Statistical Studyfor Factors Affecting the StudentsPerformance of the School of Administration and Economic atSulaimani University

Renas Abubaker Ahmed [1] &Shahla Othman Salah Qadr [2]
Huda Muhamed Saied [3]

Statistic Department-School of Administration & Economics–Sulaimani [1,2,3]
University, Sulaymanyah City, Kurdistan Region – Iraq

Abstract:

Student's issue is one of the special concern in education system, one way to reduce these problems is to identify the most probable contributing factors that effect on student's agree or disagree for their department. So that the main objective of the study is the modeling for acceptance level from students according to mathematical models to determining and investigating the various factors which are cause this acceptance. This study applied logistic regression, the data collected from Sulaimani University according to survey method which shown in the study appendix, the sample consists of a 180 subjects. The acceptance level is aresponse variable which is a binary variable with two categories: the students are agree with their department or disagree. Due to the binary nature of this response variable a logistic regression approach is suitable to analyze in this case. Among 27 variables obtained from survey, four explanatory variables found most significantly associated to response variable which are X11 (Have students long disease), X17 (Did students expect to study at their department), X18 (Are students like to continue in their department), X27 (Are the lectures appropriate with reality life).

پوختەی:

مەسەلەی خوێندکار، یەکێکە لە سەرقاڵییە تایبەتەکانی سیستەمی خوێندن، تاکە ڕێگە بۆ سنوردانان بۆ ئەو گرفتانە، بریتییە لە دیاریکردنی ئەو هۆکارانەی زۆرترین ئەگەری بەشدارییان هەیە لە کاریگەری دانان لەسەر ڕازی بوون یان ڕازینەبونی خوێندکاران بۆ ئەوەی ئیدارە بدرێن، بۆیە ئامانجی سەرەکی لە توێژینەوەکە، بریتییە لە کلێشەڕێژکردنی رئاستی ڕازیبوون لە خوێندکاران، هەڵبەت ئەمەش بەپێی چەند کلێشەیەکی ماتماتیکی بۆ لیکۆڵینەوە لەو هۆکارە جیاوازانەی دەبنە هۆی قبوڵکردن(ڕازیبوون). لەم توێژینەوەیەدا (لێژبوونەوەی لۆژیستیك =الإنحدار اللوجيستي) بەکارهاتووە، ئەو داتایانەی لە زانکۆی سلێمانیدا بە ڕێگاکانی توێژینەوەی باسکرا و لە پاشکۆی توێژینەوەکەدا کۆکراونەتەوە، نمونەی توێژینەوەکش لە ١٨٠ بابەت پێك دێت. ئاستی ڕازیبوون بریتییە لە دووەمین گۆڕاوە، لەگەڵ دوو بەشدا: ئایا خوێندکاران پێك دێن و ڕازین لە بەشەکانیان یان نا. بە ڕەچاوکردنی سروشتی دوانەیی ئەم گۆڕاوە، ڕوون بووەوە، کە ڕەچاوکردنی پڕۆگرامی لێژبوونەوەی لۆژیستیك گونجاوە بۆ شیکردنەوە لەم دۆخەدا. لەکۆی ٢٧ گۆڕاودا کە لە ڕووپێوییەکەدا دەست کەوتوون، ٤ گۆڕاوی ڕاڤەکار دۆزرانەوە، کە بەشێوەیەکی مەزن پێوەندییان بە گۆڕاوی وەڵامدانەوەوە هەیە، کە بریتییە لە (X11)، ئەویش بریتیە لە : ئایا خوێندکار نەخۆشی درێژخایەنی هەیە؟، <u>X17</u>: ئایا پێشبینی دەکرێ خوێندکارەکان لە بەشەکانیان بخوێنن؟، <u>X18</u>: ئایا پێشبینی دەکرێ خوێندکارەکان درێژە بەخوێندن بدەن لە بەشەکانیان؟، <u>X27</u>: ئایا وانەکان لەگەڵ ژینوار دا دەگونجێت؟ (ژینوار = واقع الحياە).

