FACTORS ASSOCIATED WITH THE ACADEMIC PERFORMANCE OF ENGINEERING STUDENTS OF THE UCR INTERUNIVERSITY CAMPUS FROM ALAJUELA

Sofia Bartels & Hennia Cavallini

Abstract:

The Alajuela Interuniversity campus opened in 2007, with Universidad de Costa Rica as its first academic institution, which has two engineering majors. Since recent studies have shown that students take longer to get to their major's last year because they fall behind in their curricula, it is decided to determine the factors associated with the academic performance of engineering students through Generalized Linear Models based on the theories from Tinto (1989), Vélez, Schiefelbein and Valenzuela (2001) and Blanco Vega (2011). As a result it is decided that the best model to follow is the Negative Binomial Model with dependent variable; it is taken into account the amount of approved credits and independent variables: if the student has considered to leave the university, the education of the student's household head, if the student has scholarship, if the student believes that the university is located in a dangerous area, the amount of majors the student has taken and the student's main major.

Keywords: Academic performance, Generalized Linear Models, Engineering.

I. Introduction

The Alajuela Interuniversity campus opened in 2007 according to articles 04-07 from February 13th, 2007 and 07-07 from March 6th, 2007 from the National Council of University Rectors (CONARE, by its Spanish acronym). This campus has a special characteristic because it joins the four public universities in this moment, which are the following: Universidad Nacional, Instituto Tecnológico de Costa Rica, Universidad Estatal a Distancia and Universidad de Costa Rica.

In this campus, Universidad de Costa Rica offers the following majors: Graphic Design, Industrial Engineering and Mechanical Engineering. The campus has a total of 422 active students from the engineering majors, from this amount 320 are Industrial Engineering students and 102 are Mechanical Engineering students.

The campus has some limitations because of the short time it has existing, for example it is located in a place it does not own, it is located in a Shopping Centre the universities have to rent. On the other hand, there is no dining hall where the students can get food at a lower price or subsidized.

At the beginning of 2012, an analysis was performed to determine the time the students from Mechanical Engineering took to get to the last year of their major and it reflected that the time they took was too long. There were cases of students that started the major in 2007 and there were still subjects from the third year of the major that they needed to approve. For that reason, this work's

objective is to determine the factors associated with the academic performance of engineering students from Industrial and Mechanical engineering through Generalized Linear Models. It was expected that some of the aspects mentioned before are affecting the academic performance, but maybe the impact of these variables is lower in comparison with the ones associated to the students' homes and living conditions.

Generalized Linear Models allow to determine the variables that are associated with the performance, the impact those variables have over it and the possibility to estimate the performance of one student in particular.

II. Theoretical framework

The academic performance is an important variable to consider in the university context; therefore different models have been created to try to explain the behavior of this variable through theories or related variables. This section purpose is to locate us in a theoretical basis allowing to distinguish the appropriate variables to be included in a Generalized Linear Model that determines the students' academic performance.

Tinto; (1989 cited in Romo, et al, 2000, p.1) found five constructs that are related to the lap, these are:

- Psychological
- Social or environmental
- Economic factors
- Organizational
- Interrelations

The first construct refers to the subject's own characteristics related to the personality, disposal, motivation, ability and capacity. The second component is described as the forces that establish the success or failure of the student, for example the individual social status. The third component has relation to the benefits linked to the financial resources. The fourth element is presented with aspects related to the institution; it is exemplified with the available resources of the institutions, size, and environment among others. In the case of the last variable, it refers to the aspects involved to the behavior as a result of the interaction between the environment and the subjects.

These factors are stated in other works. Sotelo, *et al* (2009) determines that there are significant and positive correlations among the auto efficiency, the assessment of the task, anxiety and guidance to the academic performance. These variables are measured in a survey where a scale in the applied questionnaire is incorporated.

"Considering the results found in this study, it is shown how a significant relationship of motivational aspect exists, of the auto efficiency in particular, with the academic performance." (Sotelo, *et al.*, 2009 p. 9).

Blanco Vega *et al.* (2011) establishes that the auto efficiency of the students is related to the motivation and, at the same time, to the academic performance because if the students have a bigger motivation then the results in the academic performance are better. Among the variables where the auto efficiency is measured, the student's concentration is influenced.

Some of the questions we find included in the questionnaire that measure the concentration are:

- *"I listen carefully when the professor clarifies a doubt to a classmate"*
- *"I pay attention when the professors teach the class"*
- *"I pay attention when a classmate makes an exposition"*

Alcover, *et al.* (2008) makes a linear regression that tries to predict the variables related to the university degrees, this author takes those degrees as a measure for the students' performance.

Among the variables that are significant in this model we find the following:

- Other university studies of the student when enrolling in the degree.
- The parents' studies.
- The student's occupation.

