Linear Possibility Model with Interactions for Ordered Categorical Data

Amin I. Adam

Abstract

The concern of this research is focussed with extending the idea of the linear possibility model for ordered categorical data to include the interactions between the categorical regressor variables. Besides the main differential effects of the categorical regressor variables, the model allow the differential effect for each level of variable to vary within the levels of other variables. This research shows how to estimate this type of model, how to resolve the problems surrounding it, and how to interpret it in a simple and straightforward way. The study also shows how to check and diagnose the estimated models using cross-validations, outliers and influential observations, and other tests. The application data for this research are collected from a random sample of students at the Omdurman Islamic University. The ordered response categorical variable for the study is the academic performance of students, which is assumed to be associated with three categorical variables and their interactions. These variables are: the specialization of the students, whether the students live with their families or not, and the educational level of their guardians. The results showed that the students whose their guardians have an intermediate level of education perform academically better when their specialization is social science and live with their families, but they seem to perform academically less than other students when they don't live with their families and their specialization is not social science. Regardless of the educational level of the guardians, all students appear to perform academically less when their specialization is social science and live with their families or just being living with their families. When they do not live with their families, their academic performance, however, seems to be the same regardless of their specialization (social or natural sciences). The data of the study are analysed by SPSS (Statistical Package for the Social Sciences) and Minitab.

Key Words: **Linear possibility model, Linear models, Regression analysis, Categorical data**.

Associate Professor in Applied Statistics, Dept. of Statistics, Faculty of Economics and Political Sciences, Omdurman Islamic University, Omdurman, P. O. Box 382, Sudan

النموذج الراجح الخطي بتفاعلات للبيانات الفئوية المرتبة

المستخلص

هذا البحث بتعميم مفهوم النموذج الراجح الخطي لتحليل البيانات الفئوية المرتبة ليشمل التفاعلات بين المتغيرات الفئوية المستقلة. فبجانب آثار الفروق الرئيسة للمتغيرات الفئوية المستقلة فإن النموذج الآن يسمح بآثار الفروق لكل مستوى لأي متغير بأن تتغير داخل مستويات المتغيرات الأخرى. يوضح البحث كيف يمكن تقدير النموذج لهذا النوع من النموذج، وكيف يمكن التغلب على المشاكل التي يمكن أن تصاحب هذا النموذج، وكيف يمكن تفسير النتائج بشكل مباشر. كما يوضح البحث كذلك كيف يمكن فحص واختبار النموذج المقدر باستخدام اختبارات الشرعية المشتركة وبفحص القيم المتطرفة والمفردات المؤثرة وغير ذلك من اختبارات. أعطت الدراسة تقديرات لأثر العلاقة بين المتغيرات محل الدراسة كما أوضحت التفسيرات المقابلة لطبيعة كون هذه المتغيرات الفئوية مرتبة. وكانت البيانات قد جمعت من عينة عشوائية من طلاب جامعة أم درمان الإسلامية. وكان المستوي التحصيل الأكاديمي بمثابة المتغير الفئوي التابع في الدراسة، ويرتبط بثلاثة من المتغيرات الفئوية وتفاعلاتها. وهذه المتغيرات هي: مساق الطلاب، إذا كان الطلاب يعيشون مع أسرهم أم لا، والمستوى التعليمي لأولياء أمورهم. أوضحت الدراسة أن الطلاب الذين يكون المستوى التعليمي لأولياء أمورهم متوسطاً يكون أداؤهم الأكاديمي أفضل عندما يكون مساقهم الأكاديمي علوم اجتماعية ويعيشون مع أسرهم، ويكون أداؤهم أقل من الذين لا يعيشون مع أسرهم والذين يكون مساقهم الأكاديمي غير العلوم الاجتماعية. وبغض النظر عن المستوى التعليمي لأولياء أمور الطلاب يكون أداء الطلاب الأكاديمي أقل عندما يكون مساقهم الأكاديمي علوم اجتماعية ويعيشون مع أسرهم أو فقط يعيشون مع أسرهم. وعندما يكون الطلاب لا يعيشون مع أسرهم يكون أداؤهم الأكاديمي متساوياً بغض النظر عن مساقهم الأكاديمي (علوم اجتماعية أو طبيعية). هذا وقد تم استخدام الحزمة الإحصائية للعلوم الاجتماعية SPSS وحزمة Minitab للمساعدة في تحليل البيانات.

1- Introduction:-

In the usual probit model the response variable is binary (zeros and ones) and. hence, the conditional expected values for the response variable are assumed not to lie outside the range of 0-1. In linear possibility model (LPM), however, the response variable for ordered categorical data is having a wider range of values (frequently more than two responses), and, consequently, the conditional expected values for this response given a number of factors will obviously lie outside the range of 0-1. Even if the responses are zeros and ones there is no guarantee that the responses will lie within the range of 0-1.

For a set of categorical variables, let us consider a response ordered categorical variable Y to be dependent on a number of k other categorical variables, ordered or otherwise. We further assume the response variable to have categories being ordered (ascendingly) as 1,2,...,c. For the c_1 categories of the first categorical regressor variable, c_2 categories of the second categorical regressor variable, ..., and c_k categories of the kth categorical regressor variable, we can write a linear possibility model for Y as

$$\begin{aligned} Y_i = \beta_0 + \beta_1^1 X1_i^1 + \beta_1^2 X1_i^2 + \ldots + \beta_1^{c_1-1} X1_i^{c_1-1} + \beta_2^1 X2_i^1 + \beta_2^2 X2_i^2 \\ + \ldots + \beta_2^{c_2-1} X2_i^{c_2-1} + \ldots + \beta_k^1 Xk_i^1 + \beta_k^2 Xk_i^2 + \ldots + \\ \beta_k^{c_k-1} Xk_i^{c_k-1} + u_i,\ i = 1,2,\ldots,n \end{aligned} \quad (1)$$

