Information Management and Business Review (ISSN 2220-3796)
Vol. 10, No. 2, pp. 17-22, June 2018

# Ordinal Logistic Regression for Students Academic Performance in Kurdistan Region of Iraq

[1]Nozad H. Mahmood, Sherko H. Murad[2], [3]Karwan Kakabra Kakamad
Kurdistan Technical Institute, Business Administration Department
Kurdistan Technical Institute, Computer Science Department
Soran University, Faculty of Arts, Psychology Department, Iraq
nozad.mahmood@kti.edu.krd, sherko.murad@kti.edu.krd, karwan.kakamad@soran.edu.iq

**Abstract:** The First Attempt Exam is the first final exam for students after their first academic year to show their academic performance. According to the outcomes of our project, some factors influence students' academic performance. These factors include (Sex, Car Ownership, Relationships, Smoking, Hours of Study, Facebook Use, Father Alive, Mother-Edu, and Father-Edu). These factors have a high effect on the results of the First Attempts Exam. Furthermore, females are more likely to pass the first attempts exam than the males. Also, students with cars, relationships, who smoking, use Facebook regularly, and whose mothers have higherlevels of education are negatively affected by the student's academic performance. On the other hand, mother alive, higher level education of father, and increasing one-hour daily study are positively affected on the students' academic performance.

**Keywords:** *Logistic Regression Model, Odd Ratio, ROC (Receiver Operating Characteristic) curve, Area Under the Curve(AUC), Students Performance*

## 1. Introduction

For several decades' institutions of higher education across the globe have experienced significant growth in both their student populations and the diversity of those students. As a result, many studies have been conducted into trying to gauge and understand the factors that contribute to student's academic performance. These predictive types of studies have become more common and important to overall planning for universities worldwide including in such areas as curriculum development and student placement. This literature review has looked at studies in such places as Australia, the United States, India, United Arab Emirates, Nigeria, Pakistan, and Iraq to identify recurring outcomes and methodologies shared by scholars working in this area. It is significant to observe that studies conducted as early as the 1980's and reported on in 2001 (McKenzie & Schweitzer, 2001) investigated such things as student's academic backgrounds and a broad litany of psychosocial factors contributing to future performance like socio-economic indicators, parents' educational levels, and gender. Most of the studies reviewed for this investigation also share many common indices used by these previous scholars. For example, (Baradwaj & Pal, 2011), (Mushtaq & Khan, 2012), (Ali et al., 2013), and (Asogwa & Oladugba, 2015) all used student socio-economic data, gender and age information, parental educational levels and a significant resource pool ranging from 100 students to over 400. Some like (McKenzie & Schweitzer, 2001) looked at several years of student's prior GPA scores and (Ali et al., 2013) investigated students actual residential/geographical backgrounds noting that place of residence and social status in India were key factors to student's future academic performance.

Other studies like (Huws, Reddy & Talcott, 2006) considered such things as student's motivational levels and reasons for pursuing higher education contradicting earlier findings that student's future success is largely dependent on preceding performance. Their findings showed that students learning or studying at the graduate level and the score secured did not have any predictive relevance to their achievements in graduate school. In the important work of (McKenzie & Schweitzer, 2001) the study established an important link between student's prior and current employment responsibilities and experience, as well as their overall outlook or "self-efficacy" in achieving above average grades and performance. This study established that student's "time management" skills are significant indices of future success prompting many universities to invest in training programs and skill building in this area. For the purposes of this study, one previous work conducted in Iraq (Gardy & Akbay, 2015) is of special note. The report "Academic Performance of Undergraduate Students at Soran University in Northern Iraq" provides an important template for cross-referencing and comparison of the present study. In the prior study, Gardy found that such factors as the educational level of the student's mother positively correlated with student's academic performance. Other factors such as the student's age, "class", employment status and significantly the size of their families played

an important part in predicting future academic performance. As a result of this work, the present study looked deeper into such issues as whether or not students own a car, are in a personal relationship, smoke, have a Facebook account and the number of hours they spend on social media. Consequently, the present state of investigations into these areas can continue to be informed by previous studies and important valuable comparisons can be made. For example, the present study found a negative correlation between students' academic performance and their mother's educational levels contradicting or questioning Gardy's earlier findings and prompting the need for further investigations. In general, the vast majority of the studies reviewed clearly show an important connection between a student's socio-economic status and their future academic performance. However, recent studies including this one, suggest that this once significant link may be changing. Perhaps, because of the revolutionary changes and access to affordable technology student's socio-economic circumstances and family status will not be the driving factor determining their future scholastic success.