The first two have a positive sign; the third variable was grouped according to the amount of hours worked by the student. The first group was made of students that worked less than 15 hours and the second one by students that worked 15 hours or more. The students that worked less than 15 hours had a better degree than those who worked 15 or more hours per week.

Vélez, Schiefelbein & Valenzuela (2001) identifies different variables that are associated with a low performance, some of them are:

- The school's distance

  "Another systematic finding is that the distance to the school is associated negatively to the performance". (Vélez, Schiefelbein, Valenzuela, 2001, p. 8.)
- Health and nutrition

  "Health and Nutrition are significant predictors of the performance". (Vélez, Schiefelbein & Valenzuela, 2001, p. 9).
- Access to books and materials

  "The access to textbooks and other instructional material is important to increase the academic performance."(Vélez, Schiefelbein, Valenzuela, 2001, p12.)

Blanco Bosco (2008) creates a linear regression model that tries to measure the performance in Math and finds significant variables in the model such as the gender, where women have better performance than men. Regarding the work variable, the students that have to work have a lower performance than those who do not work; it also showed that if the student abandoned the school, at least for a year, it affects the performance negatively.

III. Methodology

This section describes the methodology used in both the questionnaire making, which contains the variables of study, as well as in the construction of general linear models that comply with the objective.

a. Population of this study

This study takes into account all active students of the engineering major from Universidad de Costa Rica from the Alajuela Interuniversity campus. Currently, the campus offers two engineering majors: Mechanical and Industrial Engineering.

b. Sampling Method

Initially, in order to obtain the sample, the sample frame was created which is composed of the list of active students in the second semester of 2012.

It was preferred to perform a random simple sampling from the frame, where students were selected individually to minimize the risk to find duplicated cases that would have been possible if group of students from different courses would have been selected. It is known that there is an important lap regarding this.

c. Sample Size

The sampling frame is composed of 422 students; which correspond to 102 Mechanical Engineering students and 320 Industrial Engineering students. Hence the simple size is 122 students.

d. Tool construction

In order to measure the suggested objective, a questionnaire with different sections is created, 5 in total, from these sections the items composing the study variables were selected.

The procedure for the items' construction was based on the consulted bibliography.

e. Generalized Linear Models Construction

For the construction of these models, a descriptive analysis of the variable response behavior was performed, which helped to determine that what was appropriate to use were the generalized linear models of Poisson or the Negative Binomial model. For the distribution of one of the dependent variables, the possibility to use the variation of the Zero-inflated Poisson model was evaluated.

IV. Results

The main results found in the data analysis are detailed in this section.

Descriptive analysis of the variables

General characteristics of the study variables are shown in this section, as well as how to define the way in which performance was measured.

1. Dependent Variables

The variables that try to measure the academic performance of the engineering students of the Alajuela Interuniversity Campus are described in this section.

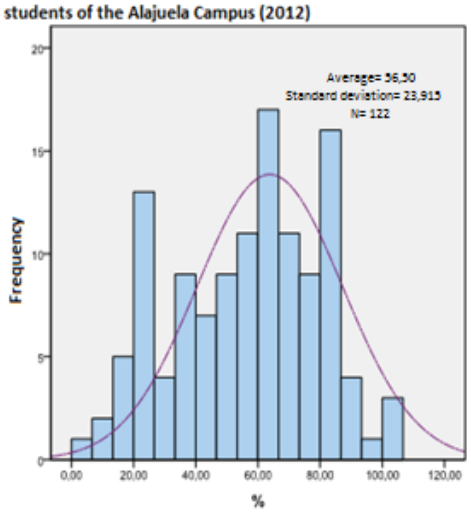
Table 1.Variables that measure the academic performance of the engineering students

Number	Name
1	Amount of approved credits
2	Amount of failed subjects

Source: Own elaboration

In table 1, a list of variables is observed. They try to measure the academic performance of the engineering students. A distinction by admission year is made for the amount of approved since, for instance, fourth year students should have more approved credits than second year students. This is also applied in the case of the variable amount of failed subjects.

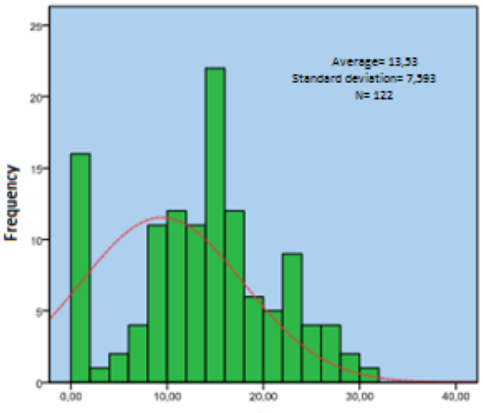
Graphic 1. Percentage of credits approved by the engineering students of the Alajuela Campus (2012)



In average, the students of the campus in the engineering majors have approved the 56,5% of the credits that they should have approved according to their syllabus. (See graphic 1)

On the other hand, the engineering students from the campus have failed 13,5% in average the subjects that they have taken in their syllabus. It is important to mention that it is represented as a percentage because the number of subjects failed from the program depends on the amount of

Graphic 2. Percentage of failed subjects from the syllabus by the engineering students of the Alajuela Campus (2012)



subjects they should have approved. (See graphic 2)

2. *Independent Variables*

Next, the independent variables are detailed. They will be used in the regression models.

