

ECONOMETRICS II TERM PAPER

Multinomial Logit Models

Instructor : Dr. Subrata Sarkar

19.04.2013

Submitted by Group 7 members:

Akshita Jain

Ramyani Mukhopadhyay

Sridevi Tolety

Trishita Bhattacharjee

Acknowledgement:

We are highly indebted to Dr. Subrata Sarkar, our instructor for Econometrics-II, for his inputs, encouragement, support and advice on this paper. We would also like to thank Ishwarya Balasubramanian for her help and important suggestions. Finally we want to convey our gratitude to IGIDR authorities for the infrastructural support and all the staff associated with CC and library. Doing this term paper was a pleasant and learning experience for us.

Abstract

The term paper is mainly intended to demonstrate the use of ‘Multinomial Logit Models’ to predict categorical data. Three types of unordered choice models have been described namely, Generalised logit models, Conditional logit models and Mixed logit models. All the models are presented through proper examples to understand the models properly. Learning important softwares like SAS and LaTeX are also the objectives of this term paper.

1 Introduction

In several areas of study, we often come across dependent variables which are nominal and have more than two categories. For instance, we might be interested in knowing whether students have stayed in mainstream education, are in vocational training or have left education altogether. This would be a three category nominal variable. We therefore need methods that can model this type of dependent variable as well. Thus we can extend the logistic regression model to do this. We call this extended method multinomial logistic regression, and refer to logistic regression for dichotomous dependent variables as binary logistic regression. If the response variable is polytomous and all the potential predictors are discrete as well, we could describe the multiway contingency table by a log linear model. But fitting a log linear model has two disadvantages:

- It has a larger number of parameters, many of which are not of interest. The log linear model describes the joint distribution of all the variables, whereas the logistic model describes only the conditional distribution of the response given the predictors.
- The log linear model is more complicated to interpret. However, the effect of X on Y is a main effect. If we are analyzing a set of categorical variables and one of them is clearly a “response” while the others are predictors, it is better to use logistic rather than log linear models.

The following factors should be considered before using multinomial logistic regression:

- The sample size and any outlying cases.
- The dependent variable should be non-metric. For instance, we may use dichotomous, nominal, and ordinal variables.
- Independent variables should be metric or dichotomous.
- Initial data analysis should be thorough and include careful univariate, bivariate and multivariate assessment.

- Multicollinearity should be evaluated.

The rest of the paper is organised as follows: Section 2 describes the theoretical framework of Multinomial Logit Models. Section 3 demonstrates the models through appropriate examples and is followed by concluding remarks in section 4.

2 Theoretical Framework

A multinomial model is a model which generalizes logistic regression by allowing for more than two discrete outcomes. It 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.

In an unordered model the polytomous response variable does not have an ordered structure, i.e., the dependent variable is nominal as it falls into any one of a set of categories which cannot be ordered in any meaningful way and for which there are more than two categories. Multinomial logit is equivalent to a series of pairwise logit models. There are basically 3 types of unordered models:

1. The generalized logit models
2. The conditional logit models
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.
- Errors are independently and identically distributed.

Note that normality, linearity, or homoscedasticity are not assumed.

In studying consumer behaviour, an individual is presented with a set of alternatives and asked to choose the most preferred alternative. Both the generalized logit and conditional logit models are used in the analysis of discrete choice data. In a conditional logit model, a choice among alternatives is treated as a function of the characteristics of the alternatives, whereas in a generalized logit model, the choice is a function of the characteristics of the individual making the choice. In many situations, a mixed model that includes both the characteristics of the alternatives and the individual is needed for investigating consumer choice. We now discuss each of the models in detail. For that consider an individual choosing among m alternatives in a choice set. The regression equation would be $y_i^* = \beta'X_i + U_i$. However y_i^* is not observable. Instead we observe an indicator Y_i with the following rule:

$$\begin{aligned} Y_i &= j && \text{if } \alpha_{j-1} < y_i^* < \alpha_j; \quad j = 1, 2, \dots, m \\ &= 0 && \text{Otherwise} \end{aligned}$$

We define m dummy variables Z_{ij} for individual i with the following rule

$$\begin{aligned} Z_{ij} &= 1 && \text{if } Y_i = j; \quad j = 1, 2, \dots, m \\ &= 0 && \text{Otherwise} \end{aligned}$$

Assuming $U_i \sim N(0, 1)$, i.e., probit, let Π_{jk} denote the probability that individual j chooses alternative k , let X_j represent the characteristics of individual j , and let Z_{jk} be the characteristics of the k th alternative for individual j .