## 2. Methodology

According to the regulations outlined by the Iraqi Higher Education system students who do not pass their final exams on the first attempt, have the ability or the right to retake the exam at a later date. In such a way, binary logistic regression model (pass in the first attempt, fail in the first attempt) is used to test the data and to illustrate the factors affecting student's performance in the University of Sulaimani. Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). In this project, logistic regression is used to model the probability of a positive outcome for a binary (fail in the first attempt = 0, pass in the first attempt = 1) outcome variable as a function of covariates, and the goal of logistic regression models is used to model the probability of the occurrence of an event depending on the value of covariates. In other words, the model is used to investigate the relationship between a binary response variable (First Attempt Exam) and a set of explanatory, or independent, variables (Sex, Age, Handwriting, Breakfast, Car, Phone types, Job, Relationship, Smoke, Alcohol, Elementary school types, Study-Hours, Sitting in the class, Dorm-Living, Sleeping-Hours, Nap-Hours, TV-Hours, Facebook, and Family variables such as Father Alive, Father Education Levels, Mother Alive, and Mother Education Levels). Furthermore, JMP software program tool was used to analyze the dataset and to explain the capacity of their impacting.

**Research area:** The University of Sulaimaniis the oldest university in Northern Iraq and is located in the city of Sulaymaniyah, which is the cultural capital city of the Kurdistan Region. The university's system of education in the Kurdistan Region is focused primarily on assisting the region students in their pursuit of higher education. The education system in Kurdistan is the same as in Iraq, and the undergraduate study is universally composed of four stages ($1^{st}$, $2^{nd}$, $3^{rd}$, and $4^{th}$) in which students need at least 4 years to complete the whole stages to get Bachelor degree except Medicine and Architectural Engineering colleges that need 6 years and 5 years respectively to get Bachelor degree.

**Data Collection:** The essential data for this study was collected by the use of questionnaires, and the simple random sampling was used to collect a sample size of 370 students from all the undergraduate students in the University of Sulaimani. The surveys were handed out to students in the ($2^{nd}$, $3^{rd}$, and $4^{th}$) years at the University of Sulaimani. The data gathered in this study was collected from all the colleges in the University of Sulaimani except the Medicine College because of the specialist nature of its students and conflicts with their study hours. The background information used in the study consists of 23 questions, and all the predictor and response variables derived from the database are given in Table 1 for reference.

**Table 1: Student Related Variables and Measures**

| Variables | | Codes |
|---|---|---|
| Response Variable | First Attempt Exam | $FirstAttemptExamPass = 1 FirstAttemptExamFail = 0$ |
| Predictor variables | Sex | $Female = 0, Male = 1$ |
| | Age | Categorical |
| | Hand-Writing | $Right = 1, Left = 0$ |
| | Breakfast | $HavingBreakfast = 1, Don'thaveBreakfast = 0$ |
| | Mobile type | $SmartPhone = 1, RegularPhone = 0$ |
| | Job | $HavingJob = 1, Don'thaveJob = 0$ |
| | Relationship | $MarriedorHavingRelationship = 1, Unmarried = 0$ |
| | Alcohol | $DrinkingAlcohol = 1, NoDrinkingAlcohol = 0$ |
| | School Type | $PrivateSchool = 1, GevernmontSchool = 0$ |
| | Place in Class | $SittingontheFirstTwoRows = 1, OtherRows = 0$ |
| | Dorms | $LiveintheDorm = 1, LivewithParents = 0$ |
| | Sleeping Hours | Categorical |
| | Nap | $TakeaNap = 1, Don'tTakeaNap = 0$ |
| | Watching TV Hours | Categorical |
| | Family Member | Categorical |
| | Mother Alive | $MotherAlive = 1, MotherDied = 0$ |
| | Father Alive | $FatherAlive = 1, FatherDied = 0$ |
| | Mother Education | $Lessthanhighschool = 1, Highschool = 2, Somecollege = 3, UniversityandHigher = 4$ |
| | Father Education | $Lessthanhighschool = 1, Highschool = 2$ $Somecollege = 3, UniversityandHigher = 4$ |
| | Smoking | $Smoking = 1, NoSmoking = 0$ |
| | Facebook | $HaveFacebookAccount = 1, Don'tHaveFacebook = 0$ |
| | Study Hours | Categorical |
| | Car | $HaveaCar = 1, Don'tHaveCar = 0$ |

## 3. Analysis and Results

The Iterations report in the Table (1) shows each iteration and the evaluated criteria that determine whether the model has converged. As the results, the model is converged gradient in 5 iterations.