Where

$$Y_i = 1,2,\ldots,c \text{ for the ordered categories of the response categorical variable}$$

$$X1_i^1 = \begin{cases} 1, & \text{for the first category of the first categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$X1_i^2 = \begin{cases} 1, & \text{for the second category of the first categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$X1_i^{c_1-1} = \begin{cases} 1, & \text{for the } (c_1 - 1) \text{ category of the first categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$X2_i^1 = \begin{cases} 1, & \text{for the first category of the second categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$X2_i^2 = \begin{cases} 1, & \text{for the second category of the second categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$X2_i^{c_2-1} = \begin{cases} 1, & \text{for the } (c_2 - 1) \text{ category of the second categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$Xk_i^1 = \begin{cases} 1, & \text{for the first category of the kth categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$Xk_i^2 = \begin{cases} 1, & \text{for the second category of the kth categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

$$Xk_i^{c_k - 1} = \begin{cases} 1, & \text{for the } (c_k - 1) \text{ category of the kth categorical regressor variable} \\ 0, & \text{otherwise} \end{cases}$$

u_i = the error term

In this model, the conditional expectation of Y given the k categorical variables is interpreted as the conditional possibility that the specific event of Y will occur. The parameter β_1^1 measures the differential effect for the first category of the first categorical variable, compared with otherwise. Likewise, β_1^2 stands for the differential effect for the second category of the first categorical, compared with otherwise. Similarly, all other parameters measure the corresponding differential effects for the specific categories of the categorical variables. The parameter β_0 represents the expected possibility of Y for situation where all the differential effects of the regressed categorical variables are absent. Any other combination of categories of the regressed categorical variables can be chosen to be represented by β_0. The differential effects would then be measured from that combination. However, the numerical variables values of the conditional means will be the same regardless of the chosen position.

For three categorical regressor variables, in particular, with the first two variables being binary (1,0) and the third one being ordinal having four categories (1,2,3, and 4), we have a complete set of all the conditional means for equation(1) as in table(1).

Table(1): The Expected Values Of Y For Model(1) with Three Assumed Categorical Regressor Variables.

X1	X2	X3	The Expected Values Of Y
1	1	1	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^1$
1	1	2	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^2$
1	1	3	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^3$
1	1	4	$\beta_0 + \beta_1^1 + \beta_2^1$
1	0	1	$\beta_0 + \beta_1^1 + \beta_3^1$
1	0	2	$\beta_0 + \beta_1^1 + \beta_3^2$
1	0	3	$\beta_0 + \beta_1^1 + \beta_3^3$
1	0	4	$\beta_0 + \beta_1^1$
0	1	1	$\beta_0 + \beta_2^1 + \beta_3^1$
0	1	2	$\beta_0 + \beta_2^1 + \beta_3^2$
0	1	3	$\beta_0 + \beta_2^1 + \beta_3^3$

0	1	4	$\beta_0 + \beta_2^1$
0	0	1	$\beta_0 + \beta_3^1$
0	0	2	$\beta_0 + \beta_3^2$
0	0	3	$\beta_0 + \beta_3^3$
0	0	4	β_0

2. The LPM With Interactions

If we could expand model(1) to allow the differential effect of each level of variable to vary within the levels of other variables, then we have

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1^1 X1_i^1 + \beta_1^2 X1_i^2 + \ldots + \beta_1^{c_1-1} X1_i^{c_1-1} + \beta_2^1 X2_i^1 + \beta_2^2 X2_i^2 \\ &+ \ldots + \beta_2^{c_2-1} X2_i^{c_2-1} + \ldots + \beta_k^1 Xk_i^1 + \beta_k^2 Xk_i^2 + \ldots + \\ &\beta_k^{c_k-1} Xk_i^{c_k-1} + \beta_{12}^{11} X1_i^1 * X2_i^1 + \beta_{12}^{12} X1_i^1 * X2_i^2 + \\ &\ldots + u_i, \ i = 1,2,\ldots,n \end{aligned} \qquad (2)$$

For the three categorical regressor variables, which mentioned in the previous section, with the first two variables being binary (1,0) and the third one being ordinal having four categories (1,2,3, and 4), the LPM with interactions can be expressed as

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1^1 X1_i^1 + \beta_2^1 X2_i^1 + \beta_3^1 X3_i^1 + +\beta_3^2 X3_i^2 + \beta_3^3 X3_i^3 \\ &+ \beta_{12}^{11} X1_i^1 * X2_i^1 + \beta_{13}^{11} X1_i^1 * X3_i^1 + \beta_{13}^{12} X1_i^1 * X3_i^2 \\ &+ \beta_{13}^{13} X1_i^1 * X3_i^3 + \beta_{23}^{11} X2_i^1 * X3_i^1 + \beta_{23}^{12} X2_i^1 * X3_i^2 \\ &+ \beta_{23}^{13} X2_i^1 * X3_i^3 + \beta_{123}^{111} X1_i^1 * X2_i^1 * X3_i^1 \\ &+ \beta_{123}^{112} X1_i^1 * X2_i^1 * X3_i^2 \\ &+ \beta_{123}^{113} X1_i^1 * X2_i^1 * X3_i^3 \\ &+ u_i, \ i = 1,2,\ldots, n \end{aligned} \qquad (3)$$

The conditional expected values, shown in the last column of table(1), are now adjusted in table(2) to accommodate the interactions effects.

Table(2): The Expected Values Of Y For Model (3) with Three Assumed Categorical Regressor Variables.

X1	X2	X3	The Expected Values Of Y
1	1	1	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^1 + \beta_{12}^{11} + \beta_{13}^{11} + + \beta_{23}^{11} + \beta_{123}^{111}$
1	1	2	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^2 + \beta_{12}^{11} + \beta_{13}^{12} + + \beta_{23}^{12} + \beta_{123}^{112}$
1	1	3	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^3 + \beta_{12}^{11} + \beta_{13}^{13} + + \beta_{23}^{13} + \beta_{123}^{113}$
1	1	4	$\beta_0 + \beta_1^1 + \beta_2^1 + \beta_{12}^{11}$
1	0	1	$\beta_0 + \beta_1^1 + \beta_3^1 + \beta_{13}^{11}$
1	0	2	$\beta_0 + \beta_1^1 + \beta_3^2 + \beta_{13}^{12}$
1	0	3	$\beta_0 + \beta_1^1 + \beta_3^3 + \beta_{13}^{13}$
1	0	4	$\beta_0 + \beta_1^1$
0	1	1	$\beta_0 + \beta_2^1 + \beta_3^1 + \beta_{23}^{11}$
0	1	2	$\beta_0 + \beta_2^1 + \beta_3^2 + \beta_{23}^{12}$
0	1	3	$\beta_0 + \beta_2^1 + \beta_3^3 + \beta_{23}^{13}$
0	1	4	$\beta_0 + \beta_2^1$
0	0	1	$\beta_0 + \beta_3^1$
0	0	2	$\beta_0 + \beta_3^2$
0	0	3	$\beta_0 + \beta_3^3$
0	0	4	β_0