2.1 The Generalized Logit Models

The generalized logit model consists of a combination of several binary logits estimated simultaneously. Here the choice is a function of the characteristics of the individual making the choice. The generalized logit model focuses on the individual as the unit of analysis and uses individual characteristics as explanatory variables. The explanatory variables which

are the characteristics of an individual, are constant over the alternatives. The probability that individual j chooses alternative k is,

$$\Pi_{jk} = \frac{\exp(\beta'_k X_j)}{\sum_{l=1}^m \exp(\beta'_l X_j)} = \frac{1}{\sum_{l=1}^m \exp[(\beta_l - \beta_k)' X_j]}$$

Where, $\beta_1, \beta_2, \dots, \beta_m$ are m vectors of unknown regression parameters (each of which is different, even though X_j is constant across alternatives). Since $\sum_{l=1}^m \Pi_{jk} = 1$, the m sets of parameters are not unique. By setting the last set of coefficients to null (that is, $\beta_m = 0$) the coefficients β_k represent the effects of the X variables on the probability of choosing the k th alternative over the last alternative. In fitting such a model, one has to estimate $m - 1$ sets of regression coefficients.

2.2 The Conditional Logit Models

In the conditional logit model, the explanatory variables Z assume different values for each alternative and the impact of a unit of Z is assumed to be constant across alternatives. The probability that the individual j chooses alternative k is

$$\Pi_{jk} = \frac{\exp(\theta' Z_{jk})}{\sum_{l=1}^m \exp(\theta' Z_{jl})} = \frac{1}{\sum_{l=1}^m \exp[\theta'(Z_{jl} - Z_{jk})]}$$

θ is a single vector of regression coefficients. The impact of a variable on the choice probabilities derives from the difference of its values across the alternatives.

2.3 The Mixed logit models

For the mixed logit model that includes both characteristics of the individual and the alternatives, the choice probabilities are

$$\Pi_{jk} = \frac{\exp(\beta'_k X_j + \theta' Z_{jk})}{\sum_{l=1}^m \exp(\beta'_l X_j + \theta' Z_{jl})}$$

$\beta_1, \dots, \beta_{m-1}$ and $\beta_m \equiv 0$ are the alternative-specific coefficients, and θ is the set of global coefficients.

2.4 Writing The Mixed Logit Model as a Generalized Logit Model

Now as assumed individuals have m choices. In this model the probability of the j th choice is specified in the following way:

$$P(Y_i = j|X_i) = \frac{e^{\beta'_j X_i}}{\sum_{j=1}^m e^{\beta'_j X_i}}$$

Here X_i includes two types of information:

1. The individual socio- economic characteristics, eg. age, income, sex etc.
2. The choice characteristics. Suppose the m choices retain no different occupations. Then X_i includes the characteristics of all the m occupations. If for the j th choice some of the occupation characteristics are irrelevant then, we simply set the corresponding co-efficient of j to zero.

Occupation choice for all m occupations Socio economic change

$$X_i = \begin{bmatrix} X_{01i} & X_{02i} & X_{03i} & X_{04i} & X_{05i} & \vdots & X_{s1i} & X_{s2i} & X_{s3i} \end{bmatrix}$$

Suppose for an occupation 1 only characteristics 1,2,3 are relevant, for occupation 2, only 2,3,5 are relevant. Then

$$\beta_1 = \begin{bmatrix} \beta_{01} & \beta_{02} & \beta_{03} & 0 & 0 & \vdots & \beta_{16} & \beta_{17} & \beta_{18} \end{bmatrix}$$

$$\beta_2 = \begin{bmatrix} 0 & \beta_{02} & \beta_{03} & 0 & \beta_{05} & \vdots & \beta_{26} & \beta_{27} & \beta_{28} \end{bmatrix}$$

Given this specification:

$$\frac{P(Y_i = j|X_i)}{P(Y_i = i|X_i)} = \frac{e^{\beta'_j X_i}}{e^{\beta'_i X_i}} = e^{(\beta'_j - \beta'_i) X_i}$$

Therefore the relative probability between j and i depends:

1. Only on the difference of β_j and β_i , hence we need some normalization. Take

$$\beta_1 = 0 \Rightarrow e^{\beta'_i X_i} = e^0 = 1 \Rightarrow P(Y_i = j | X_i) = \frac{e^{\beta'_j X_i}}{1 + \sum_{j=2}^m e^{\beta'_j X_i}}.$$

2. Independence of irrelevant alternatives. Suppose we have two basic choices, m_1 and m_2 . The individual is indifferent between the choices, then relative probability=1.

$$\frac{P(1)}{P(2)} = 1 \Rightarrow P(1) = P(2) = 0.5.$$

But say, the individual faces a choice in m_1 , i.e. m_{11} and m_{12} . The multinomial logit model will read these as three independent choices, m_{11} , m_{12} and m_2 .

$$\Rightarrow \frac{P(1)}{P(2)} = 1, \frac{P(1)}{P(3)} = 1 \text{ and } \frac{P(2)}{P(3)} = 1 \Rightarrow P(1) = P(2) = P(3) = \frac{1}{3}.$$

Since while analyzing the choice between any two choices, it ignores the characteristics of the third choice. In reality $P(1) = 0.25$, $P(2) = .25$, $P(3) = 0.5$. This means that the odds ratio between alternative 1 and 3 is 1:1 in multinomial logit structure, but it is actually 1:2. This is an obvious inconsistency and depends on the fact that 1 and 2 are correlated choices. We need to use Nested Logit Models to solve this problem.

