Journal of Scientific and Industrial Studies, Volume 7, Number 1, 2009
 ISSN: 1117 – 1693

APPLICATION OF ORDINAL LOGISTIC REGRESSION MODEL TO OCCUPATION DATA

A. A. Adepoju and M. Adegbite
Department of Statistics
University of Ibadan, Ibadan, Oyo State, Nigeria

ABSTRACT

People's occupational choices might be influenced by their parents' occupation, gender, previous experiences, ages, and their own education level. We can study the relationship of one's occupation choice with education level and father's occupation. The occupational choices will be the outcome variable which consists of categories of occupations. The regression methods are capable of allowing researchers to identify explanatory variables related to organizational programs and services that contribute to the overall staff status. These methods also permit researchers to estimate the magnitude of the effect of the explanatory variables on the outcome variable. Therefore, regression methods seem to be superior in studying the relationship between the explanatory and outcome variables. This study used ordinal logistic regression method to examine the relationship between the ordinal outcome variable, different levels of staff status in the Lagos State Civil Service of Nigeria, the explanatory variables are Gender, Indigenous status, Educational Qualification, Previous Experience and Age. The outcome variable was measured on an ordered, categorical, and three-point Likert scale as Junior staff, Middle Management staff, and Senior Management staff. Within the complete models, the logit link was the better choice because of its satisfying parallel lines assumption and larger model-fitting statistics. The study revealed that two explanatory variables namely, Education Qualification and Previous Working Experience significantly predicted the probability of an individual staff being a member of any of the three levels of staff status.
Keywords: Ordinal Logistic Regression, Proportional Odds Model, Binary Logistic, Categorical Data, Likert Scale.

INTRODUCTION

Occupation may be referred to as the principal activity (Job, employment or calling) that earns money (regular wage or salary) for a person. Human resources are indispensable assets any nation can be endowed with because they help in the coordination of all other factors of production. Labour force on the other hand. is referred to as a set of people or citizens of a country who are willing and able to make available, at any point in time, their efforts for gainful employment. However, it is not everybody in all the entire population that constitute or can constitute the Labour force of a nation. The Labour force is, to a large extent, determined by some factors, which are: the age structure of the population, physical health, religious belief and the law of the land. The supply of manpower is determined by the size of the population, its fertility patterns, its age structure, the participation rate or the population taste for employment and the extent to which the manpower is employed. At present, it is only possible to speculate on the implications of different levels of Labour force participation such as Senior Management, Middle Management and junior staff. This is because little empirical or theoretical research has been done on them. Labour force participation data can serve as summary indicators of general conditions of the population, and as basis for modelling and forecasting, hence the need for this research work. The pattern of data in this environment is mostly categorical in nature hence the need to employ the ordinal logistic regression model. This is a general approach to the analysis of more categorical responses having more than two possible values. A variety of methods has been developed for covering the various possibilities and the best known and most highly developed are methods for ordinal response variables. The advantage of using these technique compared to traditional tests of association is that the models allow the inclusion of association terms without saturation that is, the model do not require all the degrees of freedom. Furthermore, it is possible to construct more parsimonious models and also detect occupational distribution in addition to describing certain associations which are meaningful base on parameters. These parameters are the odds ratio, which are easy to interpret (Agresti, 2002). Several works have been done using ordinal Logistic Regression model. The works cut across medical, social and economical phenomena. Plank, Stephen B. and Jordan, Will J. (1997) used the logit model to predict college going behavior and found out that the probability of campus residency increased with the percentage of students living on campus in the absence of monetary constraints. Chau-Kuang Chen and John Hughes, Jr. (2004) used ordinal Logistic Regression model to analyze student satisfaction Questionnaires. The ordinal Logistic regression method was used to model the relationship between the ordinal outcome variable, e.g., different levels of student satisfaction regarding the overall college experience, and the explanatory variables concerning demographics and student learning environments in a predominantly minority health sciences center. The outcome variable for student satisfaction was measured on an ordered, categorical, and four-point Likert-scale- 'very satisfied', 'dissatisfied', 'satisfied', and 'very satisfied'. Explanatory variables included two demographics e.g., gender and ethnic groups. His research findings indicated that explanatory variable such as faculty competence and student- faculty relations were significantly associated with the satisfaction of the overall college experience. The discovery suggested that faculty members had played a major role in creating a pleasant environment to facilitate student satisfaction. Maria Halena Spyrides-Cunha. (1998) applied ordinal logistic regression model to the mapping of disease resistance genes in plants. Mapping was done by establishing a statistical association between molecular marker genotypes and quantitative variations in disease resistance. A categorical variable is considered ordinal if there is a natural ordering of the possible values, for example Low, Medium and High. A number of proposed models for this type of data are extensions of the logistic regression model. The most widely known of the ordinal logistic regression methods is called the proportional odds model. The basic idea underlying the proportional odds model is re-expressing the categorical variable in terms of a number of binary variables based on internal cut- points in the ordinal scale. For example, if y is a variable on a 4 -point scale, we can define the corresponding binary variables, y_c^* $c = 1,2,3$ by $y_c^* = 1$ if $y > c$ and y_c^* if $y \le c$. If one has a set of explanatory variables

