ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645

Methodology for Obtaining a Predictive Model Academic Performance of Students from First Partial Note and Percentage of Absence

Amelec Viloria¹ and Alexander Parody²

¹Universidad de la Costa, Barranquilla, Colombia; aviloria7@cuc.edu.co ²Universidad Autonoma del Caribe, Barranquilla, Colombia; alexander.parody@uac.edu.co

Abstract

Objectives: This study presents the methodology for a model of multiple linear regression to assess the impact of the first partial grade and the percentage of non - attendance in the final grade students. **Methods/Statistical Analysis:** Descriptive Statistics and Inferential a program Industrial engineering a university in Colombia. **Findings:** After the generation and validation of the model was obtained that it explains 83.38% of the variability of the final grade students analyzed (134 students), and this significantly high percentage as a tool to determine the outcome of a student and generate recovery strategies those with a very low projection in its final note, it should be noted that the model was rigorously validated statistically. **Application/Improvements:** This methodology is proposed as a model for similar studies in other institutions.

Keywords: College Dropout, Multiple Linear Regression, Prediction

1. Introduction

Student desertion is a scourge that affects greatly the desire of governments of developing country possessing increasingly prepared generations, they can become the engine of economic development for the welfare of society, but increasingly figures are more alarming dropout. The above is especially true in higher education, where in a country like Colombia, students who start college about half fails to finish, of which 72% desist to continue studying in their first two years of training, in a country where the number of professionals trained between 2001 and 2010 is 1'802.726, and l population approaches 48 million years¹.

2. Theoretical Foundation

One of the factors that undoubtedly influence the dropout rate of students is academic performance of them during their university studies¹, so in the process of teacher education has always been a topic of interest to determine what psychosocial factors influence student performance^{2–7} either in the context of basic vocational education and/or, but if something can evidence in these investigations it is that the weight or contribution adds each measured variable to the fact that the student has or does not perform well in the field view, it is particularly the study population, whether an educational institution⁵ or a geographical region¹² which can hardly be extrapolated

^{*}Author for correspondence

to other institutions or latitudes, situation given by the same inherent variability in the relationship of these psychosocial variables, student performance and the study site, generating said predictive model of notes present a very limited scope. But if something is common in any educational institution, is the fact that partial marks students taken during their course of study, in addition to making assistance to students in courses (when they are face), potentially turning these variables into factors that influence the final grade of the student, as partial note provides the calculation of the final mark will inevitably have a correlation with the final mark and if this is added the effect of the absence of students to the classes, you might consider a possible predictive model associated with these two variables. In this research the aim to raise is a methodology for obtaining a predictive model of student grades in any educational institution, from the first partial grade students and the percentage of non-attendance accumulated so far of making such partial note, thereby enabling to obtain a primary tool when addressing about students develop student retention strategies associated with underachievement in the institution.

3. Methodology

For an explanation of the methodology that will allow obtaining the mathematical model, a population of 137 students belonging took 5 courses Statistics (2 courses Descriptive Statistics and three groups of inferential statistics) a program of university Industrial Engineering in the city Barranquilla (Colombia) during the second half of 2014. this group of students were picked up his first partial grade and the percentage of absences presented by each of them until the definition of that note, later built a model of multiple linear regression linking these two variables with the final grade obtained by the students at the end of the course and from the percentage of variation explained by the model (adjusted R2) and residue analysis model, determine which so suitable is the model to predict the final note of students from these explanatory variables, as a last step the model information 171 schoolboys validated is corresponding to 6 courses of Descriptive Statistics and Inferential during the first half of 2015, performing just as an analysis of the waste generated by the regression model⁵.

A description of the two explanatory variables shown: **Partial Note**: Refers to the result of the first partial note that achieves the student, and in this case

corresponds to 30% of the final mark, the participation of this, in the calculation of the final note guarantee a correlation at least one minimum level with the final note, it is stated that this partial note could be a good element to reach predict the final grade.

Percentage of Absence: Refers to the percentage of absence that has had the student so far to define the first partial note, with respect to the total contact hours of the course, which for purposes of this analysis corresponds to a course of 48 contact hours.