2.5 Estimation

Maximum Likelihood estimation is used for Multinomial logit models, where

$$L_i = \prod_{j=1}^m P_{ij}^{Z_{ij}}$$

$$L = \prod_{i=1}^n \prod_{j=1}^m P_{ij}^{Z_{ij}} \rightarrow \log L = \sum_{i=1}^n \sum_{j=1}^m Z_{ij} \log P_{ij}$$

Now we use Newton-Raphson iterative method to estimate the parameters,

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

The likelihood function is globally concave and therefore guarantees the global maximum.

The variance-covariance matrix is given by $[E[\frac{-\partial^2 \log L}{\partial \beta \partial \beta'}]]^{-1}$, which can be used to do testing and inference.

3 Modelling with Examples

We model discrete choice data with different examples using the SAS software. In all the examples, subjects (individuals) are presented with choices among alternatives, and their preferences are studied in order to investigate consumer choice.

3.1 GL Models

- EXAMPLE 1: CANDY CHOICE

This example is from *SAS/STAT Software: Changes and Enhancements, Release 8.2*. There is a choice of candy in three bowls: small chocolate candy bars, lollipops, and sugar candies. The subjects are classified by gender and age (as either “child” or “teenager”)¹. In this example Chocolate, Lollipop and Sugar are the alternative-specific variables and Age and Gender are the characteristics of the individual. We investigate whether age or gender affects the choice of type of candy. A generalized logit model can be fit to relate these three factors since the choice is made on the basis of children’s age and gender. So, $k=3$ in this model which can be chocolate, lollipop and sugar. Let X_1 represent the gender of individual j , X_2 represent the age of the individual j and Z_{jk} represent the counts of the chosen candy for individual j . Chocolate’ is the reference category for the Candy variable, so the logit being modeled is $\log([\frac{Pr(Candy=lollipop)}{Pr(Candy=chocolate)}])$ and $\log([\frac{Pr(Candy=sugar)}{Pr(Candy=chocolate)}])$, one comparing choice of lollipop versus chocolate and one comparing choice of sugar versus chocolate. In fitting such a model, we estimate 2 sets of regression coefficients.

The parameter estimates² and odds ratios³ for the asymptotic analysis show that the odds of choosing a lollipop over a chocolate bar are $5(4.835 \approx 5)$ times higher for boys versus girls, and a child is $14(1/0.071 \approx 14)$ times more likely than a teenager to choose a lollipop over a chocolate bar.

¹the data sets, SAS commands and output tables are presented in the Appendix

²see Table 1.4 of Appendix

³see Table 1.5 of the Appendix.

Note in the “Analysis of Maximum Likelihood Estimates” table⁴, the dummy parameters for the class variables are labeled by their non-reference level, and that the “Candy” column indicates the non-reference response category for the logit.

In the output table, the likelihood ratio chi-square of 18.9061 with a p-value < 0.0008 tells us that our model as a whole fits significantly better than an empty model (i.e., a model with no predictors). Several model fit measures such as the AIC are listed below:

Model Fit Statistics

Two models are tested in this multinomial regression, they correspond to the two equations below:

$$\log\left(\frac{Pr(Candy = lollipop)}{Pr(Candy = chocolate)}\right) = b_{10} + b_{11}(gender = boy) + b_{12}(age = teenager)$$

$$\log\left(\frac{Pr(Candy = sugar)}{Pr(Candy = chocolate)}\right) = b_{20} + b_{21}(gender = boy) + b_{22}(age = teenager)$$

Where, b’s are the regression coefficients.

1. A one-unit increase in the variable gender is associated with a 1.5758 increase in the relative log odds of choosing lollipop vs. chocolate.
2. A one-unit increase in the variable gender is associated with a 1.5261 increase in the relative log odds of choosing sugar vs. chocolate.
3. A one-unit increase in the variable age is associated with a 2.6472 decrease in the relative log odds of choosing lollipop vs. chocolate.
4. A one-unit increase in the variable age is associated with a decrease in the relative log odds of choosing sugar vs. chocolate.
5. The overall effects of gender and age are listed under “Type 3 Analysis of Effects” and both are significant. Gender has marginal influence.

⁴see Table 1.4 of Appendix

We are interested in testing whether Genderboy_lollipop is equal to Genderboy_sugar and Ageteenager_lollipop is equal to Ageteenager_sugar, which we can now do with the test statement. From the table “Linear hypothesis testing results”⁵, we observe the effect of gender=boy for predicting lollipop vs chocolate is not different from the effect of gender=boy for predicting sugar vs. chocolate. Similarly, the effect of age=teenager for predicting lollipop vs. chocolate is not different from the effect of age=teenager for predicting sugar vs chocolate. In table “Gender Least Squares Means”⁶, the predicted probabilities are in the ‘Mean’ column. Thus, for gender=boy and age=0.5, we see that the probability of choosing the chocolate is 0.2193 and for the lollipop 0.0880. In table “Age Least Squares Means”⁷, for gender=0.5 and age=child, we see that the probability of choosing chocolate is 0.1511 and for lollipop 0.7175.

