wiley. ISBN 0-471

17082-8 (Chapter14) "Interpreting the Coefficients on Dummy Variables".

Garavaglia, S., & Sharma, A. (n.d.) (1996). *A Smart Guide to Dummy Variables: Four Applications and a Macro*. Dun & Bradstreet Murray Hill, New Jersey 07974

Freedman, D. A. (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 128.

Gujarati, D. N. (2003). *Basic econometrics.* McGraw Hill. p. 1002. ISBN 007-233542-4

Hair, J.F. et.al, (2010). *Multivariate Data Analysis. 7th Edition*. Pearson Education Inc. ISBN 978-981-0695-99-6

Huber, P. J. (1967). The behavior of maximum likelihood estimates under Nonstandard conditions. In Vol. 1 of Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 221–233. Berkeley: University of California Press.

Walker, SH and Duncan, DB (1967). "Estimation of the probability of an event as a function of several independent variables". Biometrika 54: 167–178.

Journals

Albis, M. L. F. (2013). Assistant Professor, School of Statistics, University of the Philippines Diliman and Consultant, DSWD.

Bellenas (2014). Masters in Applied Statistics, Polytechnic University of the Philippines

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. International Statistical Review 51: 279–292.

Bradshaw, T.K. (2006). Theories of Poverty and Anti-Poverty Programs in Community Development. (http://www.rupri.org/Forms/WP06-05.pdf)

Coady, Grosh and Hoddinott (2009) found that cash transfer programs most Commonly rely on some form of household (or individual) assessment, such as means testing or proxy means testing. In many of the country cases, however, household eligibility decisions (via means testing or proxy means testing) are combined with other methods of targeting, such as geographic pre-selection of priority areas

Collier, P. and Dollar, D. (2002): Aid allocation and poverty reduction. In: European Economic Review 46 (2002) (5-6): 1475-1500.

De la Briere, B. and Lindert K. (2003). Country Study on Brazil: Brazil's Cadastro Único. The World Bank. This case study is being updated based on recent collaborative work by the Ministry of Social Development (MDS) in Brazil and the World Bank (including the authors of the case study) in designing a World Bank Project to support the Bolsa Familia Program. See The World Bank (May 25, 2004). "Project Appraisal Document: Bolsa Familia Project." Report No. 28544-BR

Dershem, L. (2013). Design and Monitoring & Evaluation Advisor, Save the Children US, and technical advisor to the BOTA Foundation's Department of Monitoring & Evaluation.

Kish, L., and Frankel M. R.(1974). Inference from complex samples. Journal of the Royal Statistical Society, Series B 36: 1–37.

Gail, M. H., Tan W. Y., and Piantadosi, S..(1988). Tests for no treatment effect in randomized clinical trials. Biometrika 75: 57–64.

Haddad, l.j., Kennedy, E. and Sullivan, J. (1994): Choice of Indicators for Food Security and Nutrition. In: Food Policy 19 (3): 329-343

Johannsen, J. (2006): Operational poverty targeting in Peru – Proxy meansTesting with non-income indicators. Working Paper No. 30. In: http://www.ipc-undp.org/pub/IPCWorking Paper30.pdf. International Poverty Centre, United Nations Development Programme, Brazilia, Brazil

Johannsen, J. (2009): Operational assessment of absolute poverty by proxy Mean tests – the example of Peru. Lang, Frankfurt am Main

Kent, J. T. 1982. Robust properties of likelihood ratio tests. Biometrika 69: 19–27.

Kish, L. and Frankel, M. R. (1974). Inference from complex samples. Journal of the Royal Statistical Society, Series B 36: 1–37.

Larrañaga, O. (2003). Case Study on the Chile: Focalización de Programas Sociales en Chile: El Sistema CAS. The World Bank.

Lindert, K. (2003). Country Study on the US .Implementing Means-Tested Welfare Systems in the US. The World Bank.

Lin, D. Y., and Wei L. J. (1989). The robust inference for the Cox proportional Hazards model. Journal of the American Statistical Association 84: 1074–1078.

Orozco, M. and Hubert C. (2003). Case Study on the Mexico: La Focalización en el Programa de Desarrollo Humano Oportunidates en Mexico. The World Bank

Mapa, Dennis S. 2013:Associate Professor and Director for Research, School of Statistics and Affiliate Associate Professor, School of Economics, University of the Philippines, Diliman and Consultant, DSWD.

Neuman, W. Lawrence. 2007. Basics of Social Science Research: Qualitative and Quantitative Approaches (2nd Edition). Pearson Education, Inc. U.S.A.

Ravallion, M. (1992): Poverty comparison – a guide through methods and concepts. LSMS Working Paper 88. The World Bank. Washington, D.C.

Tarsicio, C. and Fernandez, L. (2003). Case Study on the Colombia: Targeting Social Spending to the Poor with Proxy Means Testing: Colombia's SISBEN System. The World Bank.

Viquez, Roxana (2003). Case Study on the Costa Rica: Sistema de Identificación de la Población Objetivo: SIPO en Costa Rica. The World Bank

Zeller, M., M. Sharma, C. Henry and C. Lapenu (2006): An Operational Method for Assessing the Poverty Outreach of Development Policies and Projects: Results of Case Studies in Africa, Asia, and Latin America. In: World Development 34 (3): 446-464

Internet Sources

Ahmed, F. (2004): Practices of Poverty Measurement and Poverty Profile of Bangladesh. ERD Working Paper Series No. 54. Asian Development Bank, Manila. Retrieved from http://www.adb.org/B5B8C407-30B5-4213-869E-E38834B0B162/FinalDownload/DownloadId-91E5D1EAEC7522814A005A93AE370473/B5B8C407-30B5-4213-869E-E38834B0B162/sites/default/files/publication/28341/wp054.pdf