Table 2. Variables that could influence the performance of the engineering students and how they could influence them.

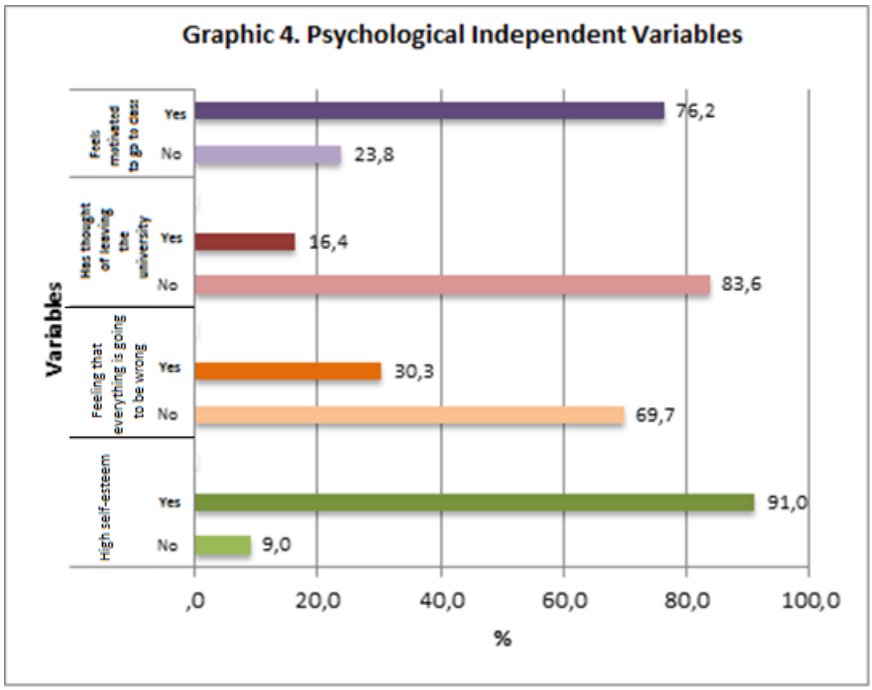
Category	Name	Symbol
Psychology	High self-esteem	+
	Feeling that everything is going to be wrong	-
	Has thought of leaving the university	-
	Feels motivated to go to class	+
Social or environmental	Female	+
	Age	-
	Household head with superior studies than high school studies	+
Economic factors	Works	-
	Has socio-economic scholarship	-
	Sometimes could not have lunch because of lack of money	-
	Can get the books and materials requested in the courses easily	+
Organizational	Believes that the library has the necessary material to support his / her studies	+
	Likes the university installations	+
	Considers that the university is located in a dangerous zone	-
Interrelations	Lives in a neighborhood where most of the neighbors studied at the university	+
	The relationship with my classmates is very good	+
	Studies Industrial Engineering	+
	Time it takes to get from the house to the university	-
	Number of majors the university has	-
	Problems to concentrate in the study	-

Source: Own elaboration

In table 2, the variables that will be considered initially as independent variables in the model are shown. These variables were chosen according to the consulted theory and they were classified in 5 components. In addition, the symbol that was expected to be obtained in the models was established.

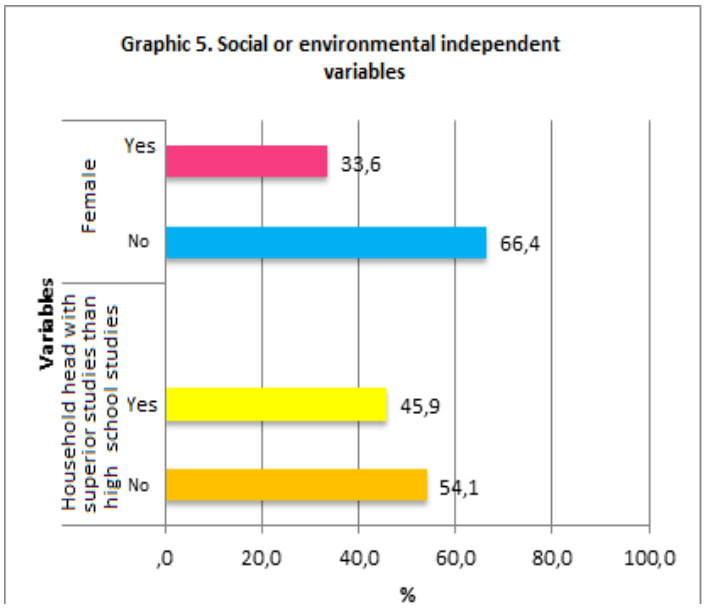
a. Psychological Variables

In this section, the variables that belong to this construct are described and they are associated to the performance according to the consulted bibliography.