- **EXAMPLE 2: TRAVEL CHOICE**

People are asked to choose between travel by auto, plane or public transit. The variables AUTOTIME, PLANTIME, and TRANTIME represent the total travel time required to get to a destination by using auto, plane, or transit, respectively. They are alternative-specific variables. The variable AGE represents the age of the individual being surveyed, and is a characteristic of the individual. The variable CHOSEN contains the individual’s choice of travel mode. The GL model here studies the relationship between the choice of transportation and age. Using the original dataset “Travel” we run the GL model⁸.

Two models are estimated in this multinomial regression. One compares choice of auto over transit and the other compares choice of plane over transit. We observe that⁹ :

⁵see Table 1.6 of Appendix

⁶see Table 1.7 of Appendix

⁷see Table 1.8 of Appendix

⁸see Appendix for SAS commands

⁹see Table 2.2 of Appendix

1. A one-unit increase in the age of the subject is associated with a 0.0710 decrease in the relative log odds of choosing auto over transit
2. A one-unit increase in the age of the subject is associated with a 0.05 decrease in the relative log odds of choosing plane over transit.

We might interpret this as a decrease in the preference for both auto and plane over public transit as the individuals age increases. This could possibly be due to the inconvenience of driving and the high cost of air travel. There are two intercept coefficients and two slope coefficients for Age. The first Intercept and the first Age coefficients correspond to the effect on the probability of choosing auto over transit, and the second intercept and second age coefficients correspond to the effect of choosing plane over transit. Using our estimates, we can predict the mode of travel choice for a hypothetical set of individuals over a range of ages from 20 to 70.

Table 1: Travel Choice- GL predictions

<i>Age</i>	$exp(\beta' X_j)$	$Prob(Auto)$	$exp(\beta' X_j)$	$Prob(Plane)$
20	5.0779	0.5156	5.5912	0.41408544
30	2.496	0.253	3.3912	0.251155515
40	1.2274	0.1246	2.0569	0.15233352
50	0.6034	0.0612	1.2475	0.0923
60	0.2966	0.0301	0.7566	0.056
70	0.1458	0.0148	0.4589	0.0339
Total=	9.847	1	13.5026	1

As we estimated, the predictions show that the probability of choosing auto over public transit decreases with age, and so does that of choosing plane over transit.

3.2 CL Models

- EXAMPLE 1: CANDY CHOICE

In this model, individuals choose their preferred alternative from a set of alternatives. In SAS the regression can be performed using the PHREG command. This command performs regression analysis of survival data, wherein subjects of the study are studied till they survive; following which no information is available on their status. We consider an example from *SAS, 1995, Logistic Regression Examples Using the SAS System, pp. (2-3)*. Chocolate candy data are available in which 10 subjects are presented with eight different chocolate candies. The subjects choose one preferred candy from among the eight types. The eight candies consist of eight combinations of:

1. dark(1) or milk(0) chocolate;
2. soft(1) or hard(0) center; and
3. nuts(1) or no nuts(0).

The dataset is first rearranged as follows:

1. The most preferred choice is said to occur at time 1 and all other choices are said to occur at later times or to be censored. In case it is not censored, i.e. the observation was chosen, the survival time is also called event time. The name of the failure time variable is t.
2. A status variable is created to denote whether the observation was censored (not chosen) or not censored (chosen). The censoring indicator variable has the value of 0 if the alternative was censored (not chosen) and 1 if not censored (chosen). Here the status variable is choose
3. A variable is created to specify the basis of censoring. The name of the variable here is subject and it refers to the individuals numbering 1,2,...,10.
4. X_1 and X_2 are explanatory variables.

The estimation is done via the PROC PHREG command. This command performs regression analysis of survival data, wherein subjects of the study are studied till they survive; following which no information is available on their status.

The estimation result for probabilities is:

$$p_j = \frac{\exp(1.386294DARK_j - 2.197225SOFT_j + 0.8472298NUTS_j)}{\sum_{j=1}^8 \exp(1.386294DARK_j - 2.197225SOFT_j + 0.8472298NUTS_j)}$$

The positive parameter estimates of DARK and NUTS mean that dark and nuts each increases the preference. The negative parameter estimate of SOFT denotes soft center.

On the basis of the estimated coefficients, we can predict the probability of a particular candy being selected on the basis of the characteristics of the candy.

Table 2: Candy Choice- CL estimates

j	Dark	Soft	Nuts	$\exp(\beta' X_j)$	p_j
1	0	0	0	1	0.054003
2	0	0	1	2.333334	0.126006
3	0	1	0	0.111056	0.005997
4	0	1	1	0.25913	0.013994
5	1	0	0	3.999999	0.216011
6	1	0	1	9.3333310	0.504025
7	1	1	0	0.444223	0.023989
8	1	1	1	1.036521	0.055975
Total				=18.51759	1

Most preferred type of candy is dark chocolate with a hard center and nuts, with probability 0.5.

To test the hypothesis that the effects of dark and soft (and, respectively, the effect of soft and nuts) are not different from zero, we run the “test” command in SAS

and observe that¹⁰ the first hypothesis can be rejected on the basis of the critical value of the chi-square, but there is no evidence to reject the second. In face of such results, care must be taken in prediction, as we could very well end up with inaccurate predictions.