$x_j, j = 1,\dots,k$, then we can consider the 3 binary logistic models corresponding to regressing each of the y_c^*'s separately against the $x's$ can be considered. The proportional odds model assumes that the true β-values are the same in all three models, so that the only difference in models is the intercept terms, $\alpha_c, c = 1,2,3$. This means that the estimates from the three binary models can be pooled to provide just one set of β estimates. Obtaining the exponent of the pooled estimate relative to a given predicator, i.e. taking λ^{β_j}, we obtain an estimate of the common odds ratio that describes the relative odds for $y > c$ for values of x_j j, differing by 1 unit. Thus, interpreting the proportion odds model is not much more difficult than a binary logistic regression.

Regression methods such as linear, logistic, and ordinal regression are useful tools to analyze the relationship between multiple explanatory variables and occupational staff status (Essien, M.C. 2003; and Hosmer, David W. and Stanley Lemeshow 2002). Despite the prevalence of linear and logistic regression analyses, researchers are experiencing the challenge of using ordinal outcome because in part, they have not been fully exposed to the mathematical theory and the application software. Nowadays, the availability of statistical software routines in the Statistical Package for the Social Science (SPSS) or Statistical Analysis System (SAS) makes it computationally possible to build an ordinal logistic regression model The application of linear, logistic, and ordinal regression methods depends largely on the measurement scale of the outcome variables and the validity of the model assumptions. The outcome variables include continuous scale, (e.g., total staff strengths), binary measure (e.g., junior management staff and senior management staff ratings), or ordered category (e.g., junior management staff, middle management staff, and senior management staff). Linear regression analysis is applicable to the outcome variable measured on a continuous scale while logistic regression analysis works well only for the binary or dichotomous outcome. In linear and logistic regression analyses, the model assumptions of normality and constant variance for the residual and the outcome data points need to be satisfied to demonstrate their appropriateness. If researchers wish to study the effects of explanatory variables on all levels of the ordered categorical outcome, an ordinal regression method must be appropriately chosen to obtain the valid results. More examples of ordinal outcomes include certain psychological measurement (e.g., levels of anxiety or depression, quality of life studies), rank scores (e.g., letter grades of the course work), and the most frequently used Likert-scale e.g., "poor", "fair", "good", and "excellent" ratings (Rubinfeld D. L 1977). It is implausible to assume the normality and homogeneity of variance for ordered categorical outcome when the ordinal outcome contains merely a small number of discrete categories. Thus, the ordinal regression model becomes a preferable modelling tool that does not assume the normality and constant variance, but require the assumption of parallel lines across all levels of the categorical outcome. The step-by-step procedures for building, evaluating, and interpreting the ordinal regression model were illustrated in this study. Essentially, the study followed four sequential protocols to create a workable model. First of all, the potential explanatory variables were examined to determine if they should be included in the model. Secondly, the outcome variable was coded or labelled as ordered, ranked, and categorical values. The explanatory variables were either a continuous or a discrete measure. Third, the complete and the reduced models along with the logit link and the complementary log-log (clog log) link were used to generate the candidate models. The complete model contained all the explanatory variables while the reduced model included a subset of the predetermined explanatory variables. The logit and the clog log links were chosen to build models based on the distribution of ordinal outcome, either evenly distributed among all categories or clustered around higher categories. Finally, the best model was chosen among all candidate models based on the model fitting statistics, the accuracy of the classification results, the validity of the model assumption, and the principle of parsimony. Clearly, the ordinal regression is a unique modelling technique in that the outcome variable is measured on the ordered categorical scale, various link functions are readily available to apply, and the validity of the model assumption for parallel lines is essentially assessed (Lehmann, E.L. 1988). This research work employs the ordinal logistic regression model to fit an appropriate model for occupational participation data of the Lagos state Civil service from 1967 to 2006.Parameter estimates for the models were obtained using both weighted least squares (WLS) and the maximum likelihood estimation (MLE) techniques. The full model was fitted and the Statistical investigation of the fitted model was carried out using the Wald's statistic and the likelihood ratio test. The Pearson statistic and Deviance statistic were used to test for the adequacy of the fitted model.

2.0 Model Description

Suppose the dependent variable $Y = \begin{cases} 1, if & junior & staff \\ 2, middle & management & staff \\ 3, senior & management & staff \end{cases}$

and let $P_1 = P(y=1)$, $P_2 = P(y=2)$, and $P_3 = P(y=3)$. The ordinal logistic regression models, the relationship between the cumulative logits of Y, that is, $Log(P_1/1-P_1) = Log[P_1/(P_2+P_3)]$ and $Log[(P_1+P_2)/(1-P_1+P_2)] = Log[(P_1+P_2)/P_3]$ and independent variables.

