# USING ORDINAL REGRESSION MODELING TO EVALUATE THE SATISFACTION OF JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY FACULTY OF SCIENCE STUDENTS

***G. M. Ombui, M. Geofrey and A. W. Gichuhi***
*Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi*
*E-mail: monariff@gmail.com*

**Abstract**

General students' satisfaction of Jomo Kenyatta University of Agriculture and Technology (JKUAT) Faculty of Science students is associated with a combination of qualitative and quantitative predictor variables. Ordinal regression statistical technique was used to model the relationship between the academic programmes, facilities and services, and the outcome variable to determine the explanatory variables that influence students' satisfaction factors that will assist in improved service delivery. Data analysis involved both descriptive and inferential analysis. The factors that were found to influence the satisfaction of the JKUAT Faculty of Science students were four; service delivery at the department office, the library services, accommodation facilities in the university hostels and accommodation facilities outside the university.

**Key words**: Ordinal regression, clog-log, link function

## 1.0 Introduction and Literature Review

An end of semester questionnaire was administered each semester with twenty seven questions using a mixture of qualitative and quantitative variables measured on five point ordinal scale to rate lecturers' per unit. The sample questions reflected key issues that the Faculty of Science administrators may use to plan in time for quality services majorly from JKUAT staff to students.

Using a multilevel modeling technique to analyse survey data, a study by Porter(2001) examined the impact that different departments have on student satisfaction in a large research university. The research finding revealed that characteristics of departments such as size, faculty contact with students, research emphasis, and proportion of female students had a significant impact on education satisfaction within student's major subjects of study.

The mean responses of student satisfaction survey conducted by Cooney(2000) revealed community college student satisfaction. The survey respondents rated highest satisfaction on responsiveness to diverse populations, registration effectiveness, and academic services, while rating the lowest satisfaction on admissions and financial aid, academic advising, and campus support services.

## 2.0 Materials and Methods

### 2.1 Introduction

Generalised linear model is a very powerful class of models which can be used to provide solutions to a wide range of statistical questions. The basic form of a generalised linear model is shown in the equation.

$$\text{Link}\left(\gamma_{ij}\right)=\theta_j-\left[\beta_1 X_{i1}+\beta_2 X_{i2}+\cdots+\beta_k X_{ik}\right] \quad \text{..........} \quad (1)$$

where, link () is the link function, Link( $\gamma_{ij}$ ) is the cumulative probability for the j$^{th}$ category for the i$^{th}$ case , $\theta_j$ - is the threshold for the j$^{th}$ category, k- is the number of regression coefficients , $\beta_1, \beta_2, \cdots, \beta_k$ - are the regression coefficients, $X_{i1},\ X_{i2}, \cdots, X_{ik}$ - are values of the predictors for the i$^{th}$ case

Three major components in ordinal regression model are:

***Location Component***

The portion of the equation (1) that includes the coefficients and predictor variables is called the location component of the model. The location is the "meat" of the model. It uses the predictor variables to calculate predicted probabilities of membership in the categories for each case.

***Scale Component***

This is an optional modification to the basic model to account for differences in variability for different values of the predictor variables. For example, if certain groups have more variability than others in their ratings, a scale component is used to account for this improved the model. The model with a scale component follows the form shown in equation (2).

$$link(\gamma_j) = \frac{\theta_j - [\beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}]}{\exp(\tau_1 z_1 + \tau_2 z_2 + \cdots + \tau_m z_m)} \quad \text{......(2)}$$

where, $\tau_1, \tau_2, \cdots, \tau_m$ are coefficients for the scale component.

$z_1, z_2, \cdots, z_m$ are m predictor variables for the scale component.

***Link Function***

**I**t is a transformation of the cumulative probabilities that allows estimation of the model. It defines what goes to the left side of the equation. It is also the link between the random component on the left side of the equation and the systematic component on the right.

