

# Binary Logistic Predictive Model in Determining Students' Intention to Take Higher Education

Asia Pacific Journal of  
Multidisciplinary Research  
Vol. 5 No.4, 138-143  
November 2017  
P-ISSN 2350-7756  
E-ISSN 2350-8442  
www.apjmr.com

**Marvin S. Daguplo**

College of Teacher Education, Southern Leyte State University, Philippines  
daguplosdv@yahoo.com

*Date Received: August 15, 2017; Date Revised: November 12, 2017*

**Abstract** –*Students' choice to take higher education is defined by the interaction among behavior and cognition, personal factors, and environmental factors. Using data mining technique and binary regression analysis, this study aims to uncover factors that significantly predict the likelihood for a student to take higher education. Analysis revealed that (i) students whose parents are educated with high income are 1.77 times more likely to pursue higher education than not; and (ii) Older Female Students are less likely to pursue higher education. The model explains that parents should encouraged their young children to embrace education to its highest level by capacitating themselves economically to provide their basic educational and financial needs. Lastly, the study concludes that the pursuit for higher education among young students can be defined primarily by their parents' educational and financial capability.*

**Keywords** –*Higher Education, Access to Education, Finance, Model, Binary Logistic Regression Analysis*

## INTRODUCTION

Intentions to pursue higher education is believed to be affected by various significant factors - personal or environmental – mostly, however, the interplay of both. Koyama [1] found that students' choice regarding college are highly affected by the social, cultural, and political capital of their family. A study also revealed that the cost of education, physical aspects and resources of Higher Education Institutions, and influences from significant people such as parents, teachers, and peers as the top three considerations in pursuing higher education [2].

Generally, literatures provide limited and comprehensive understanding of students' choice and decision to pursue higher education. Literatures revealed that students' decisions are affected by the academic reputation of the institution, assumptions, and motives [3],[4], program quality and size, price/cost, financial aid, geographic location, contact with faculty [5], and a student's individual characteristics such as academic ability and achievement [6]. Such diversity and unfamiliarity of learning domains [7], [5] heightens the need for accounting educators to develop a better understanding of the experiences and issues that students bring to their learning in higher education.

Moreover, numerous studies concluded that one's predisposition toward college is determined by

personal, familial, cultural, and environmental conditions [8]-[10]. These studies, however, are generally qualitative in nature [11]. Thus, this study was conceptualized to present a quantitative view of students' intention to pursue higher education, taking advantage of the increasing data generated from higher education repositories.

The researchers tried to uncover factors that significantly predict the likelihood for a student to take higher education. Thus, this study aims to establish a predictive model using the contextual factors like parents' education and job, student's age, sex and home address type and others, as factors influencing the student's intention to pursue higher education.

This objective was attained through the use of binary logistic regression, simply logistic regression, which predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable –to take or not to take higher education.

## FRAMEWORK, METHOD AND DESIGN OF THE STUDY

This study grounds its assumption on Bandura's Social Cognitive Theory which explains causation involving triadic interaction among behavior and cognition, personal factors, and environmental factors.

All these are assumed to operationally influence choice behaviors [12, 13]. Specifically, self-efficacy, outcome expectations, interests, and goals are thought to influence eventual activity selection and performance. From this perspective, Lent [13] stressed that physical attributes (e.g., sex and race), features of the environment, and learning experiences—influence career related interests and choice behavior.

Within this theoretical framework, the researchers examined variables, using data mining technique, retrieved from the University of California – Irvine Campus Machine Learning Repository Archive contributed by Fabio Pagnotta and Hossain Mohammad Amran from the Department of Computer Science [14]. Variable selection was done through a literature-driven approach to streamline the variables derived from the source. The sample includes 1044 secondary students around the world – the total number provided in the database retrieved at University of California – Irvine Campus Machine Learning Repository Archive.

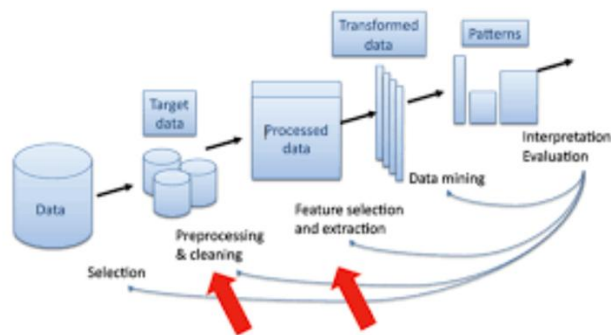


Figure 1. The Data Mining Procedure adapted in this Study

Data mining employed in this study started with the selection of the data that would answer the research question in mind. Cleaning and extraction were made to specific individuals that does not contain complete information on the identified variables for statistical testing. Data exploration were conducted at a greater depth providing more opportunities to see some variables related, and to identify other possible patterns.