The model assumes a linear relationship for each logit and parallel regression lines.

$$Log(P_1/(1-P_1)) = \text{int}ercept1 + \beta_1^{*}X_1 + \beta_2^{*}X_2 + \ldots + \beta_k^{*}X_k$$

$$Log[(P_1+P_2)/P_3] = \text{int}ercept2 + \beta_1^{*}X_1 + \beta_2^{*}X_2 + \ldots + \beta_k^{*}X_k$$

That is, the intercepts are different, but the remaining regression parameters are the same. It is easy to show that the odds $[P_1/(1-P_1)]$ and $[(P_1+P_2)/P_3]$ are proportional; Given that the intercept 1 and 2 are denoted by α_1 and α_2 respectively,

$$[P_1/(1-P_1)] = \lambda^{\alpha_1}\lambda^{\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k}$$

$$[(P_1+P_2)/P_3] = \lambda^{\alpha_2}\lambda^{\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k}$$

Proportional odds imply that odds ratios for y being Junior management staff (1) versus middle management staff or senior management staff (2 or 3) and for y being junior staff or middle management staff (1 or 2) versus senior management staff (3) are the same.

ESTIMATION OF PARAMETERS

The maximum likelihood estimation is used to obtain the estimates of the model parameters. Maximum likelihood (ML) estimates can be obtained by iterative methods such as the iterative re weighted least squares method which is found in the major statistical packages, such as GLIM4, SAS (1988), SPSS version 11.0 and above, and others. McCullagh (1980) showed how to use the Newton- Raphson method for maximum likelihood estimation in a class of models that includes cumulative logit models. For ordinal response scales, it is more suitable to form the link function using the cumulative probabilities $P_j = P(y \le j)$, [table1], instead of the response category probabilities because of the former useful properties (McCullagh and Nelder, 1989).

With the values of $\alpha_1, \alpha_2, \beta_1, \beta_2, \ldots, \beta_k$ computed, it is easy to compute predicted probabilities using the following formulas derived from the equations formulas derived from the equations above.

$$P_1 = \lambda^{\alpha_1+\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k} / \left(1+\lambda^{\alpha_1+\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k}\right)$$

$$P_1 + P_2 = \lambda^{\alpha_2+\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k} / \left(1+\lambda^{\alpha_2+\beta_1^{*}X_1+\beta_2^{*}X_2+\ldots+\beta_k^{*}X_k}\right)$$

$$P_3 = 1-(P_1+P_2)$$

If parameters β_i are positive, then P_1, predicted probability of (Y=1, Junior staff), as well as cumulative probability of (y= 1, Junior Management staff, or y = 2, middle management staff), $P_1 + P_2$ are higher for higher values of explanatory variables Xi. If parameter β_i is negative, P_1 and $P_1 + P_2$ are lower for higher values of X_i.

DATA PRESENTATION AND ORDERED LOGISTIC REGRESSION USING PLUM

PLUM, introduced in SPSS version 10, can estimate a variety of ordinal regression models including ordered logit and ordered probit. This study is limited to the ordered logit model. The ordered logit model, also known as the cumulative logit model, estimates the effects of independent variables on the log odds of having lower rather than higher scores on the dependent variable.

$$\ln\left[\frac{P(y \le j)}{P(y > j)}\right] = \alpha_j - \sum_{k=1}^{K} \beta_k X_k \; for \; j = 1 to J-1$$

In the above equation, α_j are intercepts indicating the log odds of lower rather than higher scores when all independent variables equal zero. Note that the efforts of the independent variables $\beta_k X_k$ are subtracted from rather than added to the intercepts. This is done so that positive coefficients indicate increased likelihood of higher scores on the dependent variable (Cf. Agresti 1990; 323). The intercepts for J-1 categories express the categorical nature of the dependent variable while a parallel odds restriction to let independent variables have the same effects on all cumulative logits results in a parsimonious model for ordinal data. The following summary tables show the output of how an ordered logistic model can be estimated with PLUM, SPSS version 14.0 was used to run the data. The dependent variable is respondent's staff status; the five independent variables are Sex (GENDER), indigenous status (INDIG), Educational Qualification (EDU), and Previous Experience (PREV.EXP) and Individual age (AGE). STAFF STATUS (response variables) has 3 categories: 1 = Junior staff, 2 = Middle management staff and 3 = senior management staff. GENDER is the respondent's sex (0 = female and 1 = male). INDIG is the respondent's indigenous status (0 = non- indigene and 1 = indigene). EDUC is the respondent's educational qualification (1= graduate and 0 = non graduate). PREV.EXP measures years of working experience with a range of 1 to 34. AGE also measures the respondent's years of age which range from 21 to 59. The continuous independent variables Age and working experience were recoded into categorical form so as to reduce the number of cells and zero frequencies common place. Thus PREV. EXP is changed and recoded into NEW EXPERIENCE (0 = PREV.EXP $\le$ 10 and 1 = PREV.EXP >10). The value 10 or less was chosen based on the fact that the statutory working years for a civil servant to be eligible for gratuity if leaving the service at that period is 10years and above. Age changed and recoded into BIRTH (0=AGE $\le$30 and 1=AGE > 30). The value 30 is chosen because 30years is the average Age of individual staff recruited into the service. The study results for the complete model containing all staff status revealed a number of interesting findings. Within the complete models, the logit link was the better choice because of its satisfying parallel lines assumption and larger model- fitting statistics, which will be discussed later. Using the complete model with the logit link, Table 1 shows that the two thresholds of the model equation were significantly different from zero and substantially contributed to the values of the response probability in different categories. In addition the occupational staff status of the Lagos state civil service was significantly associated with the five explanatory variables (Gender, Indigenous status, Education status, Birth and Experience).