**Table 2: Iterations of the converted model**

| Item | Objective | Relative Gradient | Norm Gradient |
|---|---|---|---|
| 0 | 256.46445681 | 9.3099272204 | 162.96855525 |
| 1 | 209.41037853 | 1.9344043844 | 12.523767342 |
| 2 | 207.39480515 | 0.2546886064 | 1.6015556473 |
| 3 | 207.36192362 | 0.005512953 | 0.078683959 |
| 4 | 207.36190842 | 3.0792292e-6 | 0.0000752681 |
| 5 | 207.36190842 | 1.109546e-12 | 3.644049e-11 |

While P-value < 0.001, shows in Table 3, the model is significant for the Chi-Squared test. In other words, full model is better than the reduced one (model with only intercept parameters), and it indicates that at least one of the parameters has an effect on the response variable (First Attempt Exam). On the other hand, A min (AIC) strategy is used for selecting among two or more competing models. In a general sense, the model for which $AIC_c$ is smallest represents the "best" with this formula $AIC_c = -2 \log likelihood + 2\ k + 2\ k\left(\frac{k+1}{n-k-1}\right)$ where k is the number of parameters in the model. Thus,

$AIC_c$ = 466.202 for full model

$AIC_c$= 563.877 for null model

As a result, the full model is the best one.

Shows the R-Square, which is the ratio of the Difference to the Reduced negative log-likelihood values. It is sometimes referred to as U, the uncertainty coefficient. R-Square ranges from zero for no improvement to 1 for a perfect fit. A Nominal model rarely has a high R-Square, and it has an R-Square of 1 only when all the probabilities of the events that occur are 1. In our case, R-square is 19.06 %

**Table 3: Whole the Model**

| Model | -Log-Likelihood | DF | Chi-square | Prob>ChiSq |
|---|---|---|---|---|
| Difference | 48.83762 | 23 | 97.67524 | <.0001* |
| Full | 207.36191 | | | |
| Reduced | 256.19953 | | | |
| | | | | |
| RSquare (U) | 0.1906 | | | |
| AICc | 466.202 | | | |
| BIC | 556.648 | | | |
| Observations | 370 | | | |

As a result of Table 4, Since Age, Hand-writing, Breakfast, Mobile Type, Job, Alcohol, School Type, Place in Class, Dorms, Sleeping, Nap, Watching-TV, Family Members, and Mom Alive show up as not significant, it indicates that these fourteen variables have not any effect on the First Attempt exam. Consequently, the regression equation is shown below

Log (odds of a First Attempt Exam) = 1.219 – 0.707 (Sex) – 1.009 (Relationship) + 1.329 (Father-Alive) – 0.505 (Mother-Edu) + 0.366 (Father-Edu)– 1.012 (Smoking) – 1.086 (Facebook) + 1.172 (Hours Study) – 1.130 (Car)

As with any regression, the positive coefficients indicate a positive relationship with the response variable (First Attempt Exam). Consequently, -0.707 is the increment to log odds of a worse outcome for males. In other words,

$$log(OR) = log\left(\frac{Oddsofmalespassthefirstattemptexam}{Oddsoffemalespassthefirstattemptexam}\right) = -0.707$$

$WhereORindicatestoOddRatio$.