***Pseudo R-Square***

In ordinal regression models, these measures were based on likelihood ratios rather than raw residuals. There are several measures intended to mimic the R-squared analysis, but none of them are an R-squared. The interpretation is not the same, but they can be interpreted as an approximate variance in the outcome. Three different methods will used to estimate the coefficient of determination. McFadden's, the ratio of the likelihoods suggests the level of improvement over the intercept model that is model without predictors offered by the full model that is the model with predictors. A likelihood falls between 0 and 1, so the log of a likelihood is less than or equal to zero. If a model has a very low likelihood, then the log of the likelihood will have a larger magnitude than the log of a more likely model. Thus, a small ratio of log likelihoods indicates that the full model is a far better fit than the intercept model. If comparing two models on the same data; **McFadden's** would be higher for the model with the greater likelihood. $R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{\mathrm{int}ercept})}$ Where, $M_{full}$= Model with predictors, $M_{intercept}$ = Model without predictors, $\hat{L}$= Estimated likelihood.

**Cox & Snell**

*L(M)* is the conditional probability of the dependent variable given the independent variables. If there are n observations in the dataset sample, then *L(M)* is the product of n such probabilities. Thus, taking the $n^{th}$ root of the product *L(M)* provides an estimate of the likelihood of each dependent value. Cox & Snell's presents the R-squared as a transformation of the $-2\ln[L(M_{Intercept})/L(M_{Full})]$ statistic that is used to determine the the convergence of a logistic regression. If the full model predicts the outcome perfectly and has a likelihood of 1, Cox & Snell's is then $1-L(M_{Intercept})^{2/n}$, which is less than one. $R^2 = 1 - \left\{\frac{L(M_{Intercept})}{L(M_{Full})}\right\}^{2/n}$ .

**Nagelkerke**

To achieve this, the Cox & Snell R-squared is divided by its maximum possible value, 1-*L(M$_{Intercept}$)$^{2/n}$*. Then, if the full model perfectly predicts the outcome and has a likelihood of 1, Nagelkerke R-squared = 1. $R^2 = \dfrac{1-\left\{\dfrac{L(M_{Intercept})}{L(M_{Full})}\right\}^{2/n}}{1-L(M_{Intercept})^{2/n}}$.

## 2.2 The Model

**Complimentary Log-log (Clog-log) Link Function**

Clog-log link function was used to predict the dependent variable category because in ordinal regression analysis. Clog-log link function was used to build the models as it is suitable for analyzing the ordered categorical data with higher categories more probable among all categories. The Clog-log link function is of the form;

$$f[\gamma_j(X)] = \log\{-\log[1-\gamma_j(X)]\} = \log\{-\log[P(Y = y_i|X)/P(Y > y_i|X)]\} = a_j + \beta X, \text{----(3)}$$

and $\gamma_j(X) = 1 - e^{-e(a_j + \beta X)}$ ……………………………………………………………………………………(4)

where, j = 1, 2, …, k - 1 and j indexes the cut-off points for all categories of the dependent variable. Since multiple explanatory variables are involved in the ordinal regression model, the linear combination of $(\alpha_1, \alpha_2, \cdots, \alpha_p)\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{pmatrix}$ is replaced with $\underline{\alpha}^{'}\underline{Y}$. The term complementary function comes from [1- $\gamma_j$ (X)]. Thus, the name of the complementary log-log link function is derived from log {-log [1- $\gamma_j$ (X)]}. Since the ordinal Clog-log model is non linear, transformation should be on the dependent variable, which equals to the linear form of $a_j + \beta X$. The ordinal regression model with the Cloglog link is called the continuation ratio model because it is a ratio of the two conditional probabilities, for example, P(Y =$y_j$ | X) to P(Y >$y_j$ | X). The model with the cloglog link is also called the proportional hazard model because the relationship between the explanatory variables and the ordinal outcome is independent of the category (Chau-Kuang Chen *et al,* 2004).

The coefficients in the ordinal regression model depict how much the Clog-log changes based on the values of the predictor variables. Statistical fittings that were analysed are; parameter estimates table with location variable that gives the coefficient for the independent variable for the specified link function in ordinal regression, factor summary table that depicts that the general question ordinal scale distribution in percentage on respondents, model fitting information table that checks the presence of a relationship between the dependent variable and combination of independent variables was based on the statistical significance of the final model, goodness of fit information table with Pearson chi-square test that gives the information about how many predicted cell frequencies differ from observed frequencies, Test of parallel lines that was designed to make judgment concerning the model adequacy.

### 2.2.1 The Assumptions

***Parallel Lines***

One of the assumptions underlying ordinal regression is that the relationship between each pair of outcome groups is the same. In other words, ordinal regression assumes that the coefficients that describe the relationship between, say, the lowest versus all higher categories of the response variable are the same as those that describe the relationship between the next lowest category and all higher categories, etc. This is called the proportional odds assumption or the parallel regression assumption. Because the relationship between all pairs of groups is the same, there is only one set of coefficients. Thus, in order to asses the appropriateness of the model proportional odds assumption is normally evaluated (O'Connell, 2000).

***Adequate Cell Count***

As per the rule of thumb, 80% of cells must have more than 5 counts. No cell should have zero count as it is considered as a missing value and excluded from the study. The large percentage of cells with missing data could lead to a decrease of actual sample size from the model construction or an inaccurate Chi-square test for the model fitting, since the model goodness-of-fit is usually dependent of chi-square test. The chi-square test normally depends on the sample size (Agresti, 2002).

## 2.3 Data Collection

A pilot sample survey was conducted before the actual survey in June 2009, with the help of trained class representatives on questionnaire distribution in their respective classes in the faculty that was used to calculate the satisfaction sample proportion of the student $\hat{p} = \frac{x}{n}$ and $\hat{p} \sim N(p, \frac{p(1.p)}{n})$ where, p is the population proportion of students' satisfaction and Z= $\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ $Z \sim N(0,1)$. To calculate a representative sample estimate n the confidence interval for population proportions formulae given by $\left[ \hat{p} \pm Z_{\alpha} \sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \right]$ was used. $\hat{p}$=0.86668, $\hat{q} = 1 - \hat{p}$ and since satisfaction is one tailed α=0.05, $Z_{0.05}$=1.645.

Error= $\hat{p} - p = Z_{0.05}\sqrt{\frac{\hat{p}\hat{q}}{n}}$, but error=0.05 then n= $\left(\frac{1.645}{0.05}\right)^2 (0.86668)(0.13332) = 125.0$

which was the sample size. Majorly, fourth years in the eight departments in the Faculty of Science were focused on as they were thought to have had more experience on university operations, compared to the students of other years' of study who might have not even known their department offices or not used most of the faculty's facilities and services to improve the study quality. The sampled departments were selected using stratified random sampling technique with the eight selected departments in the faculty being strata and simple random sampling was used in selecting the sampled students in each department.

## 3.0 Results

Factors' summary implies that over 75% of the students are above averagely satisfied. Of these, 25% are either satisfactory or very satisfactory and 25% are below average satisfaction. Overall all departments participated in equal capacity. The ratio of female to male respondents was approximately 39% to 61%. The big difference depicts the general female to male ratio in the university. The factor programme of study had Government of Kenya (GOK) sponsored students at 41.6% and Alternative Degree Programme (ADP) students at 58.4% the difference in respondents in percentages was because some of the departments like Statistics and Actuarial Sciences have more ADP students.

*Table 1: Observed distribution of general question, participation by department, gender and programme of study*

| **Variables** | | **N** | **Marginal Percentage** |
|---|---|---|---|
| **General question** | Very unsatisfactory | 7 | 5.6% |
| | Unsatisfactory | 24 | 19.2% |
| | Average | 62 | 49.6% |
| | Satisfactory | 27 | 21.6% |
| | very satisfactory | 5 | 4.0% |
| **Department** | Physics | 15 | 12.0% |
| | Medical Microbiology | 14 | 11.2% |
| | Zoology | 15 | 12.0% |
| | Statistics and Actuarial Sciences | 17 | 13.6% |
| | Pure and Applied Mathematics | 21 | 16.8% |
| | Biochemistry | 16 | 12.8% |
| | Chemistry | 13 | 10.4% |
| | Botany | 14 | 11.2% |
| **Gender** | Female | 49 | 39.2% |
| | Male | 76 | 60.8% |
| **Programme of study** | GOK | 52 | 41.6% |
| | ADP | 73 | 58.4% |

### 3.1 Model Fitting Information

The results from model fitting in the section provide results of ordinal logistic regression versus reduced model (intercept) with complimentary log-log link function. The presence of a relationship between the dependent variable and combination of independent variables is based on the statistical significance of the final model. From Table 2, the -2LL of the model with only intercept is 321.455 while the -2LL of the model with intercept and independent variables is 0.001. The difference (Chi-square statistics) is 321.455- 0.000 = 321.455 which is significant at $\alpha$=0.05, $p \leq 0.001$. The conclusion is that there is association between the dependent and independent variable(s) in complimentary Log-log link function.

*Table2: Model fitting information*

| Model | -2 Log Likelihood | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Intercept Only | 321.455 | - | - | - |
| Final | .000 | 321.455 | 34 | 0.000 |

### 3.2 Goodness-of-fit

Pearson is widely used in statistics to measure the degree of the relationship between the linear related variables. Deviance is a likelihood-ratio test used under full maximum likelihood. The deviance can be regarded as a measure of lack of fit between model and data. Generally, the larger the deviance, the poorer the fit to the data. The deviance in usually compared to deviances from other models fitted to the same data. The difference between the deviances $D_0$ and $D_1$ has a large-sample chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated. The null hypothesis states that the observed data are consistent with the fitted model. The null hypothesis is accepted and one concludes that the observed data were consistent with the estimated values in the fitted model since the p was insignificant, p = 1.00 > 0.05.

*Table 3: Goodness-of-fit*

| | Chi-Square | Df | Sig. |
|---|---|---|---|
| Pearson | 299.192 | 462 | 1.000 |
| Deviance | 238.120 | 462 | 1.000 |

### 3.3 Pseudo R-Square

**Cox & Snell's** pseudo R-squared has a maximum value that is not 1 thus the full model predicts the outcome almost perfectly as the likelihood value is 0.924. **Nagelkerke** full model perfectly predicts the outcome as its likelihood of R-squared = 1. **McFadden's** the ratio of the likelihoods suggests the model predicted the outcome perfectly as its likelihood is 1.

*Table 4: Goodness-of-fit*

| | |
|---|---|
| Cox and Snell | 0.924 |
| Nagelkerke | 1.000 |
| McFadden | 1.000 |

### 3.4 Parameter Estimates

Four factors were found to influence the satisfaction of the Jomo Kenyatta University of Agriculture and Technology Faculty of Science namely; service delivery at faculty office to the students, $p = 0.033 < 0.05$, library service's significance, $p = 0.030 < 0.05$ but the negative sign indicates that the students were less satisfied, accommodation facilities inside JKUAT hostels, $p = 0.009 < 0.05$ and accommodation facilities outside JKUAT had, $p = 0.028 < 0.05$. Since the estimated $p \leq 0.05$ the null hypothesis is rejected and conclude that the regression coefficient for the four predictor variables were found to be statistically different from zero in estimating the general question in the presence of other independent variables.

*Table 5: Parameter Estimates*

| Location parameters | Estimate | Std. Error | Wald | df | Sig. | 95% Confidence Interval lower bound | upper bound |
|---|---|---|---|---|---|---|---|
| Year of study | .114 | .191 | .355 | 1 | .551 | -.261 | .489 |
| Gender (0,1) | .385 | .301 | 1.640 | 1 | .200 | -.204 | .974 |
| Program of study (0,1) | -.939 | .873 | 1.157 | 1 | .282 | -2.650 | .772 |
| Age in years | -.136 | .082 | 2.733 | 1 | .098 | -0.297 | 0.025 |
| Reference to market | -.173 | .162 | 1.140 | 1 | .286 | -0.491 | 0.145 |
| Lecturer service delivery | .248 | .215 | 1.329 | 1 | .249 | -0.174 | 0.670 |
| Service delivery faculty | .391 | .183 | 4.561 | 1 | .033 | 0.032 | 0.749 |
| Communication skills | -.090 | .183 | .244 | 1 | .622 | -.449 | .268 |
| Computer Skills | .150 | .165 | .825 | 1 | .364 | -.173 | .473 |
| Research skills | -.094 | .155 | .373 | 1 | .542 | -.397 | .209 |
| Admission registration process | .242 | .153 | 2.498 | 1 | .114 | -.058 | .542 |
| Financial ability to pay tuition fee | -.205 | .135 | 2.291 | 1 | .130 | -.470 | .060 |
| Service delivery department office | .003 | .156 | .000 | 1 | .984 | -.303 | .309 |
| Students and subordinate staff relationship | .081 | .219 | .136 | 1 | .712 | -.348 | .510 |
| Faculty office accessibility | -.029 | .193 | .023 | 1 | .880 | -.407 | .349 |
| Course promotion | .256 | .162 | 2.520 | 1 | .112 | -.060 | .573 |
| Library service | -.358 | .164 | 4.730 | 1 | .030 | -.680 | -0.035 |
| Tutorial Service | .066 | .145 | .204 | 1 | .652 | -.219 | .350 |
| Career counseling services | .143 | .145 | .970 | 1 | .325 | -.142 | .428 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Classroom facilities | .122 | .146 | .697 | 1 | .404 | -.164 | .409 |
| JKUAT hospital facilities | -.065 | .152 | .184 | 1 | .668 | -.363 | .233 |
| Course laboratory facilities | .220 | .150 | 2.157 | 1 | .142 | -.074 | 0.513 |
| Accommodation facilities in JKUAT | .592 | .228 | 6.773 | 1 | .009 | .146 | 1.038 |
| Accommodation facilities outside JKUAT | .334 | .152 | 4.818 | 1 | .028 | .036 | 0.631 |
| JKUAT Internet facilities | -.032 | .184 | .030 | 1 | .862 | -.394 | .329 |
| JKUAT student center facilities | -.019 | .161 | .014 | 1 | .905 | -.334 | .296 |
| Faculty interdepartmental sports events | .010 | .141 | .005 | 1 | .942 | -.267 | .288 |
| [Department=1] | -.071 | .549 | .017 | 1 | .897 | -1.146 | 1.005 |
| [Department=2] | .640 | .612 | 1.093 | 1 | .296 | -.560 | 1.840 |
| [Department=3] | -.252 | .494 | .260 | 1 | .610 | -1.220 | .716 |
| [Department=4] | -.717 | .563 | 1.621 | 1 | .203 | -1.821 | .387 |
| [Department=5] | -.692 | .503 | 1.893 | 1 | .169 | -1.678 | .294 |
| [Department=6] | -.078 | .541 | .021 | 1 | .886 | -1.137 | .982 |
| [Department=7] | -.624 | .554 | 1.265 | 1 | .261 | -1.710 | .463 |
| [Department=8] | 0(a) | . | . | 0 | . | . | . |

### 3.5 Test of Parallel Lines

Test of parallel lines was designed to make judgment concerning the model adequacy. The model null hypothesis states that the slope coefficients in the model are the same across the response categories. The significance p = 1.000 >0.05 indicated that there was no significant difference for the corresponding slope coefficients across the response categories, suggesting that the model assumption of parallel lines was not violated in the model with the Complementary Log-log link.

*Table 6: Test of parallel lines*

| Model | -2 Log Likelihood | Chi-Square | df | Sig. |
|---|---|---|---|---|
| Null Hypothesis | .000 | - | - | - |
| General | .000(a) | .000 | 102 | 1.000 |

## 4.0 Conclusion and Recommendations

### 4.1 Conclusion

The explanatory variable related to the satisfaction of faculty involvement is service delivery at the department office. This was identified in the best model. Student satisfaction with faculty involvement significantly contributes to the probability of students expressing satisfaction with the general on Faculty of Science service delivery. It is evident that the Faculty of Science is one of the largest faculties within the university, thus higher student satisfaction rating regarding faculty involvement provides compelling evidence that faculty members have played a significant role in creating a pleasant environment influenced on student satisfaction generally on faculty of science service delivery.

Furthermore, the library services were significantly associated with the general satisfaction of generally on Faculty of Science service delivery. It may provide evidence that improved service delivery at the library has addressed the needs of Faculty of Science students and contributed to the fulfillment of university goal, e.g., booking of and reserving of library books online a success.

The study suggested that the accommodation facilities inside and outside the university can be improved further by providing services like internet in the hostels and provision of free transport to and from the university for students residing outside the university, and providing power backups in the halls of residence.

The goal was to obtain information from students to establish the explanatory variables that influence satisfaction that could be helpful to decision makers in Faculty of Science for improving academic programmes, facilities and services in the faculty. For example, the administrators could ensure that the faculty students themselves participate in the quality of academic programmes supported by the faculty capacity and facilities and services. Model assumption of parallel lines was checked to ensure model adequacy and it was fulfilled by the model, assuring the model goodness of fit, fitting Information and parameter estimation stability.

### 4.2 Recommendations

(i) The sample should be picked from faculties, schools and institutes to increase the sample size.

(ii) The questionnaire or otherwise could be used in other public and chartered universities using stratified random sampling and populations using regional counties in Kenya as strata.

**References**

Agresti A. (2002). Categorical Data Analysis. John Wiley and Sons: New York.

Bailey B. L., Bauman C and Lata K. A. (1998). Student Retention and Satisfaction: The Evolution of a Predictive Model. Paper presented at the Annual Forum of the Association for Institutional Research, Minneapolis, Minnesota. Eric No: ED424797

Chau-Kuang C. and Hughes J. (2004). Using Ordinal Regression Model to Analyse Student Satisfaction Questionnaires. Journal of IR Applications, **1**(1)

Cooney F. (2000). A Review of the Results and Methodology in the 1999 Noel Levitz Student Satisfaction Survey at Salt Lake Community College. Salt Lake City, Utah: Salt Lake Community College. Eric No: ED443482. JC 000 518

Damminger J. K. (2001). Student Satisfaction with Quality of Academic Advising Offered by Integrated Department of Academic Advising and Career Life Planning. Glassboro, New Jersey: Rowan University. Eric No: ED453769

Hummel T. J. and Lichtenberg J.W. (2001). Predicting Categories of Improvement among Counseling Center Clients. Paper presented at the annual meeting of the American Educational Research Association, Seattle, W. A. Eric No: ED 451434. CG 030 799

O'Connell A. A. (2000). Methods for modeling ordinal outcome variables. Measurement and Evaluation in Counseling and Development , pp 170-193.

Robins L. S., Gruppen L. D., Alexander G. L., Fantone J. C., and Davis W. K. (1997) A Prediction Model of Student Satisfaction with the Medical School Learning Environment. Journal of Academic Medicine, **72**(2)

SPSS, Inc. (2002). Ordinal Regression Analysis, SPSS Advanced Models 10.0., Chicago, IL.
Emily H. T. and Nora G. (2002) What Satisfies Students? Mining Student-Opinion Data with Regression and Decision-Tree Analysis. Stony Brook, New York: Stony Brook University, **45**,(3).

Umbach P. D. and Porter S. R. (2001). How Do Academic Departments Impact Student Satisfaction? Understanding the Contextual Effects of Departments. AIR 2001 Annual Forum Paper. *Journal of Association for Institutional Research*, Long Beach, California. Eric No: ED456789.

Wild N. (2000). Rogue Community College Student Satisfaction Survey, Management Report. Redwood and Riverside Campuses. Grant Pass, Oregon: Rogue Community College. Eric No: ED448831.