May 7th, 2015

Undertaken in an attempt to model and understand some of the factors that explain the preferences of the student community with regard to ranking 3 year v/s 4 year UG programme, this project intends to demonstrate the use of Ordered Multiple Choice Model to predict categorical data, using the software, SAS.

# Multiple Choice Models

## MODELLING PREFERENCES OVER 3Yr v/s 4Yr UNDERGRADUATE PROGRAMME

Anuva S. Agarwal,
Nipun Thakurele,
Tanushree

M.Sc. Economics
Shiv Nadar University
Course In charge - Prof. Subrata Sarkar

# CONTENTS

SHIV NADAR UNIVERSITY

# ABSTRACT

The term paper is mainly intended to demonstrate the use of ordered Multiple Choice Model to predict categorical data. Learning important softwares like SAS is also an important objective of this paper

# MOTIVATION

The furore around the much hyped FYUP in Delhi University, in the Monsoon of 2013 and its subsequent withdrawal from the system within a year was a much debated topic. The same uncertainty surrounds some of the UG batches enrolled at SNU even now.

This project was undertaken in an attempt to purely understand the factors that explain the preferences of the student community with regard to the ranking 3 year v/s 4 year UG programme.

# RESEARCH QUESTION

**The aim of this project is to find the factors driving behind the choice of whether a SNU student prefers a 4 year degree or a 3 year degree.**

The question arises how the students are deciding among the choices given to them. Intuitively we could expect that choice will depend on the satisfaction (or utility) that a student gets from that choice. He or she will choose the alternative that gives him /her highest possible satisfaction. That choice will have the highest probability of being chosen. The dependent variable Y* measures a student's utility from a 3 year programme or a 4 year's programme. As Y* is unobservable so instead a proxy variable 'Opinion' has been used. Opinion reflects the student's response to the statement **"A 3 year Undergraduate programme is better than a 4 year undergraduate programme".**

# 1. DATA

- Data for the present study has been collected using an online survey where we asked undergraduate, postgraduate and PhD scholars of Schools of social sciences and Natural Sciences about their opinion on whether a 3 year undergraduate programme is better than a 4 year undergraduate programme. They were provided four choices: Strongly Agree, Agree, Disagree and Strongly Disagree.
- Information was also collected on a student's gender. It is believed that lesser number of females pursue higher studies as compared to their male counterparts. If that is the case we believe that they will prefer a 3 year degree. The regression aims to test that.
- Students have been diversified into two groups as per their schools i.e. School of Natural sciences and School of Humanities and Social sciences. We want to see whether or not it has any significant impact on a student's decision.
- Prior to regression , we believed that the variable measuring the highest level of education of the family member plays a role in the student's decision about his plan for further studies and career and thereby affects his decision regarding choice between 3rd and fourth year.
- The career preferences also affect the choice. We have grouped the population into whether they want to go into academics, industry or want to be self-employed. Prior to regression we believed that career preferences has some impact on a student's choice. For example a student who want to pursue an academic line, anyways will choose a PhD. So, he/she might prefer a 3 year degree because in any case later he has to do each thing in a rigorous way.
- Level of study enrolled in, i.e. an UG programme or an MS or PhD programme is expected to affect there is one set of individuals who are not completely aware of the benefits at stake while comparing a 3 yr programme with a 4 yr programme. On the other hand we have the MS or PhD students who have a sure shot experience of having been through an UG programme.
- Finally, we believe that financial condition plays a role in this choice. It should be positively related to the dependent variable. Prior to regression we believed that a person a student with unsatisfactory financial condition would prefer a 3 year degree to a four year degree.

# 2. MODEL SPECIFICATION

**Dependent variable: Opinion or Yi**

$Opinion_i$= 1 for strongly agree if $Y_i^* < \alpha_1$
= 2 for agree if $\alpha_1 < Y_i^* < \alpha_2$
= 3 for disagree if $\alpha_2 < Y_i^* < \alpha_3$
= 4 for strongly disagree if $Y_i^* > \alpha_3$

Where $Y_i^*$ is a latent variable measuring the student's utility from a particular programme and $\alpha_1, \alpha_2$ and $\alpha_3$ are respective thresholds.
$Opinion_i$ = 1 implies strongly agreeing to 3 year programme whereas
= 4 implies strongly disagreeing a 3 year programme over a 4 year Programme

Note that the numbers 1,2,3 and 4 are rankings, and the values chosen are not quantitative but merely an ordering. The difference between the outcomes represented by 1 and 2 is not necessarily the same as between 2 and 3.