Data mining, as a form of exploratory data analysis, is the process of extracting patterns and relationships from data rather than testing pre-formulated hypotheses (Chong, 2016). It is, therefore, through this design that a deterministic model helpful in determining students' intention to take higher education be explored and be established through

binary logistic regression analysis. Inputs of the proposed model are combinations of personal and cognitive variables such as past failure experience in their academic subject, physical attributes (e.g. sex, health status and age), features of the environment or the social aspect (e.g. family relationship ties, family support to education, type of residence, parents' educational attainment and job) which the researchers assumed to be hypothetically factors affecting intention to take higher education.

Thus, this study hypothesizes that students' academic history, physical attributes and social support significantly affect their intention to pursue higher education.

Multivariate Statistical Analysis were performed to extract the necessary information intended for this study. Initial analysis utilized the frequency counts to determine disparity of those who intend to take higher education and not. Principal Component Analysis was performed to determine multicollinearity among variables identified as potential factors affecting students' intent to take higher education. Lastly, the Binary Logistic Regression was performed to determine significant predictors of taking higher education.

## RESULTS

Descriptive analysis starts with knowing the number of students (n=955) from the retrieved sample (n=1044) who desired to take higher education. This shows that huge number of secondary student graduates still has the innate desire to continue higher education.

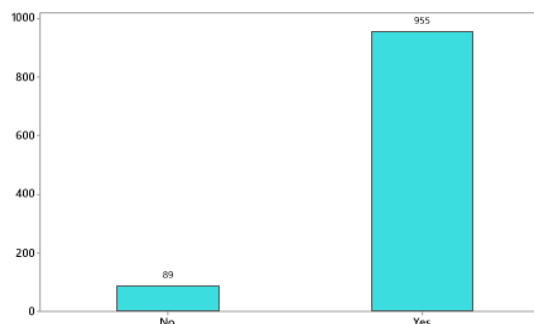


Figure 1. Number of Students who wants to take higher education.

Initially, there were eleven literature-based identified variables that is considered as potential predictors to students' intention to study higher education. Our goal is to identify which of these variables can significantly predict on the student's desire to pursue higher education.

**Table 1. Initial Selected Variables, its details, type of scale and coding**

Variables	Details	Type of Scale and Coding
Age	Age of the respondents	numeric: from 15 – 22 years old
Sex	Gender	binary: Male or Female
Address	student's home address type	binary: 'U' - urban or 'R' - rural
Health	current health status	numeric: from 1 - very bad to 5 - very good
Family size	family size	binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3
Parents' Education	Parents' education	numeric: 0 - none, 1 - primary education (4th grade), 2 5th to 9th grade, 3 secondary education or 4 higher education
Parents' Income	Parents' Income	numeric: from 1 – very low to 5 – very high
failures	number of past class failures	numeric: n if 1 <= n < 3, else 4
Family support	family educational support	binary: yes or no
Family relationship	quality of family relationships	numeric: from 1 - very bad to 5 - excellent
guardian	student's guardian	nominal: 'mother', 'father' or 'other'

To ensure that each variable has a unique contribution to the students' intention to higher education, multicollinearity among variables were assessed using Principal Component Analysis.

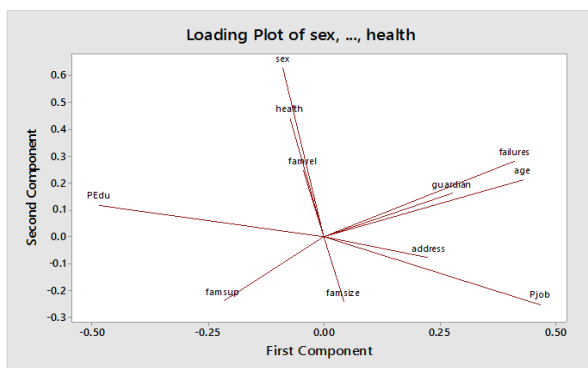


Figure 2. Loading plot of the initial components identified.