- **EXAMPLE 2: Travel Choice**

We consider another example of conditional regression model, where we study how travel time affects the individual's choice of transportation. AutoTime, PlanTime, and TranTime are alternative-specific variables, whereas Age is a characteristic of the individual. Using the same dataset as we did in Example 2 of the GL model, we need to rearrange the dataset¹¹ in order to assign a survival/failure time for each alternative.

1. The failure time is represented by choice and takes the value of 1 if it is most preferred and 2 otherwise. Choice therefore acts as time variable as well as the status variable.
2. Subject refers to the individuals numbering 1,2,...,9.
3. TravTime represents the alternative specific explanatory variables Z_1 and Z_2 .

We rearrange the dataset TRAVEL into a new data set CHOICE, where we assign a survival time for each alternative. Time=1 is assigned if that alternative was chosen and 2 if it was not chosen. We run the model using the PHREG procedure again. The estimation result for probabilities is¹²:

$$p_j = \frac{\exp(-0.26549Z_j)}{\sum_j \exp(-0.26549Z_j)}$$

The negative coefficient for TravTime indicates that higher travel time would reduce a subject's preference for that alternative.

¹⁰see Table 3.3 of Appendix

¹¹See Table 2 of Appendix

¹²See Table 3.3 of Appendix

On the basis of the estimated coefficients, we can predict the probability of a particular mode of travel being selected, given their travel times. Here, the travel times are not constant across the choices faced by individuals. If a choice of (Travel times= 4.5, 10.5 and 3.2) is given to all the subjects, our predicted probabilities would be:

Table 3: Travel Choice- CL predictions

j		$exp(\beta' X_j)$	p_j
Auto	4.5	0.302793	0.382335
Plane	10.5	0.061566	0.077739
Transit	3.2	0.4276	0.539926
	Total	=0.791959	1

Hence public transit has the highest probability of being chosen.

3.3 ML Models

To study how the choice depends on both the travel time and age of the individual, we use a mixed model that incorporates both types of variables.

We again need to reshape the dataset TRAVEL into a new dataset CHOICE¹³, creating new variables AgeAuto and AgePlane, each of which represent the products of the individual's age and his failure time for each choice. We then use the PHREG procedure for this modified dataset with the same specifications as before.

The negative coefficient for plane with transit¹⁴ as the reference category might point towards the inconvenience of travelling by plane, perhaps due to the high cost. Travel by automobile would be preferred to public transit. The negative coefficient for AgeAuto implies that, despite the general preference exhibited for auto, it becomes an undesirable option as the subject grows older. Whereas it appears that plane travel would be preferred

¹³see Table 2 of Appendix

¹⁴see Table 5.3 of Appendix

by older people. As in the conditional logit model, travel time has a negative effect on preference.

4 Conclusion

In this paper we have attempted to give an overview of the theoretical and practical aspects of the Multinomial Logit Model. We demonstrate the applications of the three types of such models through examples and showed how the estimates obtained from the regressions can be used to predict the choices or responses of individuals. This model has widespread applications in the study of consumer preferences, levels of academic achievements, gender based differences in outcomes, medical research and various areas of behavioural economics.

The assumptions underlying the Multinomial Logit Model often do not hold in practice. Independence of irrelevant alternatives in particular may be violated as we have discussed earlier. Nested logits are better fitted to describe such choices. Furthermore, the subjects of the study might be influenced by their past choices or may have some specific opinions about certain attributes or levels. This can be resolved by using Multinomial probit Models, which does not have such restrictive assumptions.

References

- SAS/STAT Software: Changes and Enhancements, Release 8.2.
- SAS, 1995, Logistic Regression Examples Using the SAS System, pp. (2-3).
- Lecture notes of Dr. Subrata Sarkar
- So Y. and Kuhfeld W. F., “Multinomial Logit Models”, 2010, presented at SUGI 20.
- Starkweather, J. and Moske, A., “Multinomial Logitistic Regression”, 2011.

Websites

- www.ats.ucla.edu
- www.wikipedia.com

APPENDIX

A. SAS Commands

1. Generalised Logit Model

1.1 Candy Choice

```
data GLEExample1;
format Candy $9.;
input Gender $ Age $ Candy $ count @@;
    datalines;
        boy child chocolate 2 boy teenager chocolate 1
        boy child lollipop 13 boy teenager lollipop 9
        boy child sugar 3 boy teenager sugar 3
        girl child chocolate 3 girl teenager chocolate 8
        girl child lollipop 9 girl teenager lollipop 0
        girl child sugar 1 girl teenager sugar 1
;

proc logistic data=GLEExample1 outest = mlogit_param;
freq count;
class Gender(ref='girl') Age(ref='child') / param=ref;
model Candy(ref='chocolate') = Gender Age / link=glogit;
run;

proc logistic data=GLEExample1 outest = mlogit_param;
freq count;
class Gender(ref='girl') Age(ref='child') / param=ref;
model Candy(ref='chocolate') = Gender Age / link=glogit;
    Test_1: test Genderboy_lollipop = Genderboy_sugar;
    Test_2: test Ageteenager_lollipop = Ageteenager_sugar;
run;

proc logistic data = GLEExample1 outest = mlogit_param;
freq count;
class Candy Gender Age / param = glm;
model Candy = Gender Age / link = glogit;
lsmeans Gender / e ilink cl;
run;

proc logistic data = GLEExample1 outest = mlogit_param;
```

```

freq count;
class Candy Gender Age / param = glm;
model Candy = Gender Age / link = glogit;
lsmeans Age / e ilink cl;
run;