3- The LPM & The Ordinary and Weighted Least Squares Methods

The form of the LPM in equation(2) or equation(3) can be written in a matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U} \quad (4)$$

where **Y** represents the column vector of n observations on the response variable Y, **β** is the column vector of k unknown parameters, **X** gives the nxk matrix of observations of the categorical regressor variables (X1, X2, ..., Xk.) with the first column of 1's representing the intercept term, and **U** accommodates the vector of the disturbances u_i. Model(4) then looks like the usual regression model and hence can be estimated by the OLS method. It is simple to apply the OLS method to model(4) but obviously this will be surrounded by some technical and logical problems. These special problems are: nonnormality of the error term u_i, heteroscedasticity, and the possibility of the estimated response outcome lying outside the bounds (1 to c for our response Y).

The OLS does not require the error term u_i to be normally distributed, but it is implicitly assumed so for the purpose of statistical inference. That is, hypothesis testing, prediction, etc. Obviously, the assumption of normality for u_i is hard to attain for an LPM because like the response variable, u_i takes on
only discrete values. So, it cannot be assumed to be normally distributed; in fact it follows a multinomial distribution. However, the violation of the normality assumption may not be as serious as it looks because we know that the OLS point estimates still remain unbiased (Johnston and DiNardo (2001)). Furthermore, and based on the central limit theorem, as the sample size increases, the OLS estimators tend to be normally distributed generally, (Gujarati(2004)).

The other problem with the error terms u_i is that their variances are no longer equal, i.e., heteroscedastic, even though $E(u_i)=0$ and $E(u_iu_j)=0$ for $i \neq j$ (no serial correlation). To see this, we take the mathematical expectation for equation(4) and, accordingly, u_i can be written as

$$u_i = Y_i - E[Y_i \mid X1_i, X2_i, \ldots, Xk_i] \quad (5)$$

and so the distribution of u_i looks as

u_i	Possibilities of Values of u_i
$1 - E[Y_i \mid X1_i, X2_i, \ldots, Xk_i]$	π_1
$2 - E[Y_i \mid X1_i, X2_i, \ldots, Xk_i]$	π_2
…	...
$c - E[Y_i \mid X1_i, X2_i, \ldots, Xk_i]$	π_c

with π_1, π_2, ..., and π_c are the possibilities of obtaining the corresponding values of u_i (same as the probability of obtaining, respectively, the values of 1, 2, ..., and c for Y_i. Therefore, the variance of u_i becomes

$$\begin{aligned} Var(u_i) &= E[u_i - (E(u_i))^2] = (E(u_i))^2 \\ &= \pi_1(1-M)^2 + \pi_2(2-M)^2 + \ldots + \pi_c(c-M)^2 \quad (6) \end{aligned}$$

where $E(u_i)=0$, by assumption, and M is $E[Y_i \mid X1_i, X2_i, \ldots, Xk_i]$. This variance depends on the conditional expectation of Y, which, of course, depends on the values taken by X1, X2,..., and Xk. Thus, ultimately the variance of u_i is heteroscedastic.

Again, the problem of heteroscedasticity is not insuperable; and even with its presence the OLS estimators are still unbiased, though not efficient (Johnston(1997)). One way of resolving the heteroscedasticity problem is to transform all the variables by dividing both sides of models (4) by

$$\sqrt{Var(u_i)} = \sqrt{w_i} \quad (7)$$

and according to Johnston(1997), Draper and Smith(1998), Maddala and Lahiri(2009) and others, we can proceed in two steps. The first step is by running the OLS method, despite the heteroscedasticity, to obtain the fitted values, which are the estimates for $E[Y_i | X1_i, X2_i, ..., Xk_i]$, and hence w_i. The second step is to use the estimated w_i to transform the variables and run the OLS method.

Before we apply these two steps, we need to mention that the LPM still faces a logical problem. This is where there is no guarantee that the fitted values of Y_i will lie within the limits (1 to c in our case), despite a priori that the conditional possibility $E[Y_i | X1_i, X2_i, ..., Xk_i]$ must fulfil this restriction. What we could do to overcome this obstacle is to estimate the LPM by the usual OLS method and to find out whether the fitted values lie between the bounds (of 1 and c). If some are less than 1, the fitted values are considered to be 1 for those cases; and if they are greater than c, they are assumed to be equal c.

4- Variables Selection Procedures & The Optimal Number Of Regressors

To avoid the unnecessary computation of all possible models, there are several procedures to identify subsets of variables that are best predictors of the dependent variables. Among these procedures are: the `stepwise' procedure, the `forward' selection, and the `backward' elimination. The stepwise method starts by entering the most significant variable in the model, i.e., the one with the smallest p-value for the associated t-test. The model is then checked for the overall significance. If the probability associated with the F-test is greater than the predetermined significance level (5% in our case), the procedure terminates. Otherwise, the variable is retained and the subsequent variable with the smallest p-value for the t is the next candidate. Variables in the model are checked for the removal and process of entering and checking continues until no more variables to be entered and/or removed.

The forward selection, on the other hand, starts with no variables in the model and then adds sequentially the most significant variables, whereas the backward elimination begins with all the variables in the model and then removes sequentially the least significant variables. The three procedures do not always end up with the same selected model but they would reinforce each other when they do.

The model to be selected, however, can be too simple, and hence might suffer from biased coefficients and prediction, or too complicated, which might have large variances for the coefficients and prediction. So Mallows' C_k criterion, suggested by Mallows(1973), is used in order to minimize the under fitting (having the important variables being left out) or overfitting (including variables that make no essential contribution) of a selected model. The criterion is given by

$$C_k = k + \frac{(s^2 - \hat{\sigma}^2)(n - k)}{\hat{\sigma}^2} \quad (8)$$

where s^2 is the residual variance for the selected model and σ^2 is an estimated residual variance for the full general model. Mallows recommended to choosing a model where $C_k \approx k$ (i.e., a model with smallest difference between C_k and k).