الملخص:

إن قضية الطالب هي واحدة من الشواغل الخاصة في نظام التعليم، طريقة واحدة للحد من هذه المشاكل هي تحديد العوامل المساهمة الأكثر احتمالا التي تؤثر على موافقة الطالب أو عدم موافقته على إدارتهم. لذلك فإن الهدف الرئيسي من الدراسة هو نمذجة لمستوى القبول من الطلاب وفقا لنماذج رياضية للتحقق من العوامل المختلفة التي تسبب هذا القبول. طبقت في هذه الدراسة الانحدار اللوجستي، البيانات التي تم جمعها من جامعة السليمانية وفقا لطريقة المسح المبينة في ملحق الدراسة، وتتكون العينة من ١٨٠ موضوعا. مستوى القبول هو المتغير الثنائي مع فئتين: الطلاب يتفقون مع قسمهم أو لا يتفقون. نظرا للطبيعة الثنائية لهذا المتغير تبين أن استجابة نهج الانحدار اللوجستي مناسبة للتحليل في هذه الحالة. من بين ٢٧ متغيرا تم الحصول عليها من المسح، وجدت أربعة متغيرات تفسيرية مرتبطة بشكل كبير لمتغير الاستجابة التي هي X11 (هل الطالب لديه مرض مزمن)، X17 (هل يتوقع الطلاب دراسة في قسمهم)، X18 (هل يرغب الطلاب في الاستمرار في قسمهم) ، X27 (هل المحاضرات مناسبة مع واقع الحياة).

1. Introduction

Logistic regression is a statistical technique for examining relationships between an outcome variable also can be called a dependent, and one or more other variables, often calledindependent variables or explanatory. Many issues in the social sciences are studied using binary logistic models. These models are useful when there are several explanatory variables and a dichotomousresponse variable. For example, given a prospective student's age, department, and other similar information, one may wish to predict the probability of that students responding to agree or disagree for their department. Multinomial logistic models consider situations where there are multiple possible discrete outcomes,then the aim of this study is to determine and investigate of the factors that behind a student's agree or disagree for their department in the higher education system, and develop a logistic regression model for this purpose. Logistic regression was used in this study to estimate the effect of the statically significant factors on student's decision.

2. Materials and Methods

2.1 Logistic regression

Logistic regression is one of the important part in statistical models category that called generalized linear model. This part includes ordinary regression and ANOVA. Logistic regression allows to predict a binary outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the response variable is binary (dichotomous), such as presence and absence or success and failure. Logistic regression does not make many of the key assumptions of linear regression and general linear models that the based on (OLS) algorithms-particularly regarding linearity, normality, homoscedasticity, and measurement level.

2.2 The Model of Logistic Regression:

The response variable in the case of logistic regression is usually binary, that is mean, when this variable can take the value 1 with a success probability θ, or can take the value 0 with failure probability 1-θ,in this situation the variable is called a Bernoulli (or binary) variable. Where the response variable has more than two casesthis case is called multinomial. We define the binary random variable as:

$$\mathrm{Z} = \begin{cases} 1 \ \text{ if the outcome is a success} \\ 0 \ \text{ if the outcome is a failure} \end{cases}$$

With probabilities $Pr(\mathrm{Z}=1)=\theta$ and $Pr(\mathrm{Z}=0)=1-\theta$ if there are n such random variable $\mathrm{Z}_1 \sim \mathrm{Z}_\mathrm{n}$ which are independent with $Pr(\mathrm{Z}_\mathrm{j}=1)=\theta_\mathrm{j}$ then their joint Pr.is

$$\theta_{j=1}^{n}\theta_{j}^{\mathrm{Z_j}}(1-\theta_\mathrm{j})^{1-\mathrm{Z_j}} = exp\left[\sum_{j=1}^{n}\mathrm{Z_j}\,log\left(\frac{\theta_\mathrm{j}}{n\theta_\mathrm{j}}\right) + \sum_{j=1}^{n}\log(1-\theta_\mathrm{j})\right] \text{......(2.1)}$$

Which is a member of the exponential family for the case where theθ_j is are all equal, we can define:

$$Y = \sum_{j=1}^{n}\mathrm{Z_j} \qquad \text{.............................(2.2)}$$

So that Y is the number of successes in n trails.