**Explanatory variables**

1. **Gender** is a dummy variable which takes value 1 for male and 0 for female
2. **School** is a dummy variable which takes values 1 for School of Humanities and Social Sciences (SHSS) and 0 for School of Natural Sciences (SNS).
3. **Course** is a dummy variable which takes value 0 for students enrolled in an Undergraduate programme and 1 for those enrolled in a Masters or PhD programme
4. **Plans** is a dummy variable which takes value 0 if the plan is to pursue higher studies in the form of a Masters and 1 if the student hasn't decided or wants to pursue a PhD or M.Phil.
5. **Family Income (INR)** has three dummy variables
   Income Level 1
   Inc_less6 = 1 if Income is Below 6,00,000
   = 0 otherwise
   Income Level 2
   Inc_6to12 = 1 if Income is between 6,00,000-12,00,000
   = 0 otherwise
   Income Level 3
   Inc_more12 = 1 if Income is Above 12 Lacs
   = 0 otherwise
   For the purpose of estimation we have included the last two dummy into our model whereas the first dummy is the base category.

6. **Career Plans** has three dummy variables
   Career Plan 1 = 1 if plan is to work in Industry
   = 0 otherwise
   Career Plan 2 = 1 if plan is to work in Academia
   = 0 otherwise
   Career Plan 3 = 1 if plan is Self - Employment
   = 0 otherwise
   For the purpose of estimation we have included the 1st and 3rd dummy into our model whereas the dummy for Academia as the career plan is the base category.
7. **Fam_Qual** is a dummy variable which takes a value 0 if the highest level of education of family member is upto bachelor's and 1 if it's masters, Mphil or PhD.

# 3. METHODOLOGY

## 4.1 Introduction

There are many settings in which the economic outcome we seek to model is a discrete choice among a set of alternatives, rather than a continuous measure of some activity. Some examples are as follows:

1. We have data on smokers and we want to find the factors which influence this habit. So, the dependent variable here is that of smokers and takes values 0 and 1. We equate "no" with 0 and "yes" with 1.These decisions are qualitative choices. The 0/1 coding is for the sake of interpretational convenience.
2. Opinions of a certain type of legislation : Let 0 represent "strongly opposed" ,1 "opposed", 2 "neutral", 3 "support", and 4 "strongly support." These numbers are ranking.
3. The occupational field chosen by an individual : Let 0 be the clerk, 1 engineer, 2 lawyer, 3 politician, and so on. These datas are merely categories giving neither a ranking nor a count.

Regression model involving nominal scale variables are an example of a class of models known as qualitative response regression models. These are further divided into:
1. Binary Choice Model: Where the dependent variable takes 2 choices
2. Multiple Choice Models: Where the dependent variable takes many values

The multiple choice model models the relationship between a polytomous response variable and a set of regressor variables, so as to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable. These polytomous response models can be classified into two distinct types, depending on whether the response variable has an ordered or unordered structure.

The salient feature of unordered model is that we can characterize the alternative choices in terms of a single latent or choice variable. Each time this latent variable crosses a specified threshold a different choice is emerged.

For unordered models we cannot explain the m choices by using a single latent variable. For this m-1 latent variables are needed to characterize m choices. It is sufficient to explain the first (m-1) choices. The last choice is just the complement of the first (m-1) choices. A (m-1) variable distribution function is used to characterize (m) multiple choices.

There are basically 3 types of unordered models:
1. The generalized logit models
2. The conditional logit models, and

3. The Mixed logit models.

All the models have the following set of assumptions:
- Data are case specific.
- Independence among the choices of the dependent variable

It is important to distinguish between whether the model is ordered or not because if the actual model is not ordered but in empirical applications we use an ordered model then estimates are seriously biased and inconsistent. On the other hand if we use an unordered model when in actuality the model is ordered, then we only loose efficiency.