For the high self-esteem variable, there is a low percentage of people (9%) that say they have a low self-esteem. The variable "feeling that everything is going to be wrong" is more distribuited in the different categories, 30,3% of the people say they feel that everything is going to be wrong. It is also noticeable that 23,8% of the interviewers say not to feel motivated to go to class. On the other hand, 16,4% of the students say having thought of leaving the university (see graphic 4).

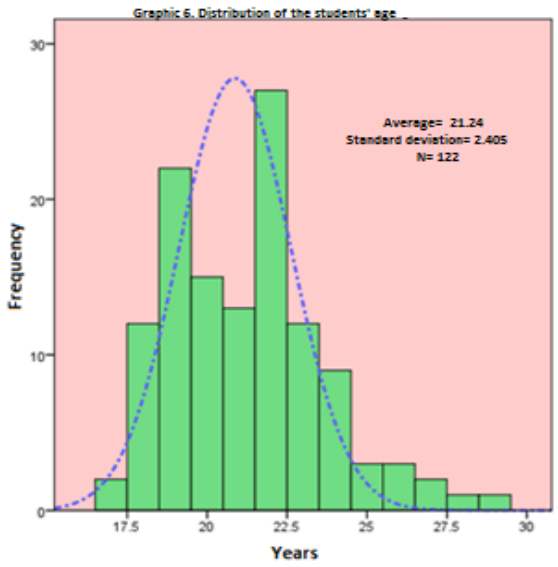
b. Social or environmental



Next, the variables classified as social or environmental are detailed.

33,6% of the students interviewed were women; this is because there is a higher proportion of men in the Mechanical Engineering major.

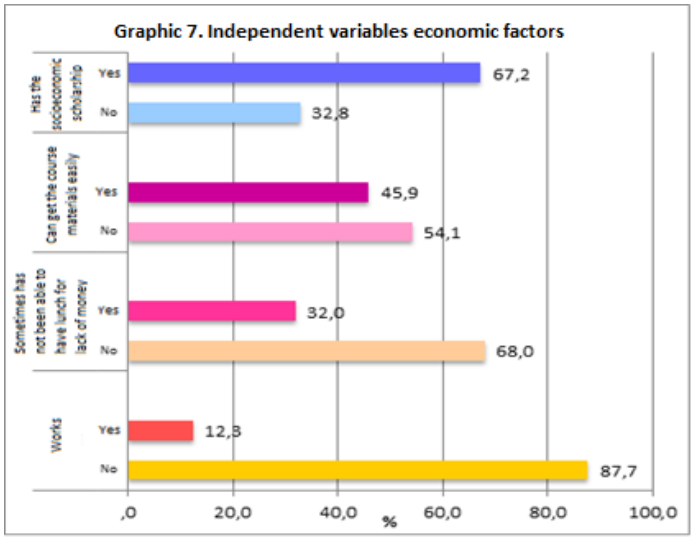
In addition, 45,9% of the students have a household head that has more than high school studies, so



that a high percentage of the students have a household head with low schooling (See graphic 5).

The age variable has a positive asymmetrical distribution, in average the students are 21,24 years old, the age range is from 17 to 29, and these high ages are due to the students' delay when failing courses from the major. (See graphic 6)

c. Economic factors



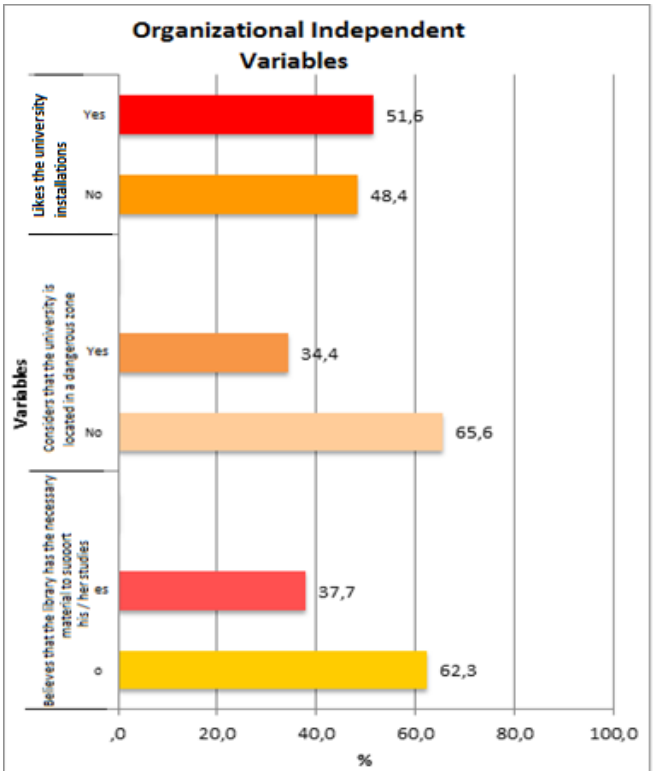
The variables included in this section are related to the economic influence that is exerted on students.