The loading plot from the PCA shows that:

- Sex, health status and quality of family relationship are highly collinear;

- Student's guardian, respondent's age, and experience of course failures are likewise correlated; and
- Student's home address type and Parent's income tend to go in almost the same direction.

Multicollinearity test leads to the elimination of some variables that are highly collinear, that is, variables that measures the same concept. Removing redundancy can help us select the best possible binary logistic model by providing the most precise estimate possible [15]. Thus, from the eleven potential predictors, only six predictors remain as input to the desired model, namely, parent's education and income, family size and support, student's sex, and age.

**Table 2. Analysis to Measure Significant Predictors of Students' Intention to Take Higher Education**

Source	DF	Adjusted Deviation	Adjusted Mean	Chi-Square	P-Value
Regression	6	102.183x	17.031	102.180	0.000
Parents Education	1	19.398	19.398	19.400	0.000
Family support	1	0.257	0.257	0.260	0.612
Family size	1	0.000	0.000	0.000	0.995
sex	1	8.228	8.229	8.230	0.004
age	1	44.154	44.154	44.150	0.000
Parents' Income	1	4.572	4.572	4.570	0.032
Error	1037	506.272	0.488		
Total	1043	608.455			

Preliminary Regression analysis revealed that Parents' Education and Income, and Students' Sex and Age are significant predictors of students' intention to take higher education (p-value < .05). Removing two nonsignificant predictors – Family support and Family size – the reduced analysis now remains only four predictors as input of the desired regression model.

**Table 3. Reduced Analysis to identify Significant Predictors of Students' Intention to Take Higher Education**

Source	DF	Adjusted Deviation	Adjusted Mean	Chi-Square	P-Value <sup>a</sup>	Odds Ratio <sup>b</sup>
Regression	4	101.925	25.481	101.930	0.000	
Parents Education	1	20.187	20.187	20.190	0.000	1.777
Sex	1	9.104	9.104	9.100	0.003	0.449
Parents Income	1	4.915	4.915	4.910	0.027	1.761
Age	1	45.344	45.344	45.340	0.000	0.525
Error	1039	506.530	0.488			
Total	1043	608.455				

<sup>a</sup>Input variable with a p-value < 0.05, significant predictor

<sup>b</sup>Odds ratios that are greater than 1 indicate that the event is likely to happen

**Regression Equation**

From table 3, the binary logistic model was derived and is now defined in the Minitab 17 output as:

$$P(1) = \exp(Y') / (1 + \exp(Y'))$$

$$Y' = 13.02 + 0.558 * P\_Education - 0.717 * sex - 0.241 P\_Income - 0.6131 * age$$

To establish the adequacy and acceptability of the model, the goodness-of-fit test was used to determine whether the predicted probabilities deviate from the observed probabilities in a way that the binomial distribution does not predict. The goodness of fit tests for model adequacy shows that the model is fit for the data (p-value > .05), thus, fail to reject the null-hypothesis that the distribution follows a binomial distribution.

**Table 4. Goodness-of-Fit Tests to Test Model Adequacy**

Test	DF	Chi-Square	P-Value <sup>a</sup>
Deviance	1039	506.53	1.000
Pearson	1039	998.67	0.811
Hosmer-Lemeshow	8	14.63	0.067

<sup>a</sup>if p-value < .05, the model does not fit the data.

Another thing to consider is how well the model reflects the observed data. There is a need to look at the number of concordant and discordant pairs in the proposed model. The proportion of concordant/discordant pairs is a measure of the level of agreement between the model predictions and the observations. Table 5 shows the proportion of concordant pairs (80%) compared with the discordant pairs (18.90%), a result that is desirable.

**Table 5. Measures of Association Between the Response Variable and Predicted Probabilities**

Pairs	Number	Percent	Summary	Measures
Concordant	67964	80.00%	Somers' D	0.61
Discordant	16044	18.90%	Goodman-Kruskal Gamma	0.62
Ties	987	1.20%	Kendall's Tau-a	0.10
Total	84995	100.00%		

Data Concordance: Concordant Pairs (80.0%) > Discordant Pairs (18.9%); Note that the higher the concordance ratio, the better is the model.