5. Cross Model Validation

Cross-validation is a method of evaluating given models by means of their predictions and to choose a model with the minimal error. We use here two forms of model validations: the `data splitting' form and the `leave one out' form. In the data splitting form, the whole data set is to be randomly split into two subsets[1]: the `estimation sample' and the `test sample'. The LPM is to be

[1]Although we let this split be even in this study, however, it does not necessarily be so.

carried out on the estimation sample and then applied to the test sample to forecast the values of the dependent variable **Y** there. In the leave one out form the **Y** value for each case is set aside and the LPM is to be estimated on the remaining (n-1) data points. The prediction is then made for the case which was left out. Thus n prediction equations are derived and n **Y** values are predicted.

If y_i is a prediction of y_i, then we let $L(y_i, y_i)$ to be the loss function (Hjorth(1994)). Accordingly, we define the cross-validation error rate (CV_{ER}) for the splitting data form as

$$CV_{ER} = \frac{1}{m} \sum_{i=1}^{m} L(y_i, \hat{y}_i) \quad (9)$$

where m is the size of the test sample and

$$L(y_i, \hat{y}_i) = \begin{cases} 0 \text{ if } y_i = \hat{y}_i \\ 1 \text{ if } y_i \neq \hat{y}_i \end{cases} \quad (10)$$

with the predicted values of y_i to be rounded to the nearest integer value. For the leave one out form, the cross-validation error rate (CV_{ER}) equals

$$CV'_{ER} = \frac{1}{n} \sum_{i1}^{n} L(y_i, \hat{y}_{-i}) \quad (11)$$

where y_{-i} is the (rounded) predicted value for subject i when it was not used in estimating the LPM and

$$L(y_i, \hat{y}_{-i}) = \begin{cases} 0 \text{ if } y_i = \hat{y}_{-i} \\ 1 \text{ if } y_i \neq \hat{y}_{-i} \end{cases} \quad (12)$$

If we square the difference between y_i and y_i (and between y_i and y_{-i}), alternatively, we get what Maddala and Lahiri(2009) and others called the PRESS (predicted sum of squares) which is given by

$$PRESS = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \quad (13)$$

for the data splitting form and by

$$PRESS' = \sum_{i=1}^{n} (y_i - \hat{y}_{-i})^2 \quad (14)$$

for the leave one out form. Dividing the PRESS value by the corresponding sample size gives the cross-validation index (CV_I), and so

$$CV_I = \frac{PRESS}{m} \quad (15)$$

and

$$CV'_I = \frac{PRESS'}{n} \quad (16)$$

for the data splitting form and the leave one out form, respectively.

The preferred model, therefore, is the one with the smallest cross-validation error rate and/or cross-validation index.

6- Fitting The LPM For the Academic performance of Students

We consider the data of our study which are collected from a randomly selected sample of 182 students from the Omdurman Islamic University. The data are cross-classified according to the academic performance of the students, which is considered as a response, and three other categorical variables. The classification of the academic performance is based on the cumulative rate out of 5, that is: a rate more than or equals 4.00 is considered "superb", a rate more than or equals 3.00 is considered "good", a rate more than or equals 2.00 is considered "fair", a rate more than or equals 1.00 is considered "weak", and a rate less than 1.00 is considered " very weak". The categories are coded as 1, 2, 3,4, and 5 for the 'very weak', 'weak', 'fair', 'good', and 'superb', respectively. The other three categorical variables are: the specialization of the students (social or natural sciences), whether the student lives with his family or not, and the education level of the guardian (primary or lower, intermediate, secondary, and university or higher level). It is clear that the categories of the academic performance are ordered and so the categories of the educational level of the guardian are ordered too. So, according to model(1), we have

$X1^1$ represents the situation where the specialization of the student is social science

$X2^1$ represents the situation where the student lives with his family

$X3^1$ represents the situation where the educational level of the guardian is university or higher

$X3^2$ represents the situation where the educational level of the guardian is secondary

$X3^3$ represents the situation where the educational level of the guardian is intermediate

In applying the OLS method to the LPM in equation(1), we obtain the results in table(3). The overall statistical significance of the model is indicated by the F-value of 2.25 which has a p-value of 0.051. And according to the partitioning of the regression sums of squares of the analysis of variance, most of the contribution came from $X3^1$ (56.6%) and $X2^1$ (42.8%). The coefficient of determination, R^2, indicates that only 6.0% of the variations in Y is explained by the variations in the regressor variables all together. The intercept of 3.782 indicates that, ignoring all the independent variables, the estimated value for Y would be 3.782 (which is roughly `good'). Along with the intercept coefficient, the parameters related to X11 and X32 are statistically significant at the 5% level (according to the attached t-values and their corresponding p-values). On the other hand, parameters related to X21, X31 and X33 turned out to be nonsignificant.

Table(3): The OLS Estimated LPM Of Model(1) For Y On X11, X21, X31, X32, And X33; Minitab Output.

Ignoring all other variables, the differential effect of X11, which is (-0.363), shows that a unit increase in X11 results in a decrease of 0.363 in the value of Y. This means that the academic performance of students tends to be less by 0.363, on average, for the students whose specialization is social science than those specialization is natural science. On the other hand, the coefficient of 0.457 attached to the variable X32 means, holding all other variables constant, the academic performance of students is higher by 0.457, on average, for the students whose the educational level of their guardians is secondary compared with whose the educational level of their guardians is primary or lower.

The expected possibility of Y for the social science specialization of students, for those whose who live with their family, and for those whose the educational level of their guardian is university or higher is 3.504 ($\beta_0 + \beta_1^1 + \beta_2^1 + \beta_3^1$). On the other hand, the expected possibility of

Y for the natural science specialization of students, for those whose who do not live with their

```
The regression equation is
Y = 3.782 - 0.363 X11 - 0.132 X21 + 0.217 X31
            + 0.457 X31 + 0.018 X33
Predictor      Coef        Stdev     t-ratio         p
Constant     3.7816       0.2073       18.24     0.000
X11         -0.3631       0.1495       -2.43     0.016
X21         -0.1320       0.1541       -0.86     0.393
X31          0.2172       0.2412        0.90     0.369
X32          0.4565       0.2285        2.00     0.047
X33          0.0180       0.2302        0.08     0.938

s = 0.9869      R-sq = 6.0%      R-sq(adj) = 3.3%

Analysis of Variance
SOURCE        DF          SS        MS        F       p
Regression     5     10.9619    2.1924     2.25   0.051
Error        176    171.4118    0.9739
Total        181    182.3736

SOURCE        DF      SEQ SS
X11            1      4.6951
X21            1      0.0568
X31            1      0.0000
X32            1      6.2040
X33            1      0.0060
```

families, and for those whose the educational level of their guardian is secondary appears to be as 4.239 $(\beta_0 + \beta_3^2)$. The complete expected possibility set of Y for the different categories of the three categorical regressor variables are given in table(4).