```

1.2 Travel Choice

```

data travel;
input AutoTime PlanTime TranTime Age Chosen $;
datalines;
  10.0 4.5 10.5 32 Plane
  5.5 4.0 7.5 13 Auto
  4.5 6.0 5.5 41 Transit
  3.5 2.0 5.0 41 Transit
  1.5 4.5 4.0 47 Auto
  10.5 3.0 10.5 24 Plane
  7.0 3.0 9.0 27 Auto
  9.0 3.5 9.0 21 Plane
  4.0 5.0 5.5 23 Auto
  22.0 4.5 22.5 30 Plane
  7.5 5.5 10.0 58 Plane
  11.5 3.5 11.5 36 Transit
  3.5 4.5 4.5 43 Auto
  12.0 3.0 11.0 33 Plane
  18.0 5.5 20.0 30 Plane
  23.0 5.5 21.5 28 Plane
  4.0 3.0 4.5 44 Plane
  5.0 2.5 7.0 37 Transit
  3.5 2.0 7.0 45 Auto
  12.5 3.5 15.5 35 Plane
  1.5 4.0 2.0 22 Auto
;

```

```

proc catmod data=travel;
  direct age;
  model chosen=age;
  title 'Multinomial Logit Model Using Catmod';
run;

```

2. Conditional Logit Model

2.1 Candy Choice

```

data CLExample1;

```

```

input subject choose dark soft nuts @@;
t=2-choose;
cards;
  1 0 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 0 1 1
  1 1 1 0 0 1 0 1 0 1 1 0 1 1 0 1 0 1 1 1
  2 0 0 0 0 2 0 0 0 1 2 0 0 1 0 2 0 0 1 1
  2 0 1 0 0 2 1 1 0 1 2 0 1 1 0 2 0 1 1 1
  3 0 0 0 0 3 0 0 0 1 3 0 0 1 0 3 0 0 1 1
  3 0 1 0 0 3 0 1 0 1 3 1 1 1 0 3 0 1 1 1
  4 0 0 0 0 4 0 0 0 1 4 0 0 1 0 4 0 0 1 1
  4 1 1 0 0 4 0 1 0 1 4 0 1 1 0 4 0 1 1 1
  5 0 0 0 0 5 1 0 0 1 5 0 0 1 0 5 0 0 1 1
  5 0 1 0 0 5 0 1 0 1 5 0 1 1 0 5 0 1 1 1
  6 0 0 0 0 6 0 0 0 1 6 0 0 1 0 6 0 0 1 1
  6 0 1 0 0 6 1 1 0 1 6 0 1 1 0 6 0 1 1 1
  7 0 0 0 0 7 1 0 0 1 7 0 0 1 0 7 0 0 1 1
  7 0 1 0 0 7 0 1 0 1 7 0 1 1 0 7 0 1 1 1
  8 0 0 0 0 8 0 0 0 1 8 0 0 1 0 8 0 0 1 1
  8 0 1 0 0 8 1 1 0 1 8 0 1 1 0 8 0 1 1 1
  9 0 0 0 0 9 0 0 0 1 9 0 0 1 0 9 0 0 1 1
  9 0 1 0 0 9 1 1 0 1 9 0 1 1 0 9 0 1 1 1
 10 0 0 0 0 10 0 0 0 1 10 0 0 1 0 10 0 0 1 1
 10 0 1 0 0 10 1 1 0 1 10 0 1 1 0 10 0 1 1 1
;

```

```

proc phreg data=CLEExample1;
  strata subject;
  model t*choose(0)=dark soft nuts;
run;

```

```

proc phreg data=CLEExample1;
  strata subject;
  model t*choose(0)=dark soft nuts;
  test1: test dark=soft=0;
  test2: test soft=nuts=0;
run;

```

2.2 Travel Choice

```

proc phreg data=choice;
  model choice*choice(2) = travtime / ties=breslow;
  strata subject;
run;

```

3. Mixed Logit Model

```

proc phreg data=choice2;
  model choice*choice(2) = auto plane ageauto ageplane travtime
  /ties=breslow;
  strata subject;
run;

```

B. Data Tables

Table 1. GL Model: Subject Preferences

Gender	Age	Candy		
		Chocolate	Lollipop	Sugar
Boy	child	2	13	3
Boy	teenager	10	9	3
Girl	child	3	9	1
Girl	teenager	8	0	1

Table 2. CL & ML: Dataset "Choice"