Aho, G., S. LariviérE and F. Martin (eds.) (1998): Poverty analysis manual – with applications in Benin. United Development Program, Université Nationale du Bénin, Université Laval, Québec, QC, CA. Retrieved from http://ageconsearch.umn.edu/B5B8C407-30B5-4213-869E-E38834B0B162/FinalDownload/DownloadId-E9EC5E08E6877E8C1030971E4CAD97A5/B5B8C407-30B5-4213-869E-E38834B0B162/bitstream/155485/2/2_van%20Edig.pdf

Albert, J.R. G., 2013. National Statistical Coordination Board, Philippine Statistical Authority, Secretary General. Retrieved from http://opinion.inquirer.net/54331/global-poverty-down-philippine-poverty-remains-high,http://www.nscb.gov.ph/pressreleases/2013/PR-201304-NS1-04_poverty.asp#sthash.1AUojCDk.dpuf

Asian Development Bank, (2009). Retrived from http://www.adb.org/publications/poverty- Philippines causes-constraints-and opportunities

Australian Agency for International Development (AusAID), Canberra, (2011) Retrieve from www.ausaid.gov.au/publications,http://ageconsearch.umn.edu/handle/794: 2007,Quarterly Journal of International Agriculture 52 (2013), No. 1; DLG-Verlag Frankfurt/M

Department of Social Welfare and Development (2010). Retrieve from http://en.wikipedia.org/wiki/Pantawid_Pamilyang_Pilipino_Program, http://nhts.dswd.gov.ph.

Dummy Variable. Retrieve from http://en.wikipedia.org/wiki/Dummy_variable_%28statistics%29

Informs (2015). Operational Research. Retrieve from https://www.informs.org/About-INFORMS/What-is-Operations-Research

Inquirer. Retrieve from http://opinion.inquirer.net/54331/global-poverty-down-philippine-poverty remains-high.

Executive Order No. 867 dated March 9, 2010 entitled "Providing for the Adoption of the National Household Targeting System for Poverty Reduction as a Mechanism for Identifying Poor Households who shall be recipients of Social Protection Programs Nationwide". Retrieve from https://www.google.com.ph/url?sa=t&rct=j&q=&esrc=s&source=web&cd=6&cad=rja&uact=8&ved=0CFsQFjAF&url=http%3A%2F%2Fwww.nscb.gov.ph%2Fncs%2F12thncs%2Fpapers%2FINVITED%2FSpecial%2520Session%2FSpecial%2520Session_%2520New%2520Proxy%2520Means%2520Test%2520(PMT)%2520Models%2520Improving%2520Targeting%2520of%2520the%2520Poor%2520for%2520Social%2520Protection.pdf&ei=k1hHU83CG4j28QWl1oBY&usg=AFQjCNFMTto1C53Y71YVAS4dLmc3TaPgw&sig2=lq8fvIOOwWFmUW69sOWwRw

Executive Issuance (2010). Retrieve from http://webcache.googleusercontent.com/search?q=cache:5RbMt5dUfL0J:www.gov.ph/section/executive-issuances/page/148/+&cd=2&hl=en&ct=clnk&gl=ph

Garavaglia, S., & Sharma, A. (n.d.) (1996). A Smart Guide to Dummy Variables: Four Applications and a Macro. Dun & Bradstreet Murray Hill, New Jersey 07974. Retrieve from http://www.ats.ucla.edu/stat/sas/library/nesug98/p046.pdf

Listahanan under Department of Social Welfare and Development. Retrieve from http://listahanan.dswd.gov.ph/2DA48234-A756-4D1A-9C79-38FB5E55CEDF/FinalDownload/DownloadId-C9A27CDC97E77579D1CEB06CD694DC14/2DA48234-A756-4D1A-9C79-38FB5E55CEDF/download/reports/profile_of_the_poor_/profile%20of%20the%20poor-021213.pdf

Logistic Regression. Retrieve from http://en.wikipedia.org/wiki/Logistic_regression

National Statistical Coordination Board. Retrieve from http://www.nscb.gov.ph/ru12/DEFINE/DEF-INC.HTM

Philippine Basic Education Blogspot (2012). Retrieve from http://philbasiceducation.blogspot.com/2012/09/overcoming-constraints-of-poverty on.html#ixzz3CySjs21l

Philippine Statistical Authority. Retrieve from http://www.bles.dole.gov.ph/dews/Terms.aspx

Polytechnic University of the Philippines. Retrieve from http://www.pup.edu.ph/profile/

Republic Act No. 8425 December 11, 1997. Retrieve from http://www.lawphil.net/statutes/repacts/ra1997/ra_8425_1997.html.

UNICEF. Retrieve from http://www.unicef.org/2DA48234-A756-4D1A-9C7938FB5E55CEDF/FinalDownload/DownloadId-

6B1DAD8B5894D9D0843004930387E3C4/2DA48234-A756-4D1A9C79-38FB5E55CEDF/socialpolicy/files/targeting-poorest.pdf

United Nations (2008): End Poverty and Hunger. Retrieve from http://www.un.org/millenniumgoals/(25.11.2008).

Ordinary Least Squares. Retrieve from http://en.wikipedia.org/wiki/Ordinary_least_squares

Per capita Income. Retrieve from http://en.wikipedia.org/wiki/Per_capita_income

World Bank (2011). Proxy Means Testing. Retrieve from http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTSAFETYNETSANDTRANSFERS/0,,contentMDK:20795790~menuPK:1552914~pagePK:210058~piPK:210062~theSitePK:282761,00.html

World Bank (1995). Proxy means tests for targeting social programs: Simulations and speculation. Retrieve from http://elibrary.worldbank.org/doi/book/10.1596/0-8213-3313-5