Table 1 : Case Processing Summary

		N	MARGINAL PERCENTAGE
STAFF STATUS	1 = JUNIOR MANAGEMENT	154	30.8%
	2 = MIDDLE MANAGEMENT	302	60.4%
	3 = SENIOR MANAGEMENT	44	8.8%
VALID		500	100%
MISSING		0	
TOTAL		500	

Frequency table of the dependent variable.

Table 2: Model Fitting Information

	-2Log Likelihood	Chi-square	df	sig
MODEL	589.642			
INTERCEPT ONLY	128.727	460.915	5	0.000
Link function; Logit				

The value of 460.915 with 5 df is the most relevant value here. This is the likelihood ratio test that all coefficients for all independent variables are equal to zero i.e. β_j's zero. This null hypothesis can be rejected since the test is highly significant.

Table 3: Goodness-Of-Fit

	Chi-square	df	Sig.
Pearson	322.032	43	0.000
Deviance	85.699	43	0.000

The additional model fitting statistic, the Pearson's Chi-square, (X^2 = 322.032 with 43 df. and P = 0.000) for the complete model with the logit link indicated that the observed data were consistent with the estimated values in the fitted model

Table 4: Pseudo R-Square

Cox and Snell	0.602
Magelkerke	0.727
Mc Fadden	0.523

The model-fitting statistic, namely the pseudo R-squared, measured the success of the model in explaining the variations in the model in explaining the variations in data. The pseudo R- squared was calculated depending upon the likelihood ratio. For example, the Mc Fadden's R- squared compared the likelihood for the intercept only model to the model goodness of fit. The interpretation of pseudo R- squared in the ordinal regression model was similar to that of the R- squared (e.g., Coefficient of the Determination) in the linear regression model. The pseudo R-squared indicated that the proportion of variations in the outcome variable was accounted for by the explanatory variables. The larger the pseudo R-squared was, the better the model fitting was. The pseudo R-squared values for Mc Fadden, Cox and Snell and Nagelkerke in the complete model with the logit link are respectively, (0.523), (0.602) and (0.727).

Table 5: Parameter Estimates

Threshold	Estimate	Std Error	Wald	df	Sig	95% confidence interval lower band	upper band
Status = 1	4.322	0.817	27.983	1	0.000	2.721	5.923
Status =2	10.718	1.009	112.814	1	0.000	8.741	12.696
Location							
Gender	0.375	0.258	2.100	1	0.147	-0.132	0.881
Indig	0.366	0.261	1.967	1	0.161	-0.145	0.876
Educ	6.818	0.529	166.356	1	0.000	5.782	7.854
Birth	0.609	0.786	0.601	1	0.430	-0.931	2.149
Experience	3.757	0.374	100.855	1	0.000	3.024	4.491

In table 5 both EDUC and PREV.EXP are statistically significant; GENDER, INDIG, and AGE are not. The estimates in the output are given in units of ordered logits or ordered log odds. So for GENDER, this implies that going from 0 to 1, we expect a 0.38 increase in the log odds of STAFF STATUS, given that all other variables in the model are held constant. For INDIG, we would say that going from 0 to 1, we would expect a 0.37 increase in the expected value of STAFF STATUS in the log odds scale, given that all of the other variables in the model are held constant. The threshold values indicate the cumulative logits when the independent variables equal zero. The positive value for STAFF STATUS = 1 means that P (STAFF STATUS $\leq$ 2)> P (STAFF STATUS >2) when all independent variables are zero. The thresholds are necessary for calculating predicted values but are relatively uninteresting.