12,3% of the students work, there is also a 32% that say that sometimes they have not been able to have lunch for lack of money. 54,1% of the students say they could not get the requested materials in the courses easily. Also, 67,2% of the students have socio-economic scholarship.

Because of what is mentioned before, it is concluded that an important percentage of students has a difficult economic situation.

a. Organizational Independent Variables

This section describes the behavior of the organizational variables that are linked to the contribution of the university in the students' performance.



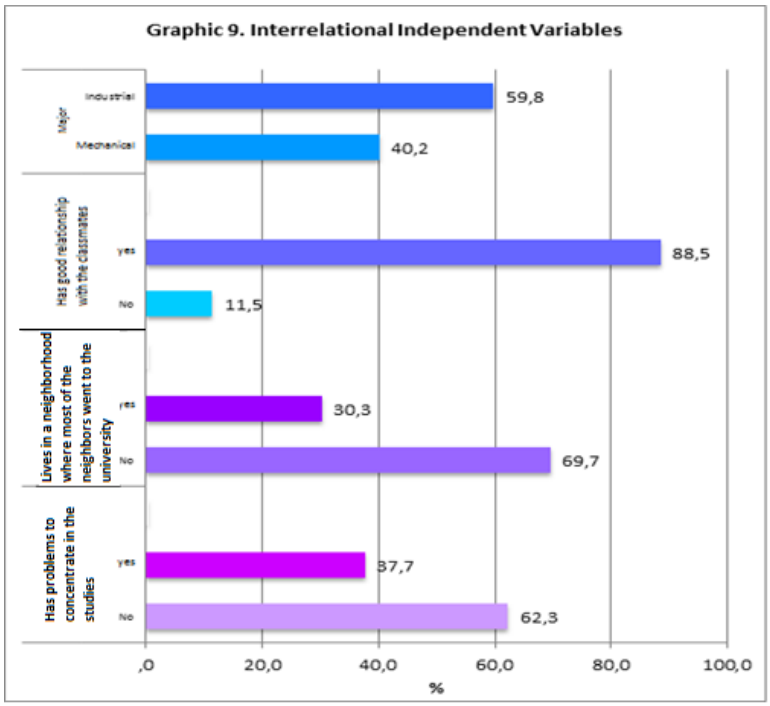
48,4% of the students do not like the university, this could be related to the university location which in this moment is located in a shopping center. In the meantime, CONARE is trying to buy the land and starts the construction.

Also 34,4% of the students consider that the university is located in a dangerous zone, this could be related that, at the moment the students were interviewed, thefts were occurring in the buses that transport the students, these buses are part of the public service system.

On the other hand, 37, 7% of the students believe that the library does not have the necessary material to support their studies; this could be because the library is very small (see graphic 8).

a. Interrelational Independent Variables

These variables have relation in how the student gets interrelated with the environment; its behavior is detailed below.

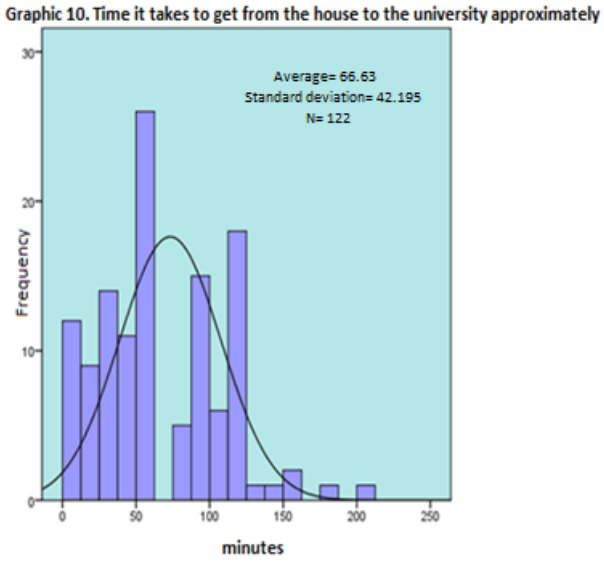


59,8% of the interviewed people study the Industrial Engineering major, there is a high proportion since the Mechanical engineering major is not open all years.