Subject	Mode	TravTime	Choice	Auto	Plane	AgeAuto	AgePlane
1	Auto	10	2	1	0	32	0
1	Plane	4.5	1	0	1	0	32
1	Transit	10.5	2	0	0	0	0
2	Auto	5.5	1	1	0	13	0
2	Plane	4	2	0	1	0	13
2	Transit	7.5	2	0	0	0	0
3	Auto	4.5	2	1	0	41	0
3	Plane	6	2	0	1	0	41
3	Transit	5.5	1	0	0	0	0
4	Auto	3.5	2	1	0	41	0
4	Plane	2	2	0	1	0	41
4	Transit	5	1	0	0	0	0
5	Auto	1.5	1	1	0	47	0
5	Plane	4.5	2	0	1	0	47
5	Transit	4	2	0	0	0	0
6	Auto	10.5	2	1	0	24	0
6	Plane	3	1	0	1	0	24
6	Transit	10.5	2	0	0	0	0
7	Auto	7	1	1	0	27	0
7	Plane	3	2	0	1	0	27
7	Transit	9	2	0	0	0	0

8	Auto	9	2	1	0	21	0
8	Plane	3.5	1	0	1	0	21
8	Transit	9	2	0	0	0	0
9	Auto	4	1	1	0	23	0
9	Plane	5	2	0	1	0	23
9	Transit	5.5	2	0	0	0	0
10	Auto	22	2	1	0	30	0
10	Plane	4.5	1	0	1	0	30
10	Transit	22.5	2	0	0	0	0
11	Auto	7.5	2	1	0	58	0
11	Plane	5.5	1	0	1	0	58
11	Transit	10	2	0	0	0	0
12	Auto	11.5	2	1	0	36	0
12	Plane	3.5	2	0	1	0	36
12	Transit	11.5	1	0	0	0	0
13	Auto	3.5	1	1	0	43	0
13	Plane	4.5	2	0	1	0	43
13	Transit	4.5	2	0	0	0	0
14	Auto	12	2	1	0	33	0
14	Plane	3	1	0	1	0	33
14	Transit	11	2	0	0	0	0
15	Auto	18	2	1	0	30	0
15	Plane	5.5	1	0	1	0	30
15	Transit	20	2	0	0	0	0
16	Auto	23	2	1	0	28	0
16	Plane	5.5	1	0	1	0	28
16	Transit	21.5	2	0	0	0	0
17	Auto	4	2	1	0	44	0
17	Plane	3	1	0	1	0	44
17	Transit	4.5	2	0	0	0	0
18	Auto	5	2	1	0	37	0
18	Plane	2.5	2	0	1	0	37
18	Transit	7	1	0	0	0	0
19	Auto	3.5	1	1	0	45	0
19	Plane	2	2	0	1	0	45
19	Transit	7	2	0	0	0	0
20	Auto	12.5	2	1	0	35	0
20	Plane	3.5	1	0	1	0	35
20	Transit	15.5	2	0	0	0	0
21	Auto	1.5	1	1	0	22	0
21	Plane	4	2	0	1	0	22
21	Transit	2	2	0	0	0	0

C. Output Tables

1. Generalised Logit Model: Candy Choice

Table 1.1: Model Fit Statistics

Model Convergence Status		
Convergence criterion (GCONV=1E-8) satisfied.		
Criterion	Intercept Only	Intercept and Covariates
AIC	125.354	114.448
SC	129.608	127.210
-2 Log L	121.354	102.448

Table 1.2: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.9061	4	0.0008
Score	16.9631	4	0.0020
Wald	12.8115	4	0.0122

Table 1.3: Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Gender	2	4.7168	0.0946
Age	2	12.2325	0.0022

Table 1.4: Analysis of Maximum Likelihood Estimates

Parameter	Candy	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept	lollipop	1	0.7698	0.5782	1.7722	0.1831	
Intercept	sugar	1	-0.9033	0.8664	1.0869	0.2972	
Gender	boy	lollipop	1	1.5758	0.7569	4.3347	0.0373
Gender	boy	sugar	1	1.5261	1.0158	2.2570	0.1330
Age	teenager	lollipop	1	-2.6472	0.7572	12.2212	0.0005
Age	teenager	sugar	1	-1.7416	0.9623	3.2754	0.0703

Table 1.5:Odds Ratio Estimates

Effect	Candy	Point Estimate	95% Confidence Limits	Wald
Gender boy vs girl	lollipop	4.835	1.097	21.313
Gender boy vs girl	sugar	4.600	0.628	33.686
Age teenager vs child	lollipop	0.071	0.016	0.313
Age teenager vs child	sugar	0.175	0.027	1.155

Table 1.6 :Linear Hypothesis Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Test_1	0.0027	1	0.9582
Test_2	1.1585	1	0.2818

Table 1.7: Gender Least Squares Means(Prediction)

Candy	Gender	Estimate	Standard Error	z Value	Pr > z	Mean	Standard Error of Mean	Lower Mean	Upper Mean
chocolate	boy	0.2480	0.5702	0.43	0.6636	0.2193	0.07682	0.06878	0.3699
chocolate	girl	1.7741	0.8127	2.18	0.0290	0.5733	0.1263	0.3257	0.8208
lollipop	boy	1.2701	0.4687	2.71	0.0067	0.6095	0.08804	0.4370	0.7821
lollipop	girl	1.2203	0.8206	1.49	0.1370	0.3295	0.1159	0.1022	0.5567