## 4.2 Ordered Model

Suppose we have m categories. Then we have m regions, and m-1 thresholds Let us denote these thresholds by $\alpha_1 > \alpha_2 > \alpha_3 \ldots\ldots > \alpha_{m-1}$ . Each time $y_i^*$ crosses a threshold, we have a different choice. However $y_i^*$ is not observable. Instead we observe an indicator $Y_i$ with the following rule:

$y_i = j$ if $\alpha_{j-1} < y_i^* < \alpha_j$ $j=1,2,3\ldots..m$
$=0$ otherwise

Given this characterization we have
$P(y_i=j/x_i) = P(\alpha_{j-1} < y_i^* < \alpha_j)$
$= P(\alpha_{j-1} < \beta'X_i + u_i < \alpha_j)$
$= P(\alpha_{j-1} - \beta'X_i < u_i < \alpha_j + \beta'X_i)$
$= F(\alpha_j + \beta'X_i) - F(\alpha_{j-1} - \beta'X_i)$
Where F is the distribution function of $u_i$
Now just as the simple binary model, different assumptions on the distribution function of $u_i$, gives us the different models.

1. If we assume $u_i \sim N(0,1)$ then it implies

$$P(y_i = j|x_i) = \Phi(\alpha_j - \beta' x_i) - \Phi(\alpha_{j-1} - \beta' x_i)$$

2. If we assume $u_i$ has a logistic distribution then, remembering the fact that logistic distribution
$F(U_i) = e^{U_i}/(1+e^{U_i})$ $f(U_i)=e^{U_i}/(1+e^{U_i})^2$

$P(y_i=j/x_i) = e^{(\alpha j-\beta'xi)}/1+ e^{(\alpha j-\beta'xi)} - e^{(\alpha j-1-\beta'xi)}/1+ e^{(\alpha j-1-\beta'xi)}$

Also note that there is need to normalize one threshold to zero. Thus if there are m choices then there would be m-1 thresholds and we can estimate m-2 thresholds uniquely or we can estimate m-1 net thresholds.

## 4.3 Estimation

One can in principle do MLE but it is difficult. Involves cross equations restrictions
We define dummy variables $Z_{ij}$ =1 if $Y_i$=j j=1,2……..m
=0 otherwise

Then assuming $U_i \sim N(0,1)$ i.e. probit

$$L_i \equiv \prod_{j=1}^{m} P_{ij} Z_{ij} = \prod_{j=1}^{m} [\Phi(\alpha_j - \beta' x_i) - \Phi(\alpha_{j-1} - \beta' x_i)]^{Z_{ij}} = \prod_{j=1}^{m} P_{ij} Z_{ij}$$

Assuming independent samples, we get the Likelihood function as

$$L_i = \prod_{i=1}^{n} \prod_{j=1}^{m} [\Phi(\alpha_j - \beta' x_i) - \Phi(\alpha_{j-1} - \beta' x_i)]^{Z_{ij}}$$

And the Log L function is given by

$$Log\ L = \sum_{i=1}^{n} \sum_{j=1}^{m} Z_{ij} \log[\Phi(\alpha_j - \beta' x_i) - \Phi(\alpha_{j-1} - \beta' x_i)]$$

1. Now we use Newton –Raphson iterative method to estimate the parameters

$$\hat{\beta}_j = \hat{\beta}_{j-y} - \left[ E\left( \frac{\partial^2 \log L}{\partial_\beta \partial_{\beta'}} \right) \right]^{-1}_{\hat{\beta}_{j-1}}$$

2. The likelihood function is globally concave and therefore guarantees that we reach the global maximum, irrespective of where we start. However, computationally, we will be faster depending on how close we start from the truth
3. The assumption Var.-Cov. matrix ix is given by:

$$\left[ E\left[ \frac{-\partial^2 \log L}{\partial_\beta \partial_{\beta'}} \right] \right]^{-1}$$

This is actually the inverse of information matrix.

Now that the β and variance is known it can be used for testing and inference.

## 4.4 Latent variable argument

Discrete dependent –variable models are often cast in the form of index function models. We view the outcome of a discrete choice as a reflection of an underlying regression. As an often cited example, consider the decision to make a large purchase.
The theory states that the consumer makes a marginal benefit –marginal cost calculation based on the utilities achieved by making the purchase and by not making the purchase and by using the money for something else. We model the difference between benefit and cost as an unobserved variable y* such that

$$\mathbf{y^* = x'\beta + \varepsilon}$$

We assume that ε has mean 0 and has either a standardized logistic with (known) variance π2/3 or a standard normal distribution with variance one. We do not observe the net benefit of the purchase, only whether it is made or not. Therefore, our observation is

$$Y=1 \text{ if } y^*>0$$
$$=0 \text{ if } y^*\leq 0$$

In this formulation, x'β is called the index function.