Table(4): The Expected Possibility Values Of Y for the Different Categories of the Three Categorical Regressor Variables.

Educational Level of the Guardian	Living with the Family			
	Yes		No	
	Specialization		Specialization	
	Social	Natural	Social	Natural
University +	3.504	3.867	3.636	3.999
Secondary	3.744	4.107	3.876	4.239
Intermediate	3.305	3.668	3.437	3.800
Primary -	3.287	3.650	3.419	3.782

The basic question, however, before accepting these estimates and the above model is: Can we trust the estimated standard errors (and hence the t-values) reported in table(3)? The answer depends generally on the existence or otherwise of heteroscedasticity. The estimated variances of u_i, given by equation(11), are all found concentrated around 1 (the largest is 1.1149 and the smallest is 1.0011).
It makes no difference dividing the variables by the square root of this variance and hence there is no need to bother about or to correct for the heteroscedasticity. To see this, table(5) gives the weighted least squares (WLS) using the two-stage procedure outlined earlier.

Table(5): The Weighted Least Squares For Model(1) For Y On X11,

```
 The regression equation is
 Y' = 3.792'- 0.367 X11' - 0.143 X21' + 0.214 X31'
          + 0.470 X32' + 0.024 X33'

 Predictor        Coef        Stdev    t-ratio         p
 constant'      3.7924       0.2073      18.29     0.000
 X11'          -0.3668       0.1511      -2.43     0.016
 X21'          -0.1428       0.1551      -0.92     0.359
 X31'           0.2143       0.2415       0.89     0.376
 X32'           0.4701       0.2315       2.03     0.044
 X33'           0.0242       0.2308       0.11     0.916

 s = 0.9607

 Analysis of Variance
 SOURCE         DF           SS         MS         F        p
 Regression      6      2436.49     406.08    439.97    0.000
 Error         176       162.44       0.92
 Total         182      2598.93

' denotes that the variable is divided by the weight √w
```

X21, X31, X32, And X33; Minitab Output.

For this model, the sums of squares are not adjusted from the means and so, the F-value seems to be inflated accordingly. However, the parameter estimates and their standard errors (and accordingly the t and p-values) along with the standard errors of the estimates thus obtained do not differ substantially from those obtained without the correction for heteroscedasticity. Accordingly, we retain the OLS estimates given in table(3).

The model in table(3), however, is still not convincing. The coefficients related to X21, X31, and X33 are all far from being important in the model. Including unimportant variables increases the standard errors of all estimates without improving prediction. We use the variables selection procedure to build a concise model that includes, potentially, the important variables.

The procedures of stepwise, forward, and backward elimination all reached the same conclusion and selected the model that includes X11 and X32 as a recommended model. The model is shown in table(6). It has an ample observed significance for the overall F-test (with a p-value of 0.008). The standard error for the coefficient related to X32 is now much lower compared with that in table(3) and, as a consequence, the corresponding t-values are now higher.

Table(6): The Final (OLS) Output Adopted By The Stepwise, Forward, And Backward Procedures For Model(1) Of Y On X11, X21, X31, X32, And X33; SPSS Output.

```
Multiple R            .22781
R Square              .05190
Adjusted R Square     .04130
Standard Error        .98284

Analysis of Variance
                     DF      Sum of Squares      Mean Square
Regression            2             9.46461          4.73230
Residual            179           172.90902           .96597

F =       4.89901       Signif F =  .0085

---------------- Variables in the Equation ----------------
Variable            B         SE B        Beta          T    Sig T
X32           .348412    .156797     .162181      2.222   .0275
X11         -.347889    .147014    -.172714     -2.366   .0190
(Constant)    3.791050    .107661                 35.213   .0000
```

The estimated R^2 of 5.2% is seen pretty low and this might look to contradict the F-value (which in turn tests the significance of the R^2). The fact is, however, the R^2 is unlikely to be high in the case of the LPM, since the response variable takes only limited values and the scatter plot of this variable with any of the independent variables is expected to be concentrated on those limited values and this results in low partial correlations and, accordingly, low multiple correlation. It is not surprising, therefore, to see the R^2 in our case as much low as 5.2% and to be significant at the same time.

The conditional possibility of Y, based on this refined model, is lower by 0.348 for all the students whose specialization is social science. Likewise, the conditional possibility of Y is higher by 0.348 for all the students whose the educational level of their guardian is secondary. For those whose the educational level of their guardian is secondary and their specialization is social science, the increment in the conditional possibility is almost 0. For those whose their specialization is natural science (i.e., when ignoring the coefficient of X11) and the educational level of their guardian is primary or lower, the conditional possibility of Y is 3.791.

7- Fitting the LPM with the Interactions

If, however, we are no longer assuming that the differential effect of each variable is constant across the levels of the other variables, then we need to fit model(4), where the interactions effects are incorporated. The estimated OLS equation along with the relevant details are shown in table(7).

Table(7): The OLS Estimates For Model (4) For Y On X1, X2 And X3 With Their Interactions; MINITAB Output.