Log odds are difficult to interpret on their own. However, it can be translated using the formulae described below in such a way Log-odds could be converted to normal odds using the exponential function

$$log(OR) = -0.707 \Rightarrow OR = e^{-0.707}$$

Hence, the odds ratio$e^{-0.707} = 0.493$ indicates that males are 0.493 times more likely to fail the First Attempts of the exam than females. Furthermore, for married students or students with having relationship the odds ratio $e^{-1.009} = 0.365$ indicates that students with having rea lationship are 0.365 times more likely to fail the first attempts of extheam. As well, the odd ratio $e^{1.329} = 3.777$ indicates that students with alive father are 3.777 more likely to pass the first attempts of exathem than those students who lost their fathers. In addition, the odd ratio $e^{-0.505} = 0.604$ indicates that students with one level higher of their mother's education are 0.604 times more likely to fail the exam. On the other hand, students with one level higher of dad's education are $e^{0.366} = 1.442$ times more likely to pass the first exam than students with lowera level of dad's education. Besides, the odd ratio $e^{-1.012} = 0.363$ indicates that students' smoker are 0.363 times more likely to fail the first attempts of exam tthehanthose smoking. Additionally, the odd ratio $e^{-1.086} = 0.338$ indicates that students who use Facebook 0.338 times more likely to fail the first attempts of exam ththean those do not use it. Further, the odd ratio $e^{1.172} = 3.288$ indicates that students with increasing one hour of study are 3.288 times more likely to pass the first attempts of exam. Lathe the stly, Car case, the odds ratio$e^{-1.130} = 0.323$ indicates that a student with having car is 0.aa 323 times more likely to fail the First Attempts of exam than the one with no having car.

**Table 4: Parameter Estimates**

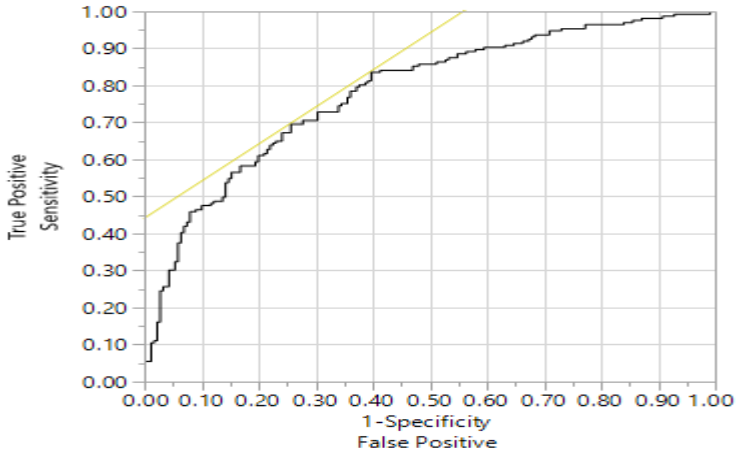
| Term | Estimate | Std Error | ChiSquare | Prob>ChiSq |
|---|---|---|---|---|
| Intercept | 1.21927692 | 2.1574109 | 0.32 | 0.5720 |
| Sex | -0.7068091 | 0.3360941 | 4.42 | 0.0355* |
| Age | -0.0089979 | 0.0718449 | 0.02 | 0.9003 |
| Hand-Writing | 0.31714648 | 0.4426766 | 0.51 | 0.4737 |
| Breakfast | 0.46665181 | 0.2706419 | 2.97 | 0.0847 |
| Mobile type | -0.1234657 | 0.3816041 | 0.10 | 0.7463 |
| Job | 0.67797553 | 0.3666101 | 3.42 | 0.0644 |
| Relationship | -1.0086692 | 0.4118749 | 6.00 | 0.0143* |
| Alcohol | 0.74493559 | 0.3951886 | 3.55 | 0.0594 |
| School Type | -0.3565235 | 0.4707394 | 0.57 | 0.4488 |
| Place in Class | -0.5063182 | 0.2718216 | 3.47 | 0.0625 |
| Dorms | -0.0244414 | 0.2988203 | 0.01 | 0.9348 |
| Sleeping | 0.03703869 | 0.0888991 | 0.17 | 0.6769 |
| Nap | 0.19725684 | 0.2603607 | 0.57 | 0.4487 |
| Watching TV | 0.02502406 | 0.06901 | 0.13 | 0.7169 |
| Family Member | -0.0447408 | 0.0709466 | 0.40 | 0.5283 |
| Mother Alive | -0.4459353 | 0.8935099 | 0.25 | 0.6177 |
| Father Alive | 1.32933275 | 0.5745795 | 5.35 | 0.0207* |
| Mother Education | -0.5047344 | 0.1558384 | 10.49 | 0.0012* |
| Father Education | 0.36603957 | 0.1414639 | 6.70 | 0.0097* |
| Smoking | -1.0120636 | 0.4140007 | 5.98 | 0.0145* |
| Facebook | -1.086195 | 0.4108137 | 6.99 | 0.0082* |
| Study Hours | 1.17204784 | 0.2811257 | 17.38 | <.0001* |
| Car | -1.1298474 | 0.2848938 | 15.73 | <.0001* |