4. Results

After statistically analyzed using the statistical software Statgraphics XVI 137 data, the following model of multiple linear regression was generated:

With the following results of tests of significance for how model and components thereof (see Table 1 and 2):

Table 1. Test of significance for model parameters

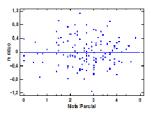
		Error	Statistical	
Parameter	Estimate	Standard	Т	Value-P
Constant	1.43843	139,929	10.2797	.0000
Percentage of Absence	0898458	.00832674	-10.79	.0000
Partial Note	605,053	.0406812	14,873	.0000

 Table 2.
 Analysis of variance for the model

Source	Sum of	Gl	Horsebit	Reason-F	Value-P
	Squares				
Model	111,299	2	55.6494	334.78	.0000
Residue	21.7757	131	166,227		
Total (Corr.)	133,075	133			

The model proved to be highly significant (p-value of 0.0 for Model, see Table 2) also obtained a R² value Adjusted 83.38% which clearly says the high percentage of the variability in final grades of students statistics of Industrial Engineering program that the model explains, without the excellent results obtained in tests of significance to model components (all p-values were in .0 for all model parameters, see Table 1). Subsequently the waste

generated by the model were calculated, these residues are the difference between the actual value of the final grade of each student and the estimated note generated by the model, with residue analysis we sought to prove that there is some kind of correlation between the waste and the various parameters of interest related to the model or construction, by means of scatter plots that link these variables (see Figure 1 and 2).



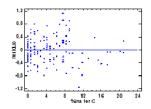
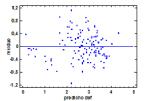


Figure 1 and 2. Note Partial versus Waste and Percentage of waste versus absences to take partial note.

The figure above indicates as waste as they increase the values of the Partial Note on students, which show that waste have a random behavior and does not depend on the value presented in Note Partial distributed.

Figure 2 indicates as waste are distributed as values increase the percentage of non - attendance of students, which show that waste have a random behavior and not dependent on the value of nonattendance used in the model.

In Figure 3, which relates the value of the waste with respect to the predicted values of the final grade students, it can be demonstrated that there is complete randomness of waste relative to the predicted values.



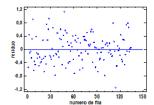


Figure 3 and 4. Waste versus Predicted final note and Waste versus row number.

Figure 4, review the behavior of waste with respect to the position of each data used in constructing the model, seeking to determine whether there is a correlation between the waste generated and the position of the data, as will be seen, there is no correlation between waste and the corresponding row numbers.

The above figures show the lack of correlation between the residuals (or prediction errors) and the various parameters of interest in the model, to complete the analysis of the suitability of the model, a Kolmogorov-Smirnov test was performed to determine that the waste follow a normal distribution with a mean equal or close to 0, resulting in that they follow a normal distribution (p-value of 91,185) with an average of 0,000,000,144 which approximates greatly to .0.

The fact that the model satisfactorily fulfilled each of the above explained evidence leads to the conclusion that it may be used with great success as a tool for prediction of the final grade of students in courses of Descriptive Statistics and Inferential in a Program university Industrial engineering. The above analysis associated with the formulation of the model can be summarized in this series of 4 steps that can be applied to obtain a model that helps predict with great success the final grade students of any course in an educational institution:

- Collect the first partial grade obtained by students in the previous study period, together with the percentage of non - attendance to the determination of the first partial note also of the final grade students at the end of the educational period.
- Generate multiple linear regression, where the variables Note Partial and Percentage of Absences are the explanatory variables and the Final Note of students is the response variable, testing the significance of both the model and its components, obtaining a p-value less than or equal to .05 in tests of significance both model, as the two explanatory variables.
- Verify that the model highly explain the variability of the final grade students, through R 2 Adjusted, it must be at least 80%. In the case that the value of adjusted R² is less than 80%, test different models to multiple linear subsequently verifying the significance of both the model and the components, in addition to the adjusted R 2.
 - Test the suitability of the model, applying the various tests to waste:
 - That the variability of the waste remains constant when a scatter plot residues are compared against: the values of the explanatory variables, the position

- of each line of data used to generate the model and against the predicted values.
- That the residues follow a normal distribution with a mean equal to 0 or very close to it.

Compliance with these requirements ensures that the model obtained is suitable to predict with much certainty the final grade students.