Predictor	Coef	Stdev	t-ratio	p
Constant	3.6250	0.3366	10.77	0.000
X11	-0.2917	0.4626	-0.63	0.529
X21	0.3750	0.4927	0.76	0.448
X31	-0.1250	0.5827	-0.21	0.830
X32	-0.0250	0.4515	-0.06	0.956
X33	0.3750	0.3907	0.96	0.339
X11*X21	-0.5083	0.7243	-0.70	0.484
X11*X31	0.0139	0.7357	0.02	0.985
X11*X32	1.3167	0.6464	2.04	0.043
X11*X33	0.1667	0.6055	0.28	0.783
X21*X31	0.1250	0.7229	0.17	0.863
X21*X32	0.2882	0.6173	0.47	0.641
X21*X33	-0.9135	0.5932	-1.54	0.125
X11*X21*X31	0.5553	0.9874	0.56	0.575
X11*X21*X32	-1.1608	0.9052	-1.28	0.202
X11*X21*X33	-0.2032	0.9275	-0.22	0.827

s = 0.9519   R-sq = 17.5%   R-sq(adj) = 10.1%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	15	31.9430	2.1295	2.35	0.004
Error	166	150.4306	0.9062		
Total	181	182.3736			

Compared with the previous fitting of table(3) where the effect due to X1 is constant -0.363, this effect depends now on conditions of X2 and X3. If we ignore the effect of X3, the X1 effect is -0.800 for the `yes' of X2 and -0.292 for the `no' of it. This means, regardless of the educational level of the guardian, the possibility Y is less by 0.508, on average, for those whose their specialization is social science when they live with their families than when they don't. If we ignore the effect of X2 instead, the effect due to X1 changes to -0.2778 for those whose the educational level of the guardian is university or higher, to 1.025 for those whose the educational level of the guardian is secondary, and to -0.125 for those whose the educational level of the guardian is intermediate. For those who live with their families, the effect due to X1 becomes: slightly smaller (-0.2308) for those whose the educational level of the guardian is university or higher, almost double (-0.6441) for those whose the educational level of the guardian is secondary, and, once more, is worse (-0.8365) for those whose the educational level of the guardian is intermediate.

The expected possibility values of Y, under all the circumstances of X1, X2, and X3, are shown in table(8). These are now more varied (range from 2.625 to 4.625) compared with those in table(4) which were concentrated around the values of 3 and 4.

Table(8): The Expected Possibility Values Of The Interaction Model(4) Of Y For All The Levels Of X1, X2, And X3.

X3	X1			
	1		0	
	X2		X2	
	1	0	1	0
1	3.769	3.222	4.000	3.500
2	3.619	4.625	4.263	3.600
3	2.625	3.875	3.642	4.000
4	3.200	3.333	4.000	3.625

The estimated model in table(7), however, suffers from the problem of heteroscedasticity, which can be seen readily from formula(4). As a consequence, we use the WLS to correct for heteroscedasticity. The WLS regression is given in table(9).

Table(9): The WLS Of The Interaction Model(4) Of Y On X1, X2 And X3; MINITAB Output.

Compared with the result obtained without the correction for the heteroscedasticity, the standard errors of estimates is now smaller. The model is now highly significant according to the related value of F-ratio2 (175.33), and so the hypothesis that the partial slope coefficients are simultaneously equal to zero is to be rejected. However, the t-values corresponding to all the levels and interactions are not statistically significant (all p-values are greater than 0.05). Hence, it looks as if there is some sort of linear relationship among the explanatory variables, i.e., multicollinearity. To correct for multicollinearity, we use, once again, the variables selection procedures to drop the collinear variable(s) and to keep the statistically significant one(s).

This time the backward elimination has recommended a different model from the one selected by both the stepwise and the forward procedures. Table(10), shows the model that selected using the stepwise and the forward procedures. Besides the constant term, this model looks simple and consists of the third order-interaction term X11'*X21'*X33' only. On the other hand, the model chosen by the backward elimination method, shown in table(11), seems more complicated. The standard errors for all the coefficients in the two models are now much lower compared with those in table(9) and thus their corresponding t-values are higher.

[2]Although this F-value is inflated as we mentioned before but since the model has a lower standard error of estimates than the corrected model, there is no doubt about the statistical significance of this model.

Predictor'	Coef	Stdev	t-ratio	p
Constant'	3.6250	0.3220	11.26	0.000
X11'	-0.2917	0.4583	-0.64	0.525
X21'	0.3750	0.4774	0.79	0.433
X31'	-0.1250	0.5661	-0.22	0.826
X32'	-0.0250	0.4327	-0.06	0.954
X33'	0.3750	0.3762	1.00	0.320
X11'*X21'	-0.5083	0.7393	-0.69	0.493
X11'*X31'	0.0139	0.7367	0.02	0.985
X11'*X32'	1.3167	0.6887	1.91	0.058
X11'*X33'	0.1667	0.5930	0.28	0.779
X21'*X31'	0.1250	0.7042	0.18	0.859
X21'*X32'	0.2882	0.6049	0.48	0.634
X21'*X33'	-0.9135	0.5777	-1.58	0.116
X11'*X21'*X31'	0.5553	0.9966	0.56	0.578
X11'*X21'*X32'	-1.1608	0.9513	-1.22	0.224
X11'*X21'*X33'	-0.2032	0.9934	-0.20	0.838

s = 0.9032

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	16	2288.29	143.02	175.33	0.000
Error	166	135.41	0.82		
Total	182	2423.70			

Table(10): The Final Output Of The Stepwise And Forward Methods Of The Interaction Model(4) Of Y On X1, X2, And X3; SPSS Output.

Multiple R	.96816
R Square	.93733
Adjusted R Square	.93663
Standard Error	.91862

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	2	2271.80256	1135.90128
Residual	180	151.89634	.84387

F = 1346.06420 Signif F = .0000

---------------- Variables in the Equation ----------------

Variable	B	SE B	Beta	T	Sig T
Constant'	-1.158854	.493902	-.044272	-2.346	.0200
X11'*X21'*X33'	3.783858	.073321	.973739	51.607	.0000

According to the stepwise and forward model, the academic performance of the students is relatively less when, concurrently, their specialization is the social science and they live with their families as well as when the educational level of their guardian is intermediate. Otherwise, there is no evidence for the academic performance of the students to be affected by any change in all these conditions. For the other students, the academic performance is also unaffected by the variations due to their specialization or due to being living with their families.