**DISCUSSION**

This paper, from the beginning, hypothesized that students' academic history, physical attributes and social support significantly affect their intention to pursue higher education. Result of the analysis revealed, however, that only physical attributes and social support significantly affect students' intention to pursue higher education. Specifically, of the eleven initially chosen variables, only students' age and sex, and parents' education and income are considered significant predictors of students' intention to take higher education.

Passing of all the criteria for model adequacy and data concordance, the regression model clearly states that:

- (i) Students whose parents are educated with high income are 1.77 times more likely to pursue higher education than not;
- (ii) Older Female Students are less likely to pursue higher education.

This result explains why parents' personal educational background has a significant effect on their children's education [16] – [19]. Perceived adult support was a significant predictor of a plan to pursue a Bachelor's degree [20] because children find strength in their struggle for academic undertakings.

A survey reported that fully 76 percent of those who had two parents with degrees entered a two- or four-year college immediately after high school; that was almost double the 37 percent of those from no-degree families who did so. Exactly half of those from no-degree families went from high school into the workforce or military; just fewer than one in five of those from two-degree families did the same [21]. All these things were made possible because children of educated parents have that trust that their parents would be of help in all academic endeavors.

Another considerable significant factor that contributes to the completion in higher education is parents' financial capacity. The model explains that parents' schooling usually defines the parametric influence of parents – “the assumption which cause many educated parents to invest more in their children's education” [22]. Family has been identified as a factor of influences to students in the decision to enroll in tertiary education. The factors include family as a resource provider, source of encouragement for tertiary education, and as role models to their children.

Parents who received higher education can model and may have a greater emphasis on children by instill that higher education are essential and important [23] for economic stability. Desforges and Abouchaar [24] states that "if students are to maximize their potential from schooling they will need the full support of their parents". Full parental support includes financial aspect which could be provided when the parents have decent-paying job.

Moreover, children who feel that their parents can provide them their financial and educational needs would tend to be motivated to pursue higher education. They will be challenging more themselves to finish higher education to reach the same status as their parents have. This parental influence develops within their children the culture of valuing education as an attempt to secure economic life.

Such parental economic modelling is a continuing challenge to every Filipino parent in the Philippines whose average monthly income is 22 Thousand Pesos [25] making the Philippines a low middle-income country. However, despite this reality, 87% of Filipinos still believed that there is a direct correlation between education attainment and the prospects of having higher income in the workplace. Education for Filipinos is seen as a leveling factor that will help them compete for better jobs and better salaries [26].

The government, therefore, must be sensitive to this reality. Programs and projects should prioritize the increase direct parallelism between education and income. Creation of jobs, academe-industry curriculum matching, and other job-related activities must be treated with great importance to sustain family basic needs for education. Failure to address such would continually increase higher education dropouts among Filipinos, proliferating other possible worst effects.

The result further indicates that students' age and sex are significant considerable factors in their choice to proceed higher education. Specifically, the model reiterates those older females are less likely to pursue higher education.

Literatures revealed that 62 million girls around the world are not in school [27] due to some reasons like – cost of education, poverty, and distance to school. Moreover, some regions in India considered education not a priority for females. Instead, they want to invest on boy's education first. When opportunity comes, women are already at their age and have lost their motivation to continue education [28]. The reasons still go back on how parents prepare

themselves to better serve their children's need for access to education.

In the Philippines, common reasons why elder females are no longer interested to pursue education are: (i) despite opportunity, cost is high; (ii) lack of personal interest – most female are their mothers' counterpart in taking care of their siblings forgetting the desire for education; and (iii) Marriage – most older females opt to marry than to enroll higher education [29].

Lastly, it is encouraged that parents should encourage their young children to embrace education to its highest level by capacitating themselves economically to provide their basic educational and financial needs. For higher education institutions, greater access to higher education through various programs that would entice less privileged but deserving students.

## CONCLUSION

The pursuit for higher education among young students can be defined primarily by their parents' educational and financial capability. This study strengthens the parents' role as children's learning models, that is, that parents continually affect children's attitude towards education. Children who are educationally cultured by successful parents inspires their children on how to take charge of their own educational journey. However, this study is conclusively limited only to American' students whose data were retrieved through data mining in the Irvine Campus Machine Learning Repository Archive. Thus, a possible study can be replicated for data sets coming from Filipino students.