Table 6: Test of Parallel Lines

Model	-2Log likelihood	Chi-square	df	Sig.
Null Hypothesis	128.727			
General	60.75	128.727	5	0.080

The test of parallel lines was designed to make judgment concerning the model adequacy. The null hypothesis states that the corresponding regression coefficients are equal across all levels of the outcome variable. The alternative hypothesis states that the corresponding regression coefficients are different across all levels of the outcome variable. The Chi-square test result (X^2 = 128.727 with df. of 5, and P=0.080) indicated that there was no significant difference for the corresponding regression coefficients across the response categories, suggesting that the model assumption of parallel lines was not violated in the complete model with the logit link Estimation of the proportional odds ratios was calculated through the use of SPSS output management system (OMS) to capture the parameter estimates and finding the exponent. The output of the proportional odds ratios results is shown below:

Table 7: The Output of the Proportional Odds Ratios

Threshold	Estimate	Std Error	Wald	df	Sig	95% confidence interval Lower bound	upper bound	EXP(B)
Status = 1	4.322	0.817	27.983	1	0.000	2.721	5.923	
Status =2	10.718	1.009	112.814	1	0.000	8.741	12.696	
Location								
Gender	0.375	0.258	2.100	1	0.147	-0.132	0.881	1.4550
Indig	0.366	0.261	1.967	1	0.161	-0.145	0.876	1.4420
Educ	6.818	0.529	166.356	1	0.000	-5.782	7.854	914.1549
Birth	0.609	0.786	0.601	1	0.430	-0.931	2.149	1.8386
Experience	3.757	0.374	100.855	1	0.000	3.024	4.491	42.6198

Column of exp (β) presents the results of the proportional odds ratios (the coefficient exponentiated). We also have the lower and upper 95% confidence intervals. We would interpret the proportional odds ratios pretty much as we would odds ratio from a binary logistic regression. We will ignore the values for STAFF STATUS = 1 and STAFF STATUS = 2, as those are the threshold and not usually reported in terms of proportional odds ratios. For GENDER, we would say that going from 0 to 1, the odds of high or senior staff status versus the combined middle and junior staff status categories are 1.46 greater, given that all of the other variables in the model are held constant. Likewise, the odds of the combined middle and senior staff status categories versus junior staff status is 1.46 times greater, given that all of the other variables in the model are held constant. For a change in INDIG, the odds of the junior and middle categories of staff status versus the senior category of staff status are 1.44 times greater, given that the other variables in the model are held constant. Because of the proportional odds assumption, the same increase, 1.60 times, is found between junior staff status and the combined categories of middle and senior staff status.Note: one of the assumptions underlying ordinal logistic regression is that the relationship between each pair of outcome groups is the same.Probabilities can be predicted using the estimated proportional odds from the logit model

$$P_1 = \frac{odds}{1+odds} = \frac{\lambda^{4.322 + 0.609Ai + 0.375Gi + 6.818Qi + 0.366indi + 3.757Ei}}{1+\lambda^{4.322 + 0.609Ai + 0.375Gi + 6.818Qi + 0.366indi + 3.757Ei}}$$

$$P_1 + P_2 = \frac{\lambda^{10.718+0.609Ai+0.375Gi + 6.818Qi + 0.366indi + 3.757Ei}}{1+\lambda^{10.718+0.609Ai+0.375Gi + 6.818Qi + 0.366indi + 3.757Ei}}$$

$$P_3 = 1 - (P_1 + P_2)$$

Since all estimated parameter coefficients β_i's are positive then P_1, predicted probability of (y = 1, Junior staff) as well as cumulative probability of (y=1, Junior staff, or y=2, middle management) $P_1 + P_2$, are higher for higher values of predictor variables' A_i, G_i, Q_i, $indi$ and E_i, where A_i =Age, G_i =Gender, Q_i = Education Qualification, $indi$ =Indigenous status and E_i = Experience

Test of Hypothesis using the WALD'S Statistic.

As stated earlier the Wald's statistic is simply the square of the z statistic and this follows a χ^2 distribution with 1 degree of freedom.

H_o (A): $\beta_1 = 0$ (status is independent of predictor AGE).

H_1 (A): $\beta_1 \neq 0$ (status is dependent on predictor AGE).

$\alpha = 0.05$

Computation: $Z = \dfrac{\beta_1}{SE(\beta_1)} = \dfrac{0.609}{0.786} = 0.77480916$

Wald's statistic, $Z^2 = 0.600329234$

$\chi^2_{0.05,1} = 3.841$

Since $Z^2 = 0.6003 < \chi^2_{0.05,1} = 3.841$, we do not reject H_o and conclude that one's status is independent of the individual's AGE at $\alpha = 0.05$

H_o (G): $B_2 = 0$ (status is independent of predictor Gender)

H_1 (G): $B_2 \neq 0$ (status is dependent on predicator Gender).

Computation: $Z = \dfrac{\beta_2}{SE(\beta_2)} = \dfrac{0.375}{0.258} = 1.453488372$

Wald's statistic, $Z^2 = 2.112628448$

We do not reject H_0 since $Z^2 = 2.113 < \chi^2_{0.05,1} = 3.841$ and conclude that status is independent of individual's Gender at $\alpha = 0.05$

H_0 (Q): $\beta_3 = 0$ (staff status is independent of Educational Qualification)

H_1 (Q): $\beta_3 \neq 0$ (staff status is dependent on Educational Qualification)

Computation: $Z = \dfrac{\beta_3}{SE(\beta_3)} = \dfrac{6.812}{0.529} = 12.88846881$

Wald's statistic, $Z^2 = 166.1126282$

We reject H_0 since $Z^2 = 166.113 > \chi^2_{0.05,1} = 3.841$ and conclude that staff status (Junior, Middle or Senior Management) is dependent on Qualification at $\alpha = 0.05$

H_0 (Ind): $\beta_4 = 0$ (staff status is independent of indigenous status)

H_1 (Ind): $\beta_4 \neq 0$ staff status is dependent on indigenous status)

$\alpha = 0.05$

Computation:

$$Z = \frac{\beta_4}{SE(\beta_4)} = \frac{0.366}{0.261} = 1.402298851$$

Wald's statistic, $Z^2 = 1.966442066$

Since $Z^2 = 1.9664 < \chi^2_{0.05,1} = 3.841$, we do not reject H_0 at $\alpha = 0.05$, conclude that staff status is independent of indigenous status.