Table(11): The Final Output Of The Backward Elimination For The Interaction Model (4) Of Y On X1, X2, And X3; SPSS Output.

```
Multiple R           .97105
R Square             .94293
Adjusted R Square    .94065
Standard Error       .88903

Analysis of Variance
                    DF      Sum of Squares      Mean Square
Regression           7          2285.38208        326.48315
Residual           175           138.31682           .79038

F =     413.07017       Signif F =  .0000

----------------- Variables in the Equation ---------------
Variable            B         SE B       Beta        T    Sig T
X11'*X21'*X32' -1.131636  .525544   -.104374   -2.153  .0327
X21'*X33'      -1.235699  .338455   -.098924   -3.651  .0003
Constant'       3.464694  .147948    .891606   23.418  .0000
X33'             .501943  .220780    .068367    2.273  .0242
X11'*X21'       -.536606  .234793   -.069761   -2.285  .0235
X21'             .662298  .204268    .127946    3.242  .0014
X11'*X32'       1.160300  .443874    .118150    2.614  .0097
```

For the backward model, however, the students whose the educational level of their guardian is intermediate performed academically better when they live with their families but they seem to perform academically less than those who do not live with their families. For those who live with their families, and regardless of their specialization, they seem to perform academically less than those who they don't. For those who live with their families with their specialization being natural science (i.e., not social science), they seem to perform academically more than those they do not live with their families. For the students whose the educational level of their guardian is secondary, and for those who live with their families, they tend to perform academically worse than those whose guardians' educational level is not secondary, but the performance they offer appears to be the same as for whose the educational level of their guardian is not secondary when their specialization is not social science and when not being living with their families. With just their specialization being social science, the student seem to perform academically worse than those specialization is not social science.

Regardless of the educational level of their guardians, according to the backward model, all the students appear to perform academically less when their specialization is social science and when they live with their families and/ or when just live with their families. When they do not live with their families, their academic performance, however, seems to be the same regardless of whether their specialization is social science or not.

It seems there is no point to apply the hierarchical principle (That is, if a higher order term exists all the lower order terms must exist as well) and include the lower-order terms of X11'*X21', X11'*X32', X21'*X32', X11', X21', and X32' to the stepwise and forward model and the terms of X21'*X32', X11', and X32' to the backward model. But as the results of table(11) shows, the lower-order terms are not necessarily significant when the higher-order term is significant. This is because the significance of a term in linear possibility models is based on the t-test which does not depend on the hierarchy.

To choose between the stepwise-forward model and the backward elimination model we use the Mallow's Ck. For the stepwise-forward model the Ck is 8.215 (with a difference of 6.215 between Ck and k) while the backward elimination model has a Ck of 1.563 (with Ck-k=-5.437). Since the backward model has a smaller absolute difference between ck and k than the stepwise-forward model, this is evidence, therefore, for choosing the backward elimination model than the stepwise-forward model, though the latter is more parsimonious.

Compared with the main factors model shown in table(6), the backward elimination model is obviously more significant and, above all, takes into account the interaction effects. However, we need to cross validate these two models to assess how well they predict in an independent sample(s) of data. In other words, to determine which of the two models have more generalizability. Table(12), along with tables (13) and (14), show the cross-validation rates and indices for these two models across the three samples: the data split sample, the leave one out sample, and the whole data set sample.

The cross-validation error rates in table(12) seem to be considerably higher than what we usually expect for all the models in the three samples. About two-thirds of the Y scores are incorrectly predicted with the main factors model and slightly lower than that with the backward elimination model. However, for a dependent variable like Y which has several outcomes (superb, good, fair, and weak), it is more likely that a predicted value will result in a mismatch than when this variable has only two outcomes (yes and no, say). This is probably because the chance for the outcome to be correctly forecasted will be smaller as the number of outcomes increase, that is, 4 out of 16 for our case compared with a 2 out of 4 chance if Y has a binary outcome.

Table(12): The Cross-validation Error Rates For The Main Factors Model Of Table(5) And The Backward Elimination Model Of Table(11).

Validation Sample	The Main Factors Model	The Backward Elimination Model
Data Split	68.1%	60.4%
Leave One Out	65.4%	64.4%
Whole Sample	65.4%	62.7%

Table(13): The Cross-validation Error Rates For The Main Factors Model And The Backward Elimination Model, Allowing For Minor Mismatch.

Validation Sample	The Main Factors Model	The Backward Elimination Model
Data Split	13.2%	16.5%
Leave One Out	11.5%	07.7%
Whole Sample	11.5%	07.1%

Table(14): The Cross-validation Indices For The Main Factors Model And The Backward Elimination Model.

Validation Sample	The Main Factors Model	The Backward Elimination Model
Data Split	1.018	0.953
Leave One Out	0.982	0.821
Whole Sample	1.124	0.760

Table(13) considers the error rates with a minor mismatch being allowed, for instance, the `superb' outcome to be predicted as `good' but not as `fair' or `weak', i.e., to be just one-level mismatch. For the two models, the error rates are now remarkably low compared with the previous ones in table(12), especially with the whole sample and the leave one out sample. It is also noted that the backward elimination model has now relatively lower error rates than its counterpart, the main effects model.

In table(12), we notice no major difference in the error rates between the split sample and whole data sample (in fact the split sample has smaller error rates than the whole sample for the backward elimination model). This is possibly because both of them are large. For the leave one out sample and the whole sample, the error rates in tables(12) and (13) are almost identical in the backward model and we would expect this, since the difference in the sample size is one.

Across the three samples, the cross-validation indices in table(14) are also lower for the backward elimination than for the main factors model. This is, therefore, further evidence to prefer the backward elimination model in table(11) than the main factors model in table(6). For this model, the whole data sample seems to be the favourite, since it has the lowest index among the three samples.

12- Summary

In this research we used the LPM techniques to analyze student academic performance data based on the cumulative rate out of 5 to be the dependent variable. The technique looks simple and can be applied by the familiar OLS method. By simple we mean the conditional mean value of the dependent variable is simply the conditional possibility of the event, given the values of the explanatory variables. However, although simple to apply as we said, this model has three main problems: nonnormality and heteroscedasticity of the error term as well as the possibility of fitted values lying outside the bounds of 2 to 5. The large sample size we have helped us not to bother with the first problem and we used the WLS method to overcome the second problem. We faced no difficulty with the third problem, as all the fitted values were within the limits of 2 to 5.

When we used the explanatory variables (X11, X21, X31, X32, and X33) without their interactions, X1 and X32 are the only variables that proved to be statistically significant. The students, therefore, appeared to perform academically better when their specialization is social sciences than otherwise. Those whose the educational level of their guardians is intermediate turned out to perform academically not quite well. Despite the importance of being living with their families, the selected model did not include this factor as being associated with the academic performance of the students.