To test the efficiency of the model, it was used to predict the final note of students who had taken courses Descriptive Statistics and Inferential during the first half of 2015, a total 171 students in 6 groups discriminated. After applying the model to generate projections of final notes on the 171 students once they took the first partial grade, they were compared with the marks obtained by these students at the end of the courses Statistics, yielding an average forecast error of .16 on a scale going from 0 to 5, which indicates the highly effective was the forecast generated by the regression model as the next step forecast errors generated by the model were analyzed, using the same methodology of analysis used with waste during validation model, obtaining that they followed the behavior of a normal distribution (Kolmogorov-Smirnov test showed a p-value of .23), with an average of .16 which is not far from 0.

5. Conclusions

Obtaining predictive model of the final grade students statistical program Industrial Engineering at the Autonomous University of the Caribbean, it is precisely what is described in its scope a particular model of regression applies to a characteristic population, but model that has only two explanatory and high R² value Adjusted variables, plus excellent performance of waste and mistakes in conducting forecasts, compared to other studies conducted that handle a greater number of explanatory variables and have an R² value set too low, or that handle a number of explanatory variables that apply only to the study population and can these variables change their relationship with the response variable to change the study population, either with respect to geographical space, the professional program training or level of study in which it will oped of the regression model⁷ making the model has a very limited scope, therefore noteworthy that the most important point of this research is that not only focuses on proposing a particular model to generate forecasts of the final grade students courses Statistics from the Autonomous

University of the Caribbean, because it is already clear that it will have a limited scope, since it would only apply (as expected) on the student population analyzed, but that a series of sequential steps is proposed that will enable the various institutions of education to obtain a sufficiently explanatory model of the final note of the students, since the first partial grade and the percentage of absences are variables directly associated with the final student academic performance. Achieving model will directly affect the implementation of strategies for improving the retention of students in different educational institutions, as will be known in advance that students would have a higher tendency to lose the subject and therefore with respect to them different strategies for obtaining an improvement in their academic performance, thereby reducing student desertion at least defection caused by poor academic performance, motivated by achieving this goal this methodology is proposed for obtaining be formulated the model beyond presenting a single model to predict the student's final grade.

As a proposal for further study arises test whether there are characteristics of student populations that affect positively or negatively on the percentage of explanation of variability (adjusted R²) of the final mark by the variables Note Partial and Percentage of Absence, in this way you can evolve increasingly in the generation of mathematical models that become effective tools for managing student retention in educational institutions.

The application of this methodology will not only increase levels of retention of college students, but a positive impact on society by having more and more individuals with a university education to help promote socially and economically weak nations in these key fields the structure of the countries.

6. References

- Ministry of National Education of Colombia. Desertion is decreasing Strengthen Human Capital. Higher Education. 2012 May; 20:7-10.
- Zuniga D, Mena B, Oliva R, Pedrals N, Padilla O and Bitran M. Prediction models Academic Performance of Medical Students of Basic and Preclinical Cycle. A Longitudinal Study. Chile Medical Journal. 2009 June; 137:1291-1300.
- Carvajal P, Mosquera J and Artamonova I. Prediction models in Mathematics I Academic Performance at the

- Technological University of Pereira. Scientia et Technica. 2009 June; 43:258-63.
- 4. Baena-Extremera A and Barn-Gallegos A. Prediction Model of Satisfaction with Physical Education and School. Ctica Psicodida Magazine. 2015; 20(1):177-92. DOI: 10.1387/RevPsicodidact.11268.
- 5. Rosario P, Lourenco A, Paiva O, Rodrigues A, Valle A and Tuero-Herrero E. Predicting Performance in Mathematics: Effect of Personal Variables, Socioeducativas and the school context. Psicothema. 2012; 24(2):289-95.
- 6. Reyes J, Escobar C, Duarte J and Ramirez P. An Application of Logistic Regression Model in the Prediction of Student Achievement. Pedagogical Studies. 2007; XXIII(2):101-20.
- 7. Nunez J, Vallejo G, Rosario P, Tuero E and Valle A. Variables Student, Professor, and Context in Academic Performance Prediction in Biology: Analysis from a multilevel perspective. Ctica Psicodida magazine. 2014; 19(1):145-72. DOI: 10.1387/RevPsicodidact.7127.