# 5. QUESTIONNAIRE

**Year of study - Academic year 2014-2015** *

- I yr
- II yr
- III yr
- IV yr

**Gender** *

- Male
- Female
- Other

**Opinion: A III yr UG programme is better than a IV yr UG programme** *

- Strongly Agree
- Agree
- Disagree
- Strongly Disagree

**School at the University** *

- School of Natural Sciences
- School of Humanities and Social Sciences

**Course enrolled in** *

---

**Course enrolled**

- BS Physics
- BS Mathematics
- BS Sociology
- BS Economics
- BS English
- BS Chemistry
- BS Biotechnology
- BS History
- Masters
- Doctoral - PhD

SHIV NADAR UNIVERSITY

**Plans in respect of further studies** *

(You're allowed to tick multiple options or write 'None' in the 'Other' option incase you've had your dose of books!)

- [ ] Masters
- [ ] M. Phil
- [ ] PhD
- [ ] Not Decided
- [ ] Other:

**Career Preferences** *

- ( ) Industry
- ( ) Academia
- ( ) Self - Employment

**Annual Family Income (INR)** *

- ( ) Upto 6,00,000
- ( ) 6,00,000 - 12,00,000
- ( ) Above 12,00,000

**Qualifications of family member with highest level of educational qualification** *

(The 'Others' option may be used to mention about a threshold, following the last item ticked, prior to the entry)

- [ ] Senior Secondary (X std.)
- [ ] Higher Secondary (XII std.)
- [ ] Bachelors
- [ ] Masters
- [ ] M.Phil.
- [ ] Ph.D.
- [ ] Other:

# 6. RESULTS

| Analysis of Maximum Likelihood Estimates | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
| Intercept | 4 | 1 | -2.4018 | 0.8939 | 7.2196 | 0.0072 |
| Intercept | 3 | 1 | 0.0521 | 0.8549 | 0.0037 | 0.9514 |
| Intercept | 2 | 1 | 1.4057 | 0.8852 | 2.5219 | 0.1123 |
| School | | 1 | -0.0689 | 0.4871 | 0.0200 | 0.8876 |
| Gender | | 1 | 0.5904 | 0.4408 | 1.7939 | 0.1804 |
| Course | | 1 | 2.2782 | 0.6050 | 14.1809 | 0.0002 |
| Plans | | 1 | 0.0971 | 0.4896 | 0.0393 | 0.8428 |
| Inc_6to12 | | 1 | -0.3869 | 0.4909 | 0.6213 | 0.4306 |
| Inc_more12 | | 1 | 2.3203 | 0.6104 | 14.4483 | 0.0001 |
| Fam_Qual | | 1 | -0.3000 | 0.5069 | 0.3503 | 0.5539 |
| Career_Ind | | 1 | -1.2354 | 0.4802 | 6.6189 | 0.0101 |
| Career_Self | | 1 | -0.6637 | 0.7186 | 0.8530 | 0.3557 |

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Point Estimate | 95% Wald Confidence Limits | |
| School | 0.933 | 0.359 | 2.425 |
| Gender | 1.805 | 0.761 | 4.282 |
| Course | 9.759 | 2.982 | 31.944 |
| Plans | 1.102 | 0.422 | 2.877 |
| Inc_6to12 | 0.679 | 0.259 | 1.777 |
| Inc_more12 | 10.179 | 3.077 | 33.676 |
| Fam_Qual | 0.741 | 0.274 | 2.001 |
| Career_Ind | 0.291 | 0.113 | 0.745 |
| Career_Self | 0.515 | 0.126 | 2.106 |

# 7. INFERENCES & CONCLUSION

**COEFFICIENTS - INTERPRETATION**

- The intercept 4 which shows strongly disagree is insignificant. The intercept 3 which shows disagree is highly significant. The intercept 2 which shows agree is highly significant. The p-value is almost 0.Thus out of the three thresholds only 2 are significant. This implies that the intercept4 can be merged with intercept 3.Also note that intercept 3 is highly significant.
- From the response profile we can see that the odds ratios for each of the explanatory variables depict the odds in favour of the opinion 'strong agreement' (taking the value 1) for the III yr programme.

