

Multinomial Logit Models

Ying So

Warren F. Kuhfeld

Abstract

Multinomial logit models are used to model relationships between a polytomous response variable and a set of regressor variables. The term “multinomial logit model” includes, in a broad sense, a variety of models. The cumulative logit model is used when the response of an individual unit is restricted to one of a finite number of ordinal values. Generalized logit and conditional logit models are used to model consumer choices. This article focuses on the statistical techniques for analyzing discrete choice data and discusses fitting these models using SAS/STAT software.*

Introduction

Multinomial logit models are used to model relationships between a polytomous response variable and a set of regressor variables. 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 ordered model, the response Y of an individual unit is restricted to one of m ordered values. For example, the severity of a medical condition may be: none, mild, and severe. The cumulative logit model assumes that the ordinal nature of the observed response is due to methodological limitations in collecting the data that results in lumping together values of an otherwise continuous response variable (McKelvey and Zavoina 1975). Suppose Y takes values y_1, y_2, \dots, y_m on some scale, where $y_1 < y_2 < \dots < y_m$. It is assumed that the observable variable is a categorized version of a continuous latent variable U such that

$$Y = y_i \Leftrightarrow \alpha_{i-1} < U \leq \alpha_i, i = 1, \dots, m$$

where $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_m = \infty$. It is further assumed that the latent variable U is determined by the explanatory variable vector \mathbf{x} in the linear form $U = -\boldsymbol{\beta}'\mathbf{x} + \epsilon$, where $\boldsymbol{\beta}$ is a vector of regression coefficients and ϵ is a random variable with a distribution function F . It follows that

$$\Pr\{Y \leq y_i | \mathbf{x}\} = F(\alpha_i + \boldsymbol{\beta}'\mathbf{x})$$

If F is the logistic distribution function, the cumulative model is also known as the proportional odds model. You can use PROC LOGISTIC or PROC PROBIT directly to fit the cumulative logit models. Although the cumulative model is the most widely used model for ordinal response data, other useful models include the adjacent-categories logit model and the continuation-ratio model (Agresti 1990).

*This chapter was presented at SUGI 20 by Ying So and can also be found in the SUGI 20 proceedings. Copies of this article (MR-2010G), the other chapters, sample code, and all of the macros are available on the Web http://support.sas.com/resources/papers/tnote/tnote_marketresearch.html.

In an unordered model, the polytomous response variable does not have an ordered structure. Two classes of models, the generalized logit models and the conditional logit models, can be used with nominal response data. The generalized logit model consists of a combination of several binary logits estimated simultaneously. For example, the response variable of interest is the occurrence or nonoccurrence of infection after a Caesarean section with two types of (I,II) infection. Two binary logits are considered: one for type I infection versus no infection and the other for type II infection versus no infection. The conditional logit model has been used in biomedical research to estimate relative risks in matched case-control studies. The nuisance parameters that correspond to the matched sets in an unconditional analysis are eliminated by using a conditional likelihood that contains only the relative risk parameters (Breslow and Day 1980). The conditional logit model was also introduced by McFadden (1974) in the context of econometrics.

In studying consumer behavior, 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.

Consider an example of travel demand. People are asked to choose between travel by auto, plane or public transit (bus or train). The following SAS statements create the data set TRAVEL. The variables `AutoTime`, `PlanTime`, and `TranTime` represent the total travel