Table 1.8: Age Least Squares Means(Prediction)

Candy	Age	Estimate	Standard Error	z Value	Pr > z	Mean	Standard Error of Mean	Lower Mean	Upper Mean
chocolate	child	0.1403	0.7067	0.20	0.8427	0.1511	0.06721	0.01939	0.2829
chocolate	teenager	1.8819	0.6583	2.86	0.0043	0.6717	0.09718	0.4813	0.8622
lollipop	child	1.6980	0.5579	3.04	0.0023	0.7175	0.08384	0.5532	0.8818
lollipop	teenager	0.7924	0.6965	1.14	0.2552	0.2260	0.08207	0.06510	0.3868

2. Generalized Logit Model: Travel Choice

Table 2.1: Maximum Likelihood Analysis of Variance

Source	DF	Chi-Square	Pr > ChiSq
Intercept	2	1.72	0.4238
Age	2	1.20	0.5478
Likelihood Ratio	34	42.18	0.1583

Table 2.2: Analysis of Maximum Likelihood Estimates

Parameter	Function Number	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	3.0449	2.4268	1.57	0.2096
	2	2.7212	2.2929	1.41	0.2353
Age	1	-0.0710	0.0652	1.19	0.2762
	2	-0.0500	0.0596	0.70	0.4013

Table 2.3: Predictions

Age	$\exp(\beta'X_i)$	Prob(Auto)	$\exp(\beta'X_i)$	Prob(Plane)
20	5.077911	0.515638	5.591234	0.414085
30	2.496526	0.25351	3.391255	0.251156
40	1.227402	0.124637	2.0569	0.152334
50	0.603445	0.061277	1.247573	0.092395
60	0.29668	0.030126	0.756691	0.05604
70	0.145861	0.014812	0.458956	0.03399
Total	9.847826	1	13.50261	1

3. Conditional Logit Model: Candy Choice

Table 3.1: Model Fit statistics

Convergence Status

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

Criterion	Without Covariates	With Covariates
-2 LOG L	41.589	28.727
AIC	41.589	34.727
SBC	41.589	35.635

Table 3.2: Testing Global Null Hypothesis: $BETA=0$

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.8618	3	0.0049
Score	11.6000	3	0.0089
Wald	8.9275	3	0.0303

Table 3.3: Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
dark	1	1.38629	0.79057	3.0749	0.0795	4.000
soft	1	-2.19722	1.05409	4.3450	0.0371	0.111
nuts	1	0.84730	0.69007	1.5076	0.219	2.333

Table 3.4: Linear Hypothesis Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
test1	7.4199	2	0.0245
test2	5.8526	2	0.0536

Table 3.5: Predictions

j	Dark	Soft	Nuts	$\exp(\beta'X_j)$	P_j
1	0	0	0	1	0.054003
2	0	0	1	2.333334	0.126006
3	0	1	0	0.111056	0.005997

4	0	1	1	0.25913	0.013994
5	1	0	0	3.999999	0.216011
6	1	0	1	9.333331	0.504025
7	1	1	0	0.444223	0.023989
8	1	1	1	1.036521	0.055975

$$\sum \exp(\beta'X_j) = 18.51759 \quad 1$$

4. Conditional Logit Model: Travel Choice

Table 4.1: Model Fit Statistics

Convergence Status

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

Criterion	Without Covariates	With Covariates
-2 LOG L	46.142	33.629
AIC	46.142	35.629
SBC	46.142	36.673

Table 4.2: Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	12.5130	1	0.0004
Score	11.9883	1	0.0005
Wald	6.7551	1	0.0093

Table 4.3: Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
TravTime	1	-0.26549	0.10215	6.7551	0.0093	0.767	TravTime

Table 4.4: Predictions

j		$\exp(\beta'x_j)$	Pj
Auto	4.5	0.302793	0.382335
Plane	10.5	0.061566	0.077739
Transit	3.2	0.4276	0.539926
		$\sum \exp(\beta'x_j)$	1
		=0.791959	

5. Mixed Logit Model: Travel Choice

Table 5.1: Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	46.142	27.464
AIC	46.142	37.464
SBC	46.142	42.687

Table 5.2: Testing Global Null Hypothesis: $BETA=0$

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.6774	5	0.0022
Score	15.3613	5	0.0089
Wald	6.8051	5	0.2355

Table 5.3: Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
Auto	1	2.50069	2.39585	1.0894	0.2966	12.191	Auto
Plane	1	-2.77912	3.52929	0.6201	0.4310	0.062	Plane
AgeAuto	1	-0.07826	0.06332	1.5274	0.2165	0.925	AgeAuto
AgePlane	1	0.01695	0.07439	0.0519	0.8198	1.017	AgePlane
TravTime	1	-0.60845	0.27126	5.0315	0.0249	0.544	TravTime