The random variable Y has the distribution binomial (n,θ)

$$P_r(Y=y) = C_y^n\theta^y(1-\theta)^{n-y} \qquad ,y=0,1,2,\ldots\ldots,n \qquad \text{.....................(2.3)}$$

If $Y_i \sim binomail\ (n_i, \theta_i)$ the log. Likelihood function is :

$$L(\theta_1, \theta_2, \dots, \theta_N; y_1, y_2, \dots, y_N) \quad \dots\dots\dots\dots\dots\dots\dots.(2.4)$$

$$= \left[\sum_{j=1}^{N} y_i \, log\ \left(\frac{\theta_i}{n\theta_i}\right) + n_i \log(1-\theta_i) + log\ C_{y_i}^{n_i}\right] \quad \dots\dots\dots\dots\dots\dots\dots.(2.5)$$

The tolerance distribution is: $\theta = \frac{\exp(\beta_1+\beta_2 X)}{1+\exp(\beta_1+\beta_2 X)}$

This gives the link function:

$$\log\left(\frac{\theta}{1-\theta}\right)(\beta_1+\beta_2 X) \quad \dots\dots\dots\dots\dots\dots\dots.(2.6)$$

2.3 Logistic Regression assumptions

The explanatory or predictor variables can take any form in logistic regression, means that, there is no assumption about the distribution of the explanatory variables in logistic regression. Or can be say they have not normal distribution, linear relationship or of assumption about hetroscedastisity for equal variance in each group.Also in logistic regression there is no linear function relationship between the explanatory and response variables, instead, the logit transformation of θ is used in logistic regression:

$$\theta = \frac{e^{(\alpha+\beta_1 X_1+\beta_1 X_1+\cdots+\beta_1 X_1)}}{1+e^{(\alpha+\beta_1 X_1+\beta_1 X_1+\cdots+\beta_1 X_1)}} \quad \dots\dots\dots\dots\dots\dots\dots.(2.7)$$

Where α = the constant of the equation and, β = the coefficient of the predictor variables. the logistic regression alternative form is:

$$\mathbf{logit}\left[\theta(\mathbf{x})\right] = \mathbf{log}\left[\frac{\theta(x)}{1-\theta(x)}\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i \quad \dots\dots\dots\dots\dots\dots\dots.(2.8)$$

The goal of use of logistic regression is to correct predict for category outcome. To obtain this goal, the model is created that includes all explanatory variables that are useful in predicting the outcome variable. There are several different options for creating this model. The variables can be entered to the model which specified by researcher, or logistic regression can test the fit of the model after each coefficient of these variables are added or deleted,this is called stepwise regression.

2.3 Hosmer-Lemshow Goodness of Fit Test:

The Hosmer-Lemshow statistic evaluates the goodness-of-fit by creating 10 ordered groups of subjects and then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted). Thus, the test statistic is a chi-square statistic with a desirable outcome of non-significance, that is mean the prediction model doesn't significantly differ from the observed. These 10 groups are created which based on their probability estimated; those with estimated probability less than 0.1 form one group, and so on, up to those with probability 0.9 to one. Each of these categories is further divided into two groups based on the actual observed outcome variable (success, failure). The expected frequencies for each of the cells are obtained from the model. If the model is good, then the subjects most success are classified in the higher deciles of risk and those with failure in the lower deciles of risk.

2.4 The Logit (Logged Odds)

The logarithm transformation of probability P and even reaching to the logarithm of Odds $\langle \ln \text{Odds} \rangle$ called (Logit Transformation). Taking the natural log of the odds eliminates the floor of lower limit (zero) much as transforming probabilities into odds eliminates the ceiling of one. Based on this, taking the natural log of odds (Logit) above zero, and below one produces negative numbers, odds equal to one produces zero, and odds above one produces positive numbers. Therefore, the logistic regression indicates to the regression models that include logit as variable in the equation, and the general formula of the model are:

$$\ln \text{Odds} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + e \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.9)$$

Based on that, the relationship between the Logit (Log Odds) and explanatory variables is a linear relationship.

$$\log \text{Odds} = \text{logit}(\theta) = b_0 + \sum_{i=1}^{j} b_i x_i \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.10)$$

2.5 Estimate logistic regression coefficients:

It is wrong to use the ordinary least squares to estimate the parameters of regression in the case of regression models with binary (dichotomous) response variable value. The dichotomous response variable makesestimation using ordinary least squares inappropriate, the maximum likelihood (ML) estimation is oneof several alternative approaches that statisticianshave developed for estimating the parametersin a mathematical model.