H_0 (E): $\beta_5 = 0$ (staff status is independent of previous experience)

H_1 (E): $\beta_5 \neq 0$ (staff status is dependent on previous experience)

$\alpha = 0.05$

Computation:

$$Z = \frac{\beta_5}{SE(\beta_5)} = \frac{3.757}{0.374} = 10.04545455$$

Wald's statistic, $Z^2 = 100.911157$

We reject H_0 since $Z^2 = 100.9111 > \chi^2_{0.05,1} = 3.841$, conclude that being staff status (junior, Middle or Senior Management) is dependent on previous experience at $\alpha = 0.05$

TEST OF HYPOTHESIS USING THE WALD'S CHI-SQUARE TEST AND SIGNIFICANT (P-VALUE) FROM TABLE

From Table 7 the Wald's test statistic for the predictor Gender is 2.100 with an associated P-value of 0.147 at $\alpha = 0.05$, we fail to reject the null hypothesis and so conclude that the regression coefficient for Gender has not been found to be statistically different from zero in estimating staff status given that the rest of the predictors are in the model. The Wald statistic for the predictor indigene is 1.967 with an associated p-value of 0.161 at $\alpha = 0.05$, we also fail to reject the null hypothesis and so conclude that the regression coefficient for Indigene has not been found to be statistically different from zero in estimating occupational staff status given that the rest of the predictors are in the model.The Wald statistic for the predictor Education Qualification is 166.356 with an associated P-value of < 0.0001. For $\alpha = 0.05$, we reject the null hypothesis and conclude that the regression coefficient for Educational Qualification is found to be statistically different from zero in estimating occupational staff status given that the rest of the predictors are in the model.The Wald statistic for the predictor prev.EXP. is 100.855 with an associated p-value of <0.0001 setting $\alpha = 0.05$, we reject the null hypothesis and conclude that the regression coefficient for previous experience has been found to be statistically different from zero in estimating staff status given that the rest of the predictors are in the model. The Wald test Statistic for the predictor Age is 0.601 with an associated p- value of 0.438 for $\alpha = 0.05$, we fail to reject the null hypothesis and conclude that the regression coefficient for Age has not been found to be statistically different from zero in estimating staff status given that the rest of the predictors are present in the model. The interpretation for a dichotomous variable such as gender, parallels that of a continuous variable; the observed difference between males and females on occupational staff status was not found to be statistically significant at the 0.05 level when controlling for the rest of predictor (p = 0.081).

SUMMARY AND FINDINGS

In this study, the principle of parsimony along with various link functions was adopted to build the occupational models and to search for the best model. Much of the time and energy was devoted to developing occupational models, checking the model assumptions, assuring the model goodness of fit, and consequently selecting the best model for the Lagos State Government. The model building itself might be partly statistical methodology and partly experience and common sense of the researchers. The ordinal regression method provides a viable alternative to analyze staff status data with the ordered categorical outcome. It does not treat an ordinal outcome as binary or dichotomous measure like logistic regression analysis, which may lead to the loss of information inherent. Also, it did not falsely assume continuous measure and the properties of normality and constant variance for linear regression to analyze few categories of ordinal outcome, which may lead to incorrect analysis. Clearly, the ordinal regression modeling is a unique statistical technique in that the ordinal outcome variable is frequently encountered in the field of education and labour force research and the model assumption of parallel lines is easily assumed and verified. It is convenient for some researchers to analyze ordinal outcome by means of logistic and linear regression analyses. By altering the measuring scale of ordinal outcome, researchers are able to analyze data and produce research findings. However, the loss of information or incorrect analysis may have occurred in some cases. For instance, when the scale of outcome categories (e.g., Junior staff, Middle management staff, and Senior management staff) is arbitrarily collapsed into a binary measure (e.g. junior and Senior management), researchers are forced to use logistic regression analysis to analyze the two levels of ordinal outcome. By doing so, important information may be lost in the resulting model. Also, while few categories of ordinal outcome are treated as continuous measure, linear regression method is used to analyze the ordinal outcome that cannot be plausibly assumed normality and constant variance. Using linear regression method to analyze the ordinal outcome, researchers may produce incorrect estimation and interpretation based on the violation of model assumptions. Therefore, if researchers wish to study the effects of explanatory variables on all levels of the ordered categorical outcome, an ordinal regression method must be appropriately chosen in order to obtain the valid research results. In this study, the ordinal regression method was used to model the relationship between the ordinal outcome variable, e.g., different levels of individual staff regarding the overall status, and the explanatory variables concerning demographics. The outcome variable for an individual staff status was measured on an ordered, categorical, and three-point Likert scale--'Junior staff', 'Middle management', and 'Senior management staff'. Explanatory variables included five demographics, e.g., gender, indigenous status, educational status, previous experience and age. The research findings indicated educational status and previous experience were significantly associated with the occupational status of the Lagos State Civil Servants. Using the ordinal regression method, researchers could identify the significant explanatory variables with their control to enhance occupational distribution regarding Lagos State Civil Service. Essentially, the four sequential protocols are performed to create an ordinal regression model. First, the explanatory variables are examined to determine if they should be included in the model. Second, the outcome variable is coded in ordered, ranked, and categorical fashion. The explanatory variables are quantified by continuous and discrete measures. Third, the complete and the reduced models as well as the logit link are used to produce the occupational models. The complete model contains all the explanatory variables in the model while the reduced model includes only a subset of the predetermined explanatory variables. Finally, the best model is chosen among all occupational Staff Status models depending upon the model fitting statistics, the accuracy of the classification results, and the validity of the model assumption.The study looks at factors that influence the decision of whether an individual belongs to any of these occupational categories- Junior, Middle or Senior Management staff based on his gender, Sex, Educational Qualification, Indigenous status, Previous working experience and Age. The study which is the prediction of occupation using ordinal logit model to model the data of the Lagos state Civil Servants. The ordinal logistic regression models the relationship between cumulative logit of y and independent variables. The fitted model is:

$$Log(P_1/1-P_1) = \underset{(0.817)}{4.322} + \underset{(0.786)}{0.609\text{Ai}} + \underset{(0.258)}{0.375\text{Gi}} + \underset{(0.529)}{6.818\text{Qi}} + \underset{(0.261)}{0.366\text{indi}} + \underset{(0.374)}{3.757\text{E}_i}$$

$$Log\left[\frac{P_1+P_2}{P_3}\right] = \underset{(1.009)}{10.718} + \underset{(0.786)}{0.609\text{Ai}} + \underset{(0.258)}{0.375\text{Gi}} + \underset{(0.529)}{6.818\text{Qi}} + \underset{(0.261)}{0.366\text{indi}} + \underset{(0.374)}{3.757\text{E}_i}$$

The fitted ordered logit model is a main effects model. The model assumes a linear relationship for each logit and parallel regression lines. It also assumes that the explanatory variables affect the response variable but with a lack of interaction i.e., the effect of a staff Gender sex on staff status for example is the same for each individual's gender sex and so on. We observed from the fitted model that both Educational Qualification and previous experience are statistically significant; Gender, Indigene and Age are not. We equally observed from the table 1 that the proportion of middle Management staff is the highest among the three categories. The proportion of senior Management staff is only higher when higher age and experience are considered. We also noticed that no junior staff is observed as being a graduate. Thus for an individual to be a member of Middle or Senior Management he must either be a graduate or having some higher working experience. Again Tables 2 to 5 present some descriptive statistics for the dependent or predictors and factor variables. Before we run our ordinal logistic model, we checked if any cells are empty or extremely small. If any are, we may have difficulty running our model. Thus, the simple cross-tabs of dependent and independent were made, since none of the cells was too small (has no cases) so the model was run. The main aim of the study which is to predict relative probability P_1 predicted probability of (y = 1, junior staff), as well as cumulative probability of (y=1, junior staff, or y= 2, middle management), $P_1 + P_2$ are higher for higher values of predictor variables, A, G, Q. Ind. and E since all our estimated parameter coefficients β_i's are all positive. The predicted probabilities P_1 as well as cumulative probability, $P_1 + P_2$ are higher for predictor variables Educational Qualification and previous working experience due to their higher values which can be seen in the Wald's column from Table 5. For model adequacy, the SPSS version 14.0 gave that the model to be significant, meaning the fitted model provides a good fit and the Pearson and Deviance statistics gave similar result.It is important to weigh the strengths and the weaknesses of the ordinal regression model so as to decide to employ this method if the strengths outweigh the weaknesses.