  The reason for that is that as in SAS we had given the command proc logistic data = work.own descending . The code "descending" reverses sorting order of the response variable. Hence, now the opinion of strongly disagreeing to 3 year programme takes the position1 , the opinion disagreeing to 3 year programme takes position 2 and agreeing to 3 year programme takes position 3.So, in this manner SAS takes strongly agreeing to 3 year programme as the base category.So, all the odds ratios that are displayed in SAS are the odds in favour of strongly agreeing to 3 year programme.
  However , it should be noted if we have not given the command descending than SAS by default had retained the natural ordering and in that case strongly disagreeing to 3 year programme had been taken as the base.So, odds ratio would have been in favour of strongly disagreeing to 3 year programme.

- The variable **school** is a dummy which takes a value 0 for School of natural sciences and 1 for school of Humanities and Social Sciences .The coefficient -0.0689 thus reflects that for the School of Humanities and Social sciences the Odds in favour of strongly agreeing to a 3 year programme reduces. Note the coefficients here only determine the direction of change and not the magnitude of change. While comparing the p-value for the coefficients we see the variable school is insignificant. For the same variable while observing the odd ratio confidence interval we see that the coefficient is insignificant.

  This insignificance seems to point to the fact that being enrolled in either the SHSS or SNS doesn't make a difference when

  For the variable **gender** we have taken female as the reference category. The estimate 0.5904 than in females just conveys the information that odds in favour of strongly agreeing is greater in male. But from the p values we see that it is insignificant. While comparing the odds ratio confidence interval for gender we find that it is insignificant.

  One possible explanation for this insignificance is that parents of females who have conceded to send their daughters to a fully residential university, away from the city/

homes already belong to the class of individuals with a progressive mindset. It is unlikely that individuals belonging to this set will succumb to the norm we had hypothesised. Thus our initial hypothesis about societal norm isn't true in this case.

- For the variable course the reference category is undergraduate. The odds in favour of strongly agree is less for PhD and Masters Students when compared to base category i.e. the undergraduate students. The p-value shows that this difference is statistically significant. While comparing the odds ratio confidence interval we find it is highly significant. This clearly explains the pro-3yr opinion of the individuals who've gone through their UG program. This pro-3yr opinion could be explained by 2 possible beliefs – one, is that fact that those in a MS or Ph.D programme right now never had the choice between a 4 yr programme and hence they have come to believe that what they have done maybe the best. Alternately, they could be of the belief that the learning curve in an UG programme doesn't portray a significant increase in value-add in the 4$^{th}$ year. As a result, the jump to a MS programme and starting a new trajectory is the most productive bet that one can land into.
- The variable plan is a dummy variable which takes value 0 for 'Masters' and 1 for 'Not decided', 'PhD' or 'M.Phil'. The estimate 0.0971 shows that for plans other than pursuing a Masters, the odds in favour of the 3yr programme increase as compared to Masters. The p-value shows that it is insignificant. Thus the plans are not a significant factor in deciding the options.

  From the odds ratio confidence interval we observe that it is not significant.
- The variable **Fam_Qual** shows that the odds in favour of strong agreement for the 3 yr programme decreases if family members have an education level of masters or more. But p-values show it is insignificant. The confidence interval of odd ratio is insignificant.
- The variable **Inc_6to12** has a negative coefficient which means that the odds ratio is expected to decrease with an increase in the income slab from 'less than 6lakhs' to 'between 6lakhs and 12 lakhs'. This decrease, though mathematically existent is statistically insignificant.
- The variable **Inc_more12** has a positive coefficient which means that the odds ratio is expected to increase significantly (relative to the base category i.e. the income slab 'less than 6 lakhs') with an increase in the income slab from the base to 'more than 12 lakhs'.

  So, over all it may be concluded that when income is greater than INR12,00,000/- the the odds ratio in favour of the 3 yr programme is expected to move upward.
- The variable **Career_Ind** has a negative coefficient i.e. the odds in strong agreement for the 3 yr undergraduate programme are expected to decrease significantly for those individuals who plan on taking up industry as a career path.
- The variable **Career_Self** has a negative coefficient too, which is however insignificant. This means that the odds ratio in favour of 'strong agreement' is not

significantly different, relative to the odds ratio expected, if the career path involves an inclination towards academia.