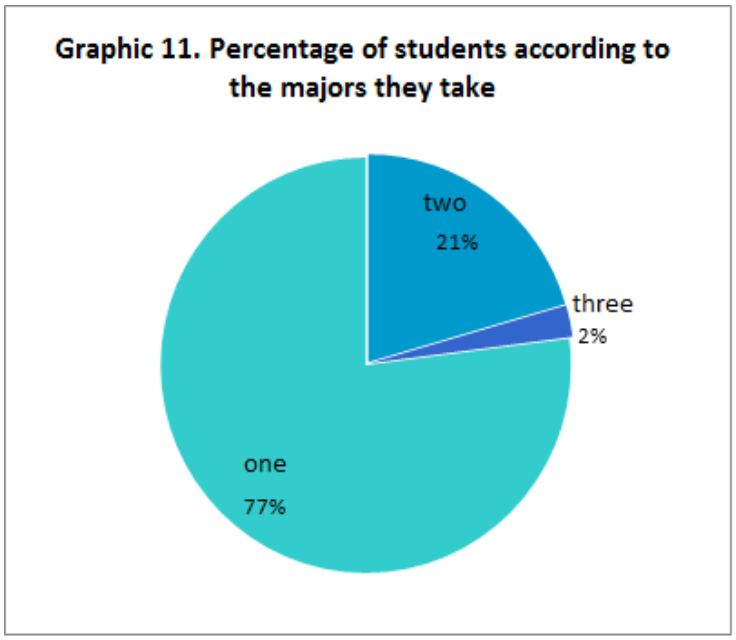
In addition, 11,5% of the students say that they do not have a good relationship with the classmates.

69,7% of the students do not live in a neighborhood where most of the neighbors went to the university. It means, they are surrounded by people with low schooling.

On the other hand, a high percentage (37,7%) says that they have problems to concentrate in the studies (See graphic 9).



The approximate time distribution in minutes that students take to get from their homes to the university is distributed relatively normal with a small symmetry to the right. In average, students take 66,63 minutes, which is equivalent to an hour and eleven minutes approximately. This could be related to the quality of the bus service (See graphic 10).



77% of the students have just one major. In general, the students with more than one major are taking the other major in San Pedro or San Ramón campus.

b. Generalized Linear Models for the performance valuation

The generalized linear model analysis is presented in this section that can be used for the dependent variables' valuation. A comparison is also done between the models in order to select the most appropriate.

Initially, it is analyzed the correlation between the variables in order to avoid the multicollinearity among the independent variables. For this purpose, the variables that present a correlation higher than 0,30 were identified.

Table 3. Variables with correlations higher than 0,30

Variables	Correlation
High self-esteem–good relationship with the classmates	0,34
Has thought of leaving the university - Age	0,31
Works–Has thought of leaving the university	0,31
Industrial - Female	0,42
Age- Works	0,42
Likes the university installations – The library has the necessary material to support his/her studies.	0,31

1. Dependent Variable: Approved credits.

According to the distribution of this variable and considering that it is the counting of approved credits, Poisson model is used at the beginning. It had to be exposed for the amount of credits that had to be approved.

Table 1. Poisson Model for the dependent variable: amount of approved credits

Variables	IRR	Dev.	P>z
High self-esteem	1.118	0.062	0.045
Has thought of leaving the university	0.856	0.032	0.000
Parents with more than high school diploma	1.360	0.035	0.000
Age	1.027	0.008	0.001
Works	0.846	0.039	0.000
Has socio-economic scholarship	1.328	0.043	0.000
Sometimes has not been able to have lunch for lack of money	1.094	0.031	0.001
The University library has the necessary material to support their studies	1.091	0.033	0.004
Considers the university is located in a dangerous zone	1.135	0.032	0.000
Likes the university installations	0.926	0.028	0.010
Amount of majors the student has	1.277	0.032	0.000
Concentration problems	0.915	0.026	0.002
Lives in a neighborhood where most of their neighbors went to the university	1.102	0.032	0.001
Industrial	1.306	0.036	0.000
Constant	0.118	0.021	0.000
ln (approved credits)	1	(exposure)	

Source: Own elaboration

This model does not have all the initial variables; some of them were correlated with other variables; for example the variable called good relationship with the classmates that was correlated with the variable high self- esteem (see table 3). Another variable eliminated was the variable called female, this one has correlation with the variable called industrial.

It can be observed that the 5 constructs related to the performance are represented in this model.

The Poisson model has an assumption that needs to be reviewed, in which the Poisson distribution has the variance equal to the average, hence a negative binomial model is designed.