STRENGTHS

The strengths of the ordinal regression model in this study are briefly described. First, many indicators concerning individual staff outcome are frequently measured on an ordinal scale. For instance, occupational levels perceived by civil servants on a Likert scale, (e.g., Junior staff, Middle management staff, and Senior management staff) are most appropriately measured by an ordinal scale. Thus, the ordinal regression model seems to have a broad marketplace to analyze diverse civil servant status outcomes. Second, comparable to linear and logistic regression models, ordinal regression model can be used to perform the following tasks: (1) to identify significant explanatory variables that influence on the ordinal outcome; (2) to describe the direction of the relationship between the ordinal outcome and the explanatory variables; and (3) to perform classifications for all levels of the ordinal outcome, and subsequently evaluate the predict validity of the regression model. Third, various link functions such as logit and cloglog links are readily available to model the effect of the explanatory variables on the ordinal outcome. Fourth, the test of parallel lines can be easily used to assess the validity of the model assumption, and the model fitting statistics (e.g., -2log likelihood ratio and pseudo R squares) can be used as criteria to screen the occupational models and choose the most appropriate one. Finally, the model assumes that the relationship between the ordinal outcome and the explanatory variables is independent of the category. This assumption implies that the corresponding regression coefficients in the link function are equal for each cut-off point (Bender and Benner, 2000). Therefore, it is easy to construct and interpret the ordinal regression model, which requires only one model assumption, and produces only one set of regression coefficients.

WEAKNESSES

Researchers need to be aware of the limitations in using ordinal regression model. The large percent of cells with missing data could lead to a decrease of actual sample size for the model construction or an inaccurate chi-square test for the model fitting. Note that the model goodness-of-fit is usually dependent on chi-square test result. The chi-square test normally depends on the sample size. Hence, if number of cells with a zero value is large, the chi-squared goodness of fit statistics may not be appropriate (Agresti, 1990). Thus, researchers are limited in how well they can assess the model goodness of fit. In addition, the logit link and clog log link in the ordinal regression analysis are not capable of selecting a subset of significant explanatory variables by means of automatic model building methods such as stepwise and back elimination procedures in SPSS command language. Therefore, researchers are obliged to rely on their own intuition and experiences to select a subset of the important or significant explanatory variables in the model. As a result, much of the time and energy is devoted to developing occupational models, checking the model assumptions, and assuring the model goodness of fit.

CONCLUSION

This study used ordinal logistic regression method to examine the relationship between the ordinal outcome variable, different levels of staff status in the Lagos State Civil Service of Nigeria, the explanatory variables are Gender, Indigenous status, Educational Qualification, Previous Experience and Age. The outcome variable was measured on an ordered, scale as Junior staff, Middle Management staff, and Senior Management staff. Within the complete models, the logit link was better because of its satisfying parallel lines assumption and larger model- fitting statistics. Education Qualification and Previous Working Experience significantly predicted the probability of an individual staff being a member of any of the three levels of staff status. The ordinal regression technique provides a viable alternative to analyze the ordinal outcome. It does not alter an ordinal outcome as binary or dichotomous measure for logistic regression analysis, which may lead to the loss of information inherent. Also, it does not falsely assume continuous measure and the properties of normality and constant variance for linear regression to analyze few categories of ordinal outcome, which may lead to incorrect analysis. Obviously, the ordinal regression modeling is a unique statistical technique in that the ordinal outcome is frequently encountered in the field of education, labour force and health and the model assumption of parallel lines is easily verified.

REFERENCES

Agresti Alan (1990): An Introduction to Categorical Data Analysis. John Wiley &Sons.

Agresti Alan (2002): Categorical Data Analysis, 2nd ed., New York: Wiley- Inter science.

Bender, R. and Benner A (2000): Calculating Ordinal Regression Models in SAS and S-Plus. Biometrical Journal 42, 6,677-699.

Chau-Kuang Chen, and John Hughes, Jr., (2001): Using Ordinal Regression model to Analyze student Satisfaction Questionnaires, Meharry Medical College.

Hosmer, David W. and Lemeshow, (1989): Applied Regression. John Wiley & Sons, New York.

Hosmer, David W. and Stanley Lemeshow (2002): Applied Logistic Regression, 2nd ed., New York, Chichester, Wiley.

Lehmann, E.L. (1988): "Two Fisher Lectures on Statistical Modeling: Model Specification." Vol 5. No.2, P 159-168.

Maria Halena Spyrides-Cunha (1998): "Proportional Odds Model Applied to Mapping of Disease Resistance Genes in Plants". Genet. Mol. Biol. Vol. 23 No.1.

McCullagh P. (1980): Regression Models for ordinal Data (with Discussion), Journal of the Royal Statistical Society, Series B 42, 109-142.

McCullagh P and Nelder (1989): A Generalized Linear Models. 2nd ed. Chapman and Hall, New York.

McFadden D. (1973): "Conditional Logit Analysis of Quantitative Behaviour." Frontiers of Econometrics, Academic Press, New York.

Nelder J.A. and Wedderburn, R.W.M. (1972): Generalized Linear Models. Journal of the Royal Statistical Society, Series A, Part 3:135: 370-84.

Plank, Stephen B. and Jordan, Will J. (1997): Reducing Talent Loss. The Impact of Information, Guidance, and Actions on Postsecondary Enrolment, Report No. 9 Eric No: ED405429

Rubinfeld D. L (1977): "Voting in a Local School Election: A Micro Analysis", Review of Economics and Statistic, Vol. 59 No.1, pp39 – 42.

SPSS, Inc. (2002): Ordinal Regression analysis, SPSS Advanced model 14.0. Chicago, IL.