- Also the global null hypothesis of all coefficients being equal to zero is rejected. So, the model is overall significant

| Testing Global Null Hypothesis: BETA=0 | | | |
|---|---|---|---|
| Test | Chi-Square | DF | Pr > ChiSq |
| Likelihood Ratio | 42.6788 | 9 | <.0001 |
| Score | 34.1398 | 9 | <.0001 |
| Wald | 33.3686 | 9 | 0.0001 |

So the **main findings of the project** are that out of the three thresholds, one is not statistically significant. The option of being strongly disagree is highly insignificant. Also some of the hypothesis prior to regression has been falsified. As an instance the variable gender came it to be statistically insignificant in a student's preference between a 3 year or a 4 year.
The variable course which depicted the preference by Master's and Phd on one hand and undergraduate on the other was statistically significant. Income as expected plays a detrimental role in this choice and so does careers.

# 8. REFERENCES


1. *ECON 522: Econometrics II. Lecture Notes by Prof. Subrata Sarkar*
2. *Damodar Gujarati, Econometrics by Example, Palgrave Macmillan*
3. *William H. Greene – Econometric Analysis, 5th Edition*
4. *http://support.sas.com/*

# 9. APPENDIX

## 9.1 SAS Output

The CONTENTS Procedure

| | | | |
|---|---|---|---|
| **Data Set Name** | WORK.OWN | **Observations** | 96 |
| **Member Type** | DATA | **Variables** | 10 |
| **Engine** | V9 | **Indexes** | 0 |
| **Created** | 05/10/2015 18:44:45 | **Observation Length** | 80 |
| **Last Modified** | 05/10/2015 18:44:45 | **Deleted Observations** | 0 |
| **Protection** | | **Compressed** | NO |
| **Data Set Type** | | **Sorted** | NO |
| **Label** | | | |
| **Data Representation** | WINDOWS_32 | | |
| **Encoding** | wlatin1 Western (Windows) | | |

**Engine/Host Dependent Information**

| | |
|---|---|
| **Data Set Page Size** | 65536 |
| **Number of Data Set Pages** | 1 |
| **First Data Page** | 1 |
| **Max Obs per Page** | 817 |
| **Obs in First Data Page** | 96 |
| **Number of Data Set Repairs** | 0 |
| **ExtendObsCounter** | YES |
| **Filename** | C:\Users\ag231\AppData\Local\Temp\SAS Temporary Files_TD5388_ECONLAB01-10_\own.sas7bdat |
| **Release Created** | 9.0401M1 |
| **Host Created** | W32_7PRO |

**Alphabetic List of Variables and Attributes**

| # | Variable | Type | Len | Label |
|---|---|---|---|---|
| **7** | Career_Ind | Num | 8 | Career_Ind |
| **8** | Career_Self | Num | 8 | Career_Self |
| **4** | Course | Num | 8 | Course |
| **9** | Fam_Qual | Num | 8 | Fam_Qual |
| **1** | Gender | Num | 8 | Gender |
| **5** | Inc_6to12 | Num | 8 | Inc_6to12 |
| **6** | Inc_more12 | Num | 8 | Inc_more12 |
| **2** | Opinion | Num | 8 | Opinion |
| **10** | Plans | Num | 8 | Plans |
| **3** | School | Num | 8 | School |



The LOGISTIC Procedure

**Model Information**

| | | |
|---|---|---|
| **Data Set** | WORK.OWN | |
| **Response Variable** | Opinion | Opinion |
| **Number of Response Levels** | 4 | |
| **Model** | cumulative logit | |
| **Optimization Technique** | Fisher's scoring | |

| | |
|---|---|
| **Number of Observations Read** | 96 |
| **Number of Observations Used** | 96 |

**Response Profile**

| Ordered Value | Opinion | Total Frequency |
|---|---|---|
| **1** | 4 | 35 |
| **2** | 3 | 38 |
| **3** | 2 | 13 |

SHIV NADAR UNIVERSITY

**Response Profile**

| Ordered Value | Opinion | Total Frequency |
|---|---|---|
| **4** | 1 | 10 |

Probabilities modeled are cumulated over the lower Ordered Values.