Noted that the maximum likelihood (ML) is suitable for all linear and non-linear unlike the least squares method that does not fit except the linear models. However, themaximum likelihood (ML) method can be applied in the estimationof complex nonlinear as well as linear models.In particular,maximum likelihood (ML) estimation is the preferredestimation method for logistic regression because logistic model is nonlinear.Maximum Likelihood method is the Iterative method which depends on many Iterate of themathematical operations,until access to the best estimates of the coefficients, can be interpret the data observation through maximum likelihood method.

To describe themaximum likelihood (ML) method, can be introduce likelihood function (L). This is a function of theunknown parameters in one's model and, thus,can alternatively be denoted as $L(\theta)$, where θ denotes the collection of unknown parametersbeing estimated in the model, denoted here asθ_1, θ_2, up through θ_q, where q is the number ofindividual components.The procedure begins with an expression forthe likelihood of observing the pattern of occurrences (Y = 1) andNonoccurrences (Y = 0) of an event or characteristic in a given sample. This expression, termed the likelihood function, depends on unknown logistic regression parameters, maximum likelihood estimation finds the model parameters that give the maximum value for the likelihood function. The probability of that class was either P, if yi = 1 or 1 – p, if yi = 0 the likelihood is then:

$$L(\beta_0 + \beta) = \pi_{i=1}^{n} P(x_i)^{y_i} [1 - P(x_i)]^{1-y_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.11)$$

$$L(\beta_0 + \beta) = \sum_{i=1}^{n} y_i \, log \, P(x_i) + (1 - y_i) \log 1 - P(x_i)$$

$$= \sum_{i=1}^{n} \log 1 - \mathrm{P}(\mathrm{x_i}) + \sum_{i=1}^{n} \mathrm{y_i} \log \frac{\mathrm{P}(\mathrm{x_i})}{1-\mathrm{P}(\mathrm{x_i})}$$

$$= \sum_{i=1}^{n} - \log 1 + e^{(\beta_0+\beta \mathrm{x_i})} + \mathrm{y_i} \sum_{i=1}^{n} \mathrm{y_i}\, (\beta_0 + \beta \mathrm{X_i})$$

$$\frac{\partial L}{\partial \beta_j} = -\sum_{i=1}^{n} \mathrm{y_i} \frac{1}{1+e^{(\beta_0+\beta \mathrm{x_i})}} e^{(\beta_0+\beta \mathrm{x_i})} \mathrm{X_{ij}} + \sum_{i=1}^{n} \mathrm{y_i}\ \mathrm{X_{ij}}$$

$$= \sum_{i=1}^{n} (\mathrm{y_i} - \mathrm{P}(\mathrm{x_i}; \beta_0 + \beta)\, \mathrm{X_{ij}} \qquad \text{……………………..….(2.12)}$$

The maximum likelihood function in logistic regression is:

$$\mathrm{L} = \prod_{\mathrm{l}=1}^{\mathrm{m_1}} \mathrm{P}(\mathrm{x}_i) \prod_{\mathrm{l}=\mathrm{m_1}+1}^{\mathrm{n}} [1 - \mathrm{P}(\mathrm{x_i})] \qquad \text{…………………..….(2.13)}$$

Cases which have the status   Cases which do not have the status

Where n is the total number of cases, m_1 is the number of cases which have the status and Π indicates to the multiplication and is similar to the sum sign Σ, which means that the function is multiplication probability values for each case.We can use these products (previous equation) by assuming that we have independent observations on all subjects. The probability of obtaining the data for the 1th case when (y = 1) is given by P(Xi), where P(X) is the logistic model formula for individual X. The probability of the data for the 1th non-case where (y = 0) is given by 1 – P(Xi).

$$\mathrm{P}(\mathrm{x_i}) = \text{Logistic model}$$

$$\therefore \mathrm{P}(\mathrm{x_i}) = \frac{1}{1+\mathrm{e}^{-(\mathrm{a}+\sum \beta_\mathrm{i} \mathrm{x_i})}} \qquad \text{……………………..….(2.14)}$$

$$\therefore \mathrm{P}(\mathrm{x_i}) = \frac{\mathrm{e}^{(\mathrm{a}+\sum \beta_\mathrm{i} \mathrm{x_i})}}{1+\mathrm{e}^{(\mathrm{a}+\sum \beta_\mathrm{i} \mathrm{x_i})}}$$