## REFERENCES

- [1] Koyama, J. P. (2007). Approaching and attending college: Anthropological and ethnographic accounts. *Teachers College Record*, 109(10), 2301-2323.
- [2] Haur, L. S. (2009). *Higher education marketing concerns: Factors influenceing Malaysian students' intention to study at Higher Education Institutions* (Doctoral dissertation, University of Malaya).
- [3] McInnis, C. (2001). Researching the first year experience: where to from here?. *Higher Education Research & Development*, 20(2), 105-114.
- [4] Biggs, J. B. (2011). *Teaching for quality learning at university: What the student does*. McGraw-Hill Education (UK).
- [5] Ridley, M. (2004). *Evolution*. Malden.

- [6] Kallio, R. E. (1995). Factors influencing the college choice decisions of graduate students. *Research in Higher Education*, 36(1), 109-124.
- [7] Lowe, H., & Cook, A. (2003). Mind the gap: are students prepared for higher education? *Journal of further and higher education*, 27(1), 53-76.
- [8] Rosa, M. L. D. L. (2006). Is opportunity knocking? Low-income students' perceptions of college and financial aid. *American Behavioral Scientist*, 49(12), 1670-1686.
- [9] Bers, T. H., & Galowich, P. M. (2002). Using survey and focus group research to learn about parents' roles in the community college choice process. *Community College Review*, 29(4), 67-82.
- [10] Butner, B., Caldera, Y., Herrera, P., Kennedy, F., Frame, M., & Childers, C. (2001). The College Choice Process of African American and Hispanic Women: Implications for College Transitions. *Journal of College Orientation and Transition*, 9(1), 24-32.
- [11] National Postsecondary Education Cooperative (2008) retrieved at <https://goo.gl/hWjDwD> on August 7, 2017
- [12] Bandura, A. (1989). Social cognitive theory. W: Annals of child development. Vol. 6. Six theories of child development. Vasta R (red.).
- [13] Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of vocational behavior*, 45(1), 79-122.
- [14] University of California – Irvine (2008). UCI Machine Learning Repository. Student Alcohol Consumption Data Set retrieved from <https://goo.gl/QRTrzW> on September 18, 2016.
- [15] Paul, K. (n.d.) Multicollinearity: Causes, Effects and Remedies. Retrieved at <https://goo.gl/z7Bnfy> on February 20, 2012.
- [16] Gratz, K. L., & Roemer, L. (2008). The relationship between emotion dysregulation and deliberate self-harm among female undergraduate students at an urban commuter university. *Cognitive behaviour therapy*, 37(1), 14-25.
- [17] Dubow, E. F., Boxer, P., & Huesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill-Palmer quarterly (Wayne State University Press)*, 55(3), 224.
- [18] Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, 19(2), 294.
- [19] Dearing, E., McCartney, K., & Taylor, B. A. (2006). Within-child associations between family income and externalizing and internalizing problems. *Developmental psychology*, 42(2), 237.
- [20] Zebrak, K. A., Le, D., Boekeloo, B. O., & Wang, M. Q. (2013). Predictors of intent to pursue a college health science education among high achieving minority 10th graders. *Current issues in education (Tempe, Ariz.)*, 16(2).
- [21] Brownstein, R. (2014). Are college degrees inherited. *The Atlantic*.
- [22] Gang, I. N., & Zimmermann, K. F. (2000). Is child like parent? Educational attainment and ethnic origin. *Journal of Human Resources*, 550-569.
- [23] Chenoweth, E., & Galliher, R. V. (2004). Factors influencing college aspirations of rural West Virginia high school students. *Journal of research in rural education*, 19(2), 1-14.
- [24] Desforges, C., & Abouchaar, A. (2003). *The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review* (Vol. 433). Nottingham: DfES publications.
- [25] Philippine Statistics Authority (2015). Income and Expenditure. Retrieved at <https://goo.gl/ANeNK9> on August 10, 2017.
- [26] Lucas, DL (2017). Filipinos lead global average in education survey. Retrieved at <https://goo.gl/azV2zL> on August 10, 2017.
- [27] USAID (2017). Let girls learn. Retrieved at <https://goo.gl/uqnP9a> on August 10, 2017.
- [28] Patil, A. (2014). Why Many Indians Can't Stand to Use the Toilet. Retrieved at <https://goo.gl/wta4Qc> on AUGUST 10, 2017
- [29] Olchandra, R. (2011). 7-Year Development Plan Guide for Philippines Signed

#### COPYRIGHTS

Copyright of this article is retained by the author/s, with first publication rights granted to APJMR. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4>).