**Model Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Score Test for the Proportional Odds Assumption**

| Chi-Square | DF | Pr > ChiSq |
|---|---|---|
| 28.7724 | 18 | 0.0512 |

**Model Fit Statistics**

| Criterion | Intercept Only | Intercept and Covariates |
|---|---|---|
| **AIC** | 244.284 | 219.605 |
| **SC** | 251.977 | 250.377 |
| **-2 Log L** | 238.284 | 195.605 |

**Testing Global Null Hypothesis: BETA=0**

| Test | Chi-Square | DF | Pr > ChiSq |
|---|---|---|---|
| **Likelihood Ratio** | 42.6788 | 9 | <.0001 |
| **Score** | 34.1398 | 9 | <.0001 |
| **Wald** | 33.3686 | 9 | 0.0001 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|
| **Intercept** | **4** | 1 | -0.1236 | 0.6239 | 0.0392 | 0.8430 |
| **Intercept** | **3** | 1 | 2.3304 | 0.6783 | 11.8048 | 0.0006 |
| **Intercept** | **2** | 1 | 3.6840 | 0.7608 | 23.4476 | <.0001 |
| **School** | | 1 | -0.0689 | 0.4871 | 0.0200 | 0.8876 |
| **Gender** | | 1 | 0.5904 | 0.4408 | 1.7939 | 0.1804 |

**Analysis of Maximum Likelihood Estimates**

| Parameter | DF | Estimate | Standard Error | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|
| **Course** | 1 | -2.2782 | 0.6050 | 14.1809 | 0.0002 |
| **Plans** | 1 | 0.0971 | 0.4896 | 0.0393 | 0.8428 |
| **Inc_6to12** | 1 | -0.3869 | 0.4909 | 0.6213 | 0.4306 |
| **Inc_more12** | 1 | 2.3203 | 0.6104 | 14.4483 | 0.0001 |
| **Fam_Qual** | 1 | -0.3000 | 0.5069 | 0.3503 | 0.5539 |
| **Career_Ind** | 1 | -1.2354 | 0.4802 | 6.6189 | 0.0101 |
| **Career_Self** | 1 | -0.6637 | 0.7186 | 0.8530 | 0.3557 |

**Odds Ratio Estimates**

| Effect | Point Estimate | 95% Wald Confidence Limits | |
|---|---|---|---|
| **School** | 0.933 | 0.359 | 2.425 |
| **Gender** | 1.805 | 0.761 | 4.282 |
| **Course** | 0.102 | 0.031 | 0.335 |
| **Plans** | 1.102 | 0.422 | 2.877 |
| **Inc_6to12** | 0.679 | 0.259 | 1.777 |
| **Inc_more12** | 10.179 | 3.077 | 33.676 |
| **Fam_Qual** | 0.741 | 0.274 | 2.001 |
| **Career_Ind** | 0.291 | 0.113 | 0.745 |
| **Career_Self** | 0.515 | 0.126 | 2.106 |

**Association of Predicted Probabilities and Observed Responses**

| | | | |
|---|---|---|---|
| **Percent Concordant** | 78.2 | **Somers' D** | 0.572 |
| **Percent Discordant** | 21.0 | **Gamma** | 0.577 |
| **Percent Tied** | 0.8 | **Tau-a** | 0.394 |
| **Pairs** | 3139 | **c** | 0.786 |

## 9.2 SAS Codes

- **SAS Code with reverse sorting (results in this project base on this code)**

```
libname a 'D:\Econometrics_Project\SAS_Data';
proc import out= work.own
datafile='D:\Econometrics_Project\SAS_Data.xlsx'
dbms=excel replace;
sheet=sheet1;
getnames=yes;
run;

proc contents data=work.own descending;
run;
proc logistic data = work.own;
model Opinion=School Gender Course Plans Inc_6to12 Inc_more12
Fam_Qual Career_Ind Career_Self;
run;
```

- **SAS Code without reverse sorting i.e. natural ordering**

```
libname a 'D:\Econometrics_Project\SAS_Data';
proc import out= work.own
datafile='D:\Econometrics_Project\SAS_Data.xlsx'
dbms=excel replace;
sheet=sheet1;
getnames=yes;
run;

proc contents data=work.own;
run;
proc logistic data = work.own;
model Opinion=School Gender Course Plans Inc_6to12 Inc_more12
Fam_Qual Career_Ind Career_Self;
run;
```