Table 2.Negative Binomial Model for the dependent variable amount of approved credits

Variables	IRR	Dev.	P>z
High self-esteem	1.010	0.146	0.945
Has thought of leaving the university	0.801	0.085	0.036
Parents with more than high school diploma	1.385	0.105	0.000
Age	1.003	0.019	0.872
Works	0.886	0.109	0.324
Has socio-economic scholarship	1.315	0.120	0.003
Sometimes has not been able to have lunch for lack of money	1.158	0.097	0.079
The University library has the necessary material to support their studies	1.062	0.088	0.467
Considers the university is located in a dangerous zone	1.191	0.095	0.029
Likes the university installations	0.934	0.076	0.398
Amount of majors the student has	1.314	0.106	0.001
Concentration problems	0.885	0.069	0.116
Lives in a neighborhood where most of their neighbors went to the university	1.140	0.097	0.124
Industrial	1.411	0.113	0.000
Constant	0.196	0.083	0.000
ln (approved credits)	1	(exposure)	
Alpha natural algorithm	-2.182	0.167	
alpha	0.113	0.019	

Source: Own elaboration

This model makes the following parametrical hypothesis test:

$$H_0: \lambda = \sigma^2$$

$$H_a: \lambda \neq \sigma^2$$

The value p of this test gave 0,000 so that with 95% certainty, the null hypothesis that the average is the same as the variance is rejected. Therefore, the Poisson model could not be used. Also, notice that some variables in the negative binomial model are not significant, but they will not be eliminated because they are theoretically substantive to represent the model and the size of the sample can be influencing the results.

Now a valuation of a Logit model will be done grouped by the amount of credits the students should have approved.

Table3.Logit Model for the dependent variable amount of approved credits.

Variables	Odd Ratio	Dev.	P>z
High self-esteem	0.697	0.038	0.000
Has thought of leaving the university	2.056	0.085	0.000
Parents with more than high school diploma	1.061	0.013	0.000
Age	0.744	0.050	0.000
Works	2.052	0.099	0.000
Has socio-economic scholarship	1.254	0.056	0.000
Sometimes has not been able to have lunch for lack of money	1.137	0.053	0.006
The University library has the necessary material to support their studies	1.375	0.060	0.000
Considers the university is located in a dangerous zone	0.891	0.041	0.013
Likes the university installation	1.879	0.081	0.000
Amount of majors the student has	0.787	0.034	0.000
Concentration problems	1.310	0.061	0.000
Lives in a neighborhood where most of their neighbors went to the university	1.880	0.083	0.000
Industrial	0.043	0.011	0.000

Source: Own elaboration

This model has all the significant variables and coincides with the Poisson model in that aspect.

In order to determine which model is the best, a comparison will be done by using the Akaike and Bayes criteria.

Table 4. Comparison of the Models for the dependent variable amount of approved credits

Model	Observations	AIC	BIC
Poisson	122	1409.80	1451.86
Negative Binomial	122	1061.65	1106.51
Grouped logistic	12620	15953.62	16057.82

Source: Own elaboration

In table 4, the model that has the lower AIC and BIC is the negative binomial, so that it would be recommended to use this model. When doing the projections for one person only the variables with significant results would be taken into account. With this model we can conclude that in average the students that have thought of leaving the university have 0,801 times the amount of credits from those who have not. In average, the students that have a household head with more than high school studies have 1,385 times the amount of credits from those who do not.

In addition, the students with scholarship have, in average, 1,315 times the credits from those who do not. Also for an increase in a major, the students increase the number of approved credits in 34% in average.

Finally, the industrial engineering students have in average 1,41 times the credits that the mechanical engineering students have, keeping the other variables constant. (See table 2)

2. *Dependent Variable amount of failed subjects*

The Poisson model should be used for this variable according to its distribution. (See graphic 2), when modeling this one, significant variables are obtained which are presented in table 4.

Table 4. Poisson Model for the dependent variable amount of failed subjects.

Variables	IRR	Dev.	P>z
Household head with more than high school studies	0.768	0.065	0.002
Has scholarship	0.822	0.072	0.026
Amount of majors	0.838	0.067	0.026
Constant	0.219	0.030	0.000
ln (subjects from the program)			

Source: Own elaboration

This model does not incorporate significant variables of all the components of the bibliography, it only gets to incorporate the social, interrelation and economic factors through a variable.

To prove the assumption of the required Poisson regression:

$$H_O: \lambda = \sigma^2$$

$$H_a: \lambda \neq \sigma^2$$

A parametric test was done, the negative binomial model is built. With a significance of 5% there is no sufficient statistic evidence to reject the null hypothesis that the average and the variance are equal (p=0,50).

Table 5. Negative Binomial Model for the dependent variable amount of failed subjects

Variables	IRR	Dev.	P>z
Household head with more than high school studies	0.768	0.065	0.002
Has scholarship	0.822	0.072	0.026
Amount of majors	0.838	0.067	0.026
Constant	0.219	0.030	0.000
ln(subjects from the program)	1.000	(exposure)	
Ln(alpha)	-15.796	645.637	
alpha	0.000	0.000	

In table 5 a summary of the model is shown; it can be observed that the same significant variables in the model are significant for the Poisson model.