$$\therefore \mathrm{L} = \frac{\prod_{\mathrm{l}=1}^{\mathrm{n}} \exp(\mathrm{a}+\sum_{\mathrm{i}=1}^{\mathrm{k}} \beta_\mathrm{i} \mathrm{x_{il}})}{\prod_{\mathrm{l}=1}^{\mathrm{n}} [1+\exp(\mathrm{a}+\sum_{\mathrm{i}=1}^{\mathrm{k}} \beta_\mathrm{i} \mathrm{x_{il}})]} \qquad \text{……………………..….(2.15)}$$

to avoid multiplication of probabilities (and typically having to deal with exceedingly small numbers).the likelihood function can be converted into a logged likelihood function. Since

$$\ln(\mathrm{X} * \mathrm{Y}) = \ln\mathrm{X} + \ln\mathrm{Y}$$

$$\ln(\mathrm{X^z}) = \mathrm{Z} * \ln\mathrm{X}$$

Taking the natural log of both sides of the likelihood equation gives the log likelihood function:

$$\mathrm{LL} = \ln\mathrm{L} = \sum [\mathrm{Y_i} * \ln\mathrm{P_i} + (1-\mathrm{Y_i}) * \ln(1-\mathrm{P_i})] \qquad \text{……………………..….(2.16)}$$

The log likelihood function sums the formerly multiplicative terms. If the likelihood function varies between zero and one, the log likelihood function is vary from negative infinity to zero.

2.6 Wald Test:

A Wald test is used to test the statistical significance of each coefficient (β) in the model. A Wald test is calculating a statistic Z, as follows:

$$Z = \frac{\hat{\beta}}{SE} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(2.17)$$

This z value is then squared, a Wald statistic is yielding with a distribution of chi-square. However, several authors have identified problems with the use of the Wald statistic. Warns that for large coefficients, and standard error is inflated, if the Wald statistic is lowering (chi-square) value.States that the likelihood-ratio test is more reliable for small sample sizes than the Wald test.

2.7 Likelihood-Ratio Test:

The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the full model (L1) over the maximized value of the likelihood function for the simpler model (L0). The likelihood-ratio test statistic equals:

……………………..….(2.18)

This log transformation of the likelihood functions yields a chi-squared. This test is recommended to use when building a model through backward stepwise elimination.

$$-2\log\left(\frac{L_0}{L_1}\right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1)$$

3. Data Analysis and Results

3.1 Data Description:

The data set used in this research consists of a sample of 180 students and was obtained fromSulaimani University / Administration and Economic School – Kurdistan Reign / Iraq, by using survey. Only students were examined who had been in a fourth stage for academic year 2013-2014. Since the goal of the study was to investigate the factors which might effecton the students during their education, the description of the survey variables are stated as below:

3.2 Variables of the study

3.2.1 Response variable:

The response variable is the student acceptance level thatis a binary variable with two categories, which are students are agree with their department or disagree.

3.2.2 Explanatory variables:

For the explanatory variables (independent variables), some of variables are nominal. Since some of the nominal variables have several levels and there should be an identifier (number: 1, 2, 3,……., n), these variables are:

X1: Student’s department.
X2: Student’s Gender.
X3: Student’s Age.
X4: Student’s Number of siblings.
X5:Student’s sequence in their family.
X6: Student’s family income.
X7: Average baccalaureate degree of student.
X8: Student’s housing type.
X9: Does student accept their department.

X10: Does studenthave a job besides education.
X11: Does studenthave a long-term disease.
X12: Are student’s father survival in life.
X13: Are student’s mother survival in life.
X14: Arestudent’s parent divorced?
X15:Do any members of the student’s family havea long-term disease.
X16:Are student’s family sexist.
X17:Did studentsexpect to study at their department.
X18:would students like to continue in their department.
X19:Arethe lectures time suitable for students.
X20:Do the student’s lecturers have theability to apply their lectures.
X21: Could students contact with their lecturers easily.
X22:Are student’s lecturers punctual.
X23: Are student’s lecturers respectful towards their students.
X24: Are students benefiting from lectures.
X25: Do students enterthe class on time.
X26: Are student’s references available in the college library.
X27: Are the lectures appropriate with real life.