Ultimately, plotting the pairs of the true positive rate (Sensitivity) versus the false positive rate (1-Specificity) on a scatter plot, shown in Figure 1, provides a ROC (Receiver Operating Characteristic) curve, and the area under the curve(AUC) clarifies an overall measure of fit of the model. As a general rule:

If $AUC < 0.5$ No discrimination
If $0.50 \leq AUC < 0.70$ Poor discrimination
If $0.70 \leq AUC < 0.8$ Acceptable discrimination
If $0.80 \leq AUC < 0.90$ Excellent discrimination
If $AUC \geq 0.90$ Outstanding discrimination

Using First Attempt = "1" to be the positive level, $AUC = 0.78447$
Comparing this value with the general rule, the model is considered acceptable discrimination.

**Figure 1: shows the plot between the sensitivity and (1 – specificity)**



## 4. Conclusion

The objective of this study is to illustrate the factors that affect students' academic performance in the Iraqi University of Sulaimani; According to the results, the factors of (Sex, Car, Relationship, Smoking, Hours Study, Facebook, Father-Alive, Mother-Education, and Father-Education) have a high effect on the First Attempts of Final Exam, in which for Sex, the females are0.493 times more likely to pass the first attempts exam than the males. Furthermore, students who have cars and who are in relationships, as well as students who smoke, use Facebook and whose mothers have higher levels of education are 0.323, 0.365, 0.363, 0.338 and 0.604 times more likely to fail the first attempts exam respectively. On the other hand, factors such as father alive, higher level education of father, and increasing one-hour daily study are positively significant to pass the first attempt of the finalexam with 3.777, 1.442, and 3.288 individually. This project depicted some fascinating and weird points that should be copied more seriously and carefully with the results. First of all, the results show that apart from the advantages and utility of technology and fortune, there are still some uselessness issues for someone who does not knowingly use them in a careful and right way, such that car ownership and using Facebooks. Next, the outcomes illustrated uncanny result while mothers with higher education levels, having a negative effect on their university students' test results. the conclusion regarding relationship between mother education levels and students might due to Iraqi father sovereignty, or might be because of educated mothers constantly have their own jobs as the men, and it makes them busier and reduces their times to care for children compared to uneducated mothers who always stay at home with greater conscious of children.

### Recommendations

- In order to find out whether the effective university model is applicable to all universities in Iraq and Kurdistan Region, on the other hand, to find other factors that influence academic performance, similar studies should be conducted in other universities of the country.
- It is necessary for Iraqi communities in general to preserve a warm and friendly relationship with their students irrespective of their gender differences.
- Parents should provide a great academic atmosphere for their students and improve the level of care of their children.

### References

Ali, S., Haider, Z., Munir, F., Khan, H. & Ahmed, A. (2013). Factors contributing to the student's academic performance: A case study of Islamia University Sub-Campus. *American Journal of Educational Research*, 1(8), 283-289.

Asogwa, O. C. & Oladugba, A. V. (2015). Of Students Academic Performance Rates Using Artificial Neural Networks (ANNs). *American Journal of Applied Mathematics and Statistics*, 3(4), 151-155.

Baradwaj, B. K. & Pal, S. (2011). Mining Educational Big Data to Analyze Students' Performance. *International Journal of Advanced Compute,* 2(3).

Gardy, I. H. S. & Akbay, C. (2015). Academic Performance of Undergraduate Students at Soran University in Northern Iraq.

Hijazi, S. T. & Naqvi, S. M. M. (2006). Factors Affecting Students' performance. *Bangladesh e-journal of Sociology*, 3(1).

Huws, N., Reddy, P. & Talcott, J. (2006). Predicting university success in psychology: are subject-specific skills important? *Psychology Learning & Teaching*, 5(2), 133-140.

McKenzie, K. & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first-year Australian university students. *Higher education research & development*, 20(1), 21-33.

Mushtaq, I. & Khan, S. N. (2012). Factors Affecting Students' Academic Performance. *Global Journal of Management and Business Research,* 12 (9), 17-22. Retrieved on October 2015.https://globaljournals.org/GJMBR_Volume12/3-Factors-Affecting-Students-Academic.