When we used the explanatory variables with their interactions we obtained two different models using the variable selection procedures (stepwise, forward, and backward elimination). The first model, which was selected by the `forward' and `stepwise' procedures, consists of the interaction term of X11*X21*X32 only.

According to this model, the students tend to perform academically better when they, simultaneously, have a social science specialization and live with their families as well as their guardians having an intermediate level of education. The second model, which was obtained by the `backward' elimination procedure, has the interaction terms of X11*X21*X32, X21*X33, X11*X32, and X11*X21 along with the main factors of X21 and X33. When compared with all the previous models, this backward model appeared to be the most significant model and the one with the most reasonable Ck Mallow's, in addition, it was the one with the smallest cross-validation error rates and indices. We refined the model by omitting some outlier cases after which the model resulted in a lower standard error of estimate than before and, so, it became more significant. The interaction term of X11*X21 is dropped because of its high correlation with the other terms in the model. The term X21*X32 is also deleted because of its statistical nonsignificance. According to this model, the students whose their guardians have an intermediate level of education perform academically better when their specialization is social science and live with their families but they seem to perform academically less than other students when they don’t live with their families and their specialization is not social science. Without being living with their families, and regardless of their specialization, they also perform academically less than the other students. When they live with their families but their specialization is not social science, the students perform academically better than the other students. Regardless of the educational level of the guardians, according to the backward model, all students appear to perform academically less when their specialization is social science and live with their families or just being living with their families. When they do not live with their families, their academic performance, however, seems to be the same regardless of their specialization (social or natural sciences).

13. References:

1. Adam, Amin. I. (1993). Analyzing Categorical data: Various Solutions and Different Conclusions. Young Statistical Meeting, Liverpool.

2. Adam, Amin. I. (1996). Analysis of Categorical Data from a Case Study of Child Safety. Unpublished PhD thesis: University of Keele, Dept. of Mathematics, England.

3. Adam, Amin. I. (2010). Concepts on the Chi-square Test of Independence for Analyzing Categorical data. Journal of the Faulty of Economics & Political Science, Omdurman I. University, Sudan, Vol.4:131-143.

4. Adam, Amin. I. (2010). Local-Local, Local-Global, Global-Local and Global-Global Odds Ratios for Categorical data. J. of Economics and Political and Statistical Sciences, Omdurman I. University, Vol. 5:165-186.

5. Adam, Amin. I. (2010). Measures of Associations for Ordered Categorical Data: Different Measures but Similar Conclusions. J. of Economics and Political and Statistical Sciences, Omdurman I. University, Vol. 6:180-198.

6. Adam, Amin. I. (2011). Linear Possibility Model for Ordered Categorical Data: A Way of Analysis to Regression Analysis. J. of Economics and Political and Statistical Sciences, Omdurman I. University, Vol. 7:124-145.

7. Agresti, A. (2002). Categorical Data Analysis, 2nd ed. Wiley, New York.

8. Agresti, A. (2007). An Introduction to Categorical Data Analysis. Wiley, New York.

9. Agresti, A. (2010). Analysis of Ordinal Categorical Data. Wiley, New York.

10. Baglivo, J., Oliver, D. & Pagano, M. (1992). Methods for Exact Goodness-of-Fit Tests. Journal of the American Statistical Association 87:464-469.

11. Becker, M. P. & Clogg, C. C. (1989). Analysis of Sets of Two-Way Contingency Tables Using Association Models. Journal of the American Statistical Association 84:142-151.

12. Bilder, C. & Loughin, T. M. (2007). Modeling Association Between Two or More Categorical Variables that Allow for Multiple Categorical Choices. Communications in Statistics 36:433-451.

13. Draper, N. R. & Smith, H. (1998). Applied Regression Analysis, 3rd ed. Wiley, New York.

14. Everitt, B. S. (1977). The Analysis of Contingency Tables. Chapman & Hall, London.

15. Eye, A. V. & Bogat, G. A. (2009). Analysis of Intensive Categorical Longitudinal Data. Springer, New York.

16. Fan, Y. (2008). Strategic Groups and cluster Analysis. Henry Stewart, London.

17. Fienberg, S. E. (2007). The Analysis of Cross-classified Categorical Data. Springer, New York.

18. Freeman, D.H. (1987). Applied Categorical Data Analysis. Marcel Dekker, New York.

19. Greenland, S. (1991). On the Logical Justification of Conditional Tests for Two-by-Two Contingency Tables. American Statistician 45:248-251.

20. Gujarati, D. N. (2004). Basic Econometrics, 4th ed. McGraw-Hill, New York.

21. Hjorth, J. S. U. (1994). Computer Intensive Statistical Methods: Validation Model Selection and Bootstrap. Chapman, London.

22. Imrey, P. B. & Koch, G. G. (2005). Categorical Data Analysis. Wiley, New York.

23. Johnston, J. (1984). Econometric Methods, 3rd ed. McGraw-Hill, New York.

24. Johnston, j. & DiNardo, J. (2001). Econometric Methods, 4th ed. McGraw-Hill, New York.

25. Liu, I. & Agresti, A. (2005). The Analysis of Ordinal Categorical Data: An Overview and a Survey of Recent Development. Sociedad de Estadistica e Investigacion OperativaTe Vol.14 No. 1:1-73.

26. Maddala, G. S. & Lahiri, K.(2009). Introduction to Econometrics. Wiley, New York.

27. Mallows, C. L. (1973). Some Comments on Cp. Technometrics 15:661-675.

28. McCullagh, P. (1980). Regression Models for Ordinal Data. J. Roy. Statist. Soc. B 42:109-142.

29. Ott, R. L. & Longnecker M. (2008). An Introduction to Statistical Methods and Data Analysis, 6th ed. Brooks/Cole, Bolmont, U.S.A.

30. Powers, D. A. (2008). Statistical Methods for Categorical Data Analysis, 2nd ed. Emerald, Bingley, U.K.

31. Simono, J.(2003). Analyzing Categorical Data. Springer, New York.

32. Stevens, J. (1992). Applied Multivariate Statistics for the Social Sciences, 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum.

33. Weisberg, S. (1980). Applied Linear Regression. New York: Wiley.