3.3Application Part:

To test null hypothesis, which provides that the effect of all the factors of logistic regression model is equal to zero - the researcher tested it, where the results of the chi square test for the significance of the difference in the values of the logarithm likelihood function to logistic regression model by independent variables and no independent variables described as follows:

Table (3-1): test statistics significant to the model as all

	Chi-square	df	Sig.
Model	96.738	19	.000

It is clear from the table (3-1) that the chi square value is equal to (96.738) at the level of statistically equal to (0.000) , this means that the model which includes the explanatory variables explain classification of the acceptance level of students, and also predicts to it better than the model that does not include those explanatory variables, and therefore, the null hypothesis which provides that all factors logistic regression model is equal to zero is rejected, and that the test outputs prove that there is at least one factor of the coefficients of explanatory variables included in the model is not equal to zero, that is means, that there is at least one variable from the explanatory variables included in the model has the contribution , impact and importance in the classification of the acceptance level of students.

There are two different models according to logarithm likelihood function (-2LL); the first model is a model which included a constant only, where the value of the logarithm likelihood function (-2LL) for this model is equal to (249,533) that symbolized by (Do). But after insert explanatory variables under study the value of the logarithm likelihood function has become (152.795) which symbolized is (Dm).Since the difference between (Do) and (Dm) is equal to new method which is called (Gm) that has the chi square distribution,and can be calculated as follows: -

$$G_M = D_O - D_M = 249.533 - 152.795 = 96.738 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

It is the same value of chi square, which appeared in the table (4) previously.

Also there are many indicators (pseudo R square) interview, the first of them is the (RL square) which is called (Mc Fadden R square) and is calculated as follows:

$$R_L^2 = \frac{G_M}{D_O} = \frac{(D_O - D_M)}{D_O} = \frac{96.738}{249.533} = 0.388 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.2)$$

This means that, the logistic regression model which was containing the explanatory variables that is contributes by (38.8 %) to reduce the logarithm likelihood function value for the model which includes a constant value only without any independent variables.The second is the statistic (Rc square) is calculated as follows: -

$$R_C^2 = \frac{G_M}{G_M + N} = \frac{96.738}{96.738 + 180} = 0.350 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.3)$$

Since this measure has no direct meaningful and no clear in the percentage of reduction logarithm likelihood function, such as in the case statistic (RL square), with that interested by such an interpretation nearest to the concept of reduce sum squares residuals is known in the linear regression analysis. Another type of pseudo R square is (Rm square) calculated as follows: -

$$R_M^2 = 1 - \left(\frac{L_O}{L_M}\right)^{2/N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.4)$$

Note that the formula (Rm square) can't take one, therefore the adjusting this formula which symbolized by the (Rn square) and allow to take one is equal to dividing the (Rm square) on the largest value can take statistic (Rm square) as below: -

$$R_N^2 = \frac{R_M^2}{1 - (L_O)^{2/N}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.5)$$

It is noted that the values of statistics (Rm square) and (Rn square) previously yield in the output of statistical package SPSS under the name Cox & Snell R Square and Nagelkerke R Square respectively:

Table (3-2): practice significant measurement for logistic regression model

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	152.795	.416	.554

To test the second null hypothesis which provided that there was no statistically significant differences between the observation values and expected values that was fitted model, the (Hosmer and Lemeshow test) is used for this test. In this test, the probability value is calculated from the distribution Chi square to test the fitting logistic model.

Table (3-3):Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.078	8	.929

Indicate table (3-3) to the value of chi square test for (Hosmer and Lemeshow) equals to (3.078) when the degree of freedom equals to(8) and level of significance equals (0.929), so the value of chi square test is not statistically significant, this indicates the inability to reject null hypothesis above that the expected data model user identify the data observation.

3.4 Classifications tables: -

The following table shows outputs of the analysis specifications at the cut value 0.5

Table (3-4): classification table for logistic regression model

Observed		Predicted		
		Acceptance Level Variable		Percentage Correct
		Disagree	Agree	
Acceptance Level Variable	Disagree	26	25	51,0
	Agree	15	114	88,4
Overall Percentage				77,8

a. The cut value is ,500

Is clear that the sensitivity of the model, which is a percentage of correct predictions in the group disagreecan be calculated as follows: $\mathrm{SE} = 51\%$. And the Specifity model, which is a percentage of correct predictions in the group agree is calculated as follows : $\mathrm{SP} = 88.4\%$. In general, the percentage of correct classification (Hit rate), which is equal to the number of correct predictions on the total number of the sample study was calculated as follows: $\text{Hit Ratio} = 77.8$.