The distribution of the variable answer presents some cases with zero value that can cause sub dispersion (see graphic 2); therefore it is convenient to contrast the Poisson model with the Zero-inflated Poisson model, because with the previous test we know that it is appropriate to use the Poisson model.

Table 6. Poisson Model inflated with zeros for the dependent variable amount of failed subjects.

Variables	IRR	Dev.	P>z
Household head with more than high school studies	0.803	0.068	0.009
Has scholarship	0.851	0.075	0.067
Amount of majors	0.822	0.066	0.014
Constant	0.219	0.030	0.000
Ln(subjects from the program)	1.000	(offset)	
Inflated Model			
Household head with more than high school studies	15.774	3671.543	0.997
Has scholarship	15.759	3538.251	0.996
Amount of majors	-15.78664	3386.929	0.996
Constant	-18.16172	6121.334	0.998

Source: Own elaboration

In table 6, the Zero-inflated Poisson model is summarized, this one makes the test:

$$H_0: LL(Poisson\ Model) = LL(Model\ with\ overdispersion)$$

$$H_a: LL(Poisson\ Model\ ) \neq LL(Model\ with\ overdispersion)$$

With a significance of 5% there is no sufficient statistic evidence to reject the null hypothesis that the models are equal (p=0,17), this has relation to the contrast of the model in table 5, where differences between the models were not found either; no over dispersion was found.

Table 5. Comparison of the Models for the dependent variable amount of failed subjects

Model	Observations	AIC	BIC
Poisson	122	511.800	523.016
Negative Binomial	122	513.800	527.820
Zero-inflated Poisson model	122	510.965	533.397

Source: Own elaboration

In table 5, a contrast of the models is done. By using the Akaike criteria, it should be chosen the Poisson model or the Zero-inflated Poisson model since the difference between both is only one point. In the case of the Bayes criteria, there are more than three points of difference between the models, so that we should select the one that has the lower BIC, in this case the Poisson model.

When selecting the Poisson model we can conclude that in average the students that have a household head with more than high school studies have 0.768 times the amount of failed subjects from those who do not.

The students that have scholarship have in average 0.822 times the failed subjects of the students that do not have scholarship. Also due to an increase in one major, the students decrease in 17.8% the number of failed subjects keeping the other variables constant (see table 2).

V. Conclusions

It can be proven that the students present particular characteristics that should be considered; as it is the fact that a high percentage has a feeling that everything is going to be wrong, this could be related to the fact that they are in average 21, 24 years old, they are almost finishing their studies, but there are students who are 29, when they are supposed to finish their major approximately between 23-24 years old.

Another characteristic that is reflected in the data is that the students present a difficult economic situation. A high percentage has socio-economic scholarship and some students say that sometimes they were not able to have lunch because of lack of money (32%). A high percentage of the students live in neighborhoods where most of the neighbors do not have university studies and they live a family unit where a high percentage does not have a household head with studies higher than high school.

The university limitations are visible since a considerable percentage of the students (48,4%) do not like the installations of the university which could be related to the sense of belonging to the campus.

In the construction of the negative binomial model with the dependent variable amount of approved credits, we have as the more significant variables to determine the performance, from the variable amount of approved credits, the following: if they have thought of leaving the university, the household head education, if he/she has socio-economic scholarship, the opinion that the university is located in a dangerous zone, the amount of majors the university has and the student's main major. Even we do not have any theoretical justification why students with better performance say the university is located in a dangerous zone, it might be thought that maybe they spend more time at the university, so they know about the dangers. However this could be further investigated.

The Poisson model with dependent variable amount of failed subjects has as significant variables the household head education, if the student has socio-economic scholarship and the amount of majors the student has.

As it can be noticed, both models find significant variables in common; but because theoretical variables should be incorporated from the bibliography, it is recommended to use the first model to measure the performance.

The university can examine which variables can be modified to help the students to have a better performance; for instance, provide support to students that have thought of leaving the university.

VI. Bibliography

Alcover, *et al*. (2008), *Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos*. Universidad de Politécnica de Valencia, España.

Blanco Bosco (2008). Factores escolares asociados a los aprendizajes en la educación primaria mexicana. *Revista electrónica Iberoamericana sobre Calidad, Eficiencia y Cambio en Educación*, 6 (1).

Blanco Vega, *et al*. (2011), *Análisis psicométrico de la escala Autoeficacia en conductas académicas en universitarios de primer ingreso*. Costa Rica.

Sotelo, *et al* (2009). Relaciones entre variables motivacionales y rendimiento académico en estudiantes universitarios. *X Congreso Nacional de Investigación Educativa*, México.

Tinto; (1989 cited in Romo, et al, 2000, chapter 4-p.1). *Deserción, Rezago y Eficiencia Terminar en la IES, Asociación Nacional de Universidades de Educación superior*, México D.F.

Vélez, Schiefelbein & Valenzuela (2001). *Factores que afectan el rendimiento académico en la educación primaria***.**