3.5 Interpretation of logistic regression coefficients:

The researcher can be interpret of coefficient logistic regression as follows:

Table (3-5): Wald statistic for variables in the model

	B	S.E.	Wald	df	Sig.	Exp(B)	95,0% C.I.for EXP(B)	
							Lower	Upper
X27	-1,062	,406	6,839	1	,009	,346	,156	,766
X18	-1,059	,272	15,127	1	,000	,347	,203	,591
X17	-1,397	,382	13,370	1	,000	,247	,117	,523
X10	1,744	1,082	2,596	1	,107	5,719	,686	47,706
X11	1,202	,654	3,379	1	,066	3,328	,923	11,996
X2	,536	,532	1,017	1	,313	1,710	,603	4,851
X3	-,123	,172	,516	1	,473	,884	,631	1,238

X4	-,132	,106	1,564	1	,211	,876	,712	1,078
X6	,334	,320	1,093	1	,296	1,397	,746	2,616
X7	,056	,058	,918	1	,338	1,058	,943	1,186
X8	,405	,607	,444	1	,505	1,499	,456	4,928
X9	,248	,614	,164	1	,686	1,282	,385	4,268
X12	-1,048	,945	1,230	1	,267	,351	,055	2,234
X13	1,695	1,052	2,597	1	,107	5,444	,693	42,754
X15	,107	,910	,014	1	,907	1,112	,187	6,614
X20	-,633	,497	1,621	1	,203	,531	,200	1,407
X24	,035	,409	,007	1	,932	1,035	,465	2,306
X25	,083	,358	,053	1	,818	1,086	,539	2,190
X26	-,348	,376	,857	1	,355	,706	,338	1,475

Table (3-5) is explain the coefficient of all explanatory variable in the study with their Wald text and Exp(B) for all of them, but after removing non- significant explanatory variables in the model, can be interpret the logistic regression parameters for remaining variables in the model as following table (3-6):

Table (3-6): Wald statistic for variables in the model

	B	S.E.	Wald	Df	Sig.	Exp(B)
X27	-,626	,315	3,957	1	,047	,535
X18	-,844	,238	12,584	1	,000	,430
X17	-1,090	,307	12,574	1	,000	,336
X11	,992	,540	3,374	1	,066	2,698
X10	2,943	,606	23,614	1	,000	18,972

From table (3-6) it is clear, where the variable X27 increase by one degree will lead to the decreasing the Odds by (46.5%), the X27 variable parameter value is equal to (-.626), means that the Logit value is increasing by (.626) when the degrees of the X27 variable decreased by one degree, after controlling the impact of theother explanatory variables. The value of the Wald test for variable X27 at degree of freedom equal to one is (3.957) and by statistical significant level equal to (0.047), this means that the variable X27 has predictive ability for classifying of the acceptance level of students.

Sothen an increase in the variable X18 by one degree will lead to the decreasing the Odds by (57%). And the X18 variable parameter value is equal to (-0.844), and this means that the value of Logit decreasing by (0.844) where the degrees of the X18 variable increased by one degree. The value of the Wald test for this variable at degree of freedom equal to one is (12.584) and by statistical significance level is equal to (0.000) , this means that the logistic regression coefficient value for variable X18 differ from zero. also this variable has significant statistical of the predictive ability for classifying of the acceptance level of students.

The next variable is X17 the parameter value of this variable is equal to -1.09, and the value of the Wald test for variable X17 at degree of freedom equal to one is (12.574) and by statistical significant level equal to (0.000), this means that this variable also has statistical significant of the predictive ability for classifying of the acceptance level of students. Then an increase in the variable X17 by one degree will lead to the decreasing the Odds by (66.4%).

And the effect of X11 variable was very small so that that parameter value for this variable it was small too, and the Wald test for X11 variable was also very small, this means that this variable has a poor significant statistical of the predictive ability on classifying of the acceptance level of students.

Finally the X10 variable parameter value is equal to 2.943, means that the Logit value is increasing by (0.94) when the degrees of the drinking variable increased by one degree, the value of the Wald test for this variable at degree of freedom equal to one is (23.614) at the level equal to (0.000), this means that this variable has a strong statistically significant of the predictive ability on classifying of the acceptance level of students.
Therefore the acceptance level models will be:

$$\text{Logit} = -.626\,\text{X27} - .844\,\text{X18} - 1.09\,\text{X17} + .992\,\text{X11} + 2.943\,\text{X10} \quad \ldots\ldots(3.6)$$

Where:-

X27: Are the lectures appropriate with reality life.
X18: Are students like to continue in their department.
X17: Did students expect to study at their department.
X11: Have students long disease.
X10: Have students another job except education.

4.1Discussion of Results

The intent was to provide a demonstration of a model that can be used to assess the most important contributing factors for classifying of the acceptance level according to students idea incollege of administration and economic -Sulaimani university.The results presented in this study show that the model provided a reasonable statistical fit, twenty sevenvariables were used as explanatory variables in the development process. Using the concept of deviance together with wald statistic, the study variables were subjected to statistical testing. Only four variables were included in the model, namely, X11, X17, X18, X27. That is indicate to the long disease of students, students expect to receive their department, future study at the same department of student, and appropriate the lectures with reality life are the most important factors behind acceptance level of students at higher education system.

4.2 Recommendations

To solve the problem of acceptance system, it is recommended to change the system to the system does not depend on the central acceptance, therefore the government has to pay attention for the acceptance system and renovate them.The result of the study shows that the long disease of students, students expect to receive their department, future study at the same department of student, and appropriate the lectures with reality life have a great role to explain and determine the acceptance level from students, therefore the government has to pay attention for the appropriate the lectures with reality life with the other factors according to student's understood to improve acceptance system.

5. References

1- Agresti, Alan. "**An Introduction to Categorical Data Analysis**". John Wiley and Sons, Inc. 1996.

2- Al-Ghamdi, A. S. "**Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity**". Journal of Accident Analysis and Prevention (AAP), Vol. 34, No. 6, pp. 729-741, 2002.

3- Eliason, Scott R. "**Maximum Likelihood Estimation Logic and Practice**". Sage University Paper series on Quantitative Applications in the Social Sciences. No.07-096). Beverly Halls, CA: Sage, 1993.

4- Fan, Xitao and Wang, Lin. "**Comparing Linear Discriminant Function with Logistic Regression for the Two-Group Classification Problem**". Paper Presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

5- Horton, Nicholas J. and Laird, Nan M. "**Maximum Likelihood Analysis of Logistic Regression Models with Incomplete Covariate Data and Auxiliary Information**". Biometrics. Vol.57, pp.34-42, March 2001.

6- Hosmer, David W. & Lemeshow, Stanely, "**Applied Logistic Regression**". 2nd edition. New York: Johnson Wiley & Sons, Inc., 2000.

7- Houston, Walter M. & Woodruff, David J. "**Empirical Bayes Estimates of Parameters from the Logistic Regression Model**". ACT Research Report Series 97-6. American Coll. Testing Program, Iowa City, IA. 1997.

8- Kadhim, A.H. And Muslim, B.S. "**The Advance Econometrics Theory & Application**". Baghdad, 2002.

9- Kleinbaum, David & Klein, Mitchel, "**Logistic Regression: a Self-learning Text**". USA: Springer, 2002.

10- Lea, Stephen, "**Multivariate Analysis II: Manifest Variables Analysis. Topic 4: Logistic Regression and Discriminant Analysis**". University of EXETER, Department of Psychology. Revised 11th March, 1997. Available at: http://people.exeter.ac.uk/SEGLea/multvar2/disclogi.html

11- Menard, Scott, "**Applied Logistic Regression Analysis**". Sage University Paper series on Quantitative Applications in the Social Sciences. No.07-106). Beverly Halls, CA: Sage, 1995.

12- Newsom. **Data Analysis II: Logistic Regression**. Fall 2003.

13- Pample, Fred C. "**Logistic Regression A primer**". Sage University Paper series on Quantitative Applications in the Social Sciences. No.07-132). Beverly Halls, CA: Sage, 2000.

14- Tabachnick, Barbara and Linda Fidell. "**Using Multivariate Statistics**", Third edition. Harper Collins. 1996.

15- Walker, Jonathan, "**Methodology Application: Logistic Regression Using the Codes Data**". Developed For Department of Transportation, National Highway Traffic Safety Administration (NHTS), Washington DC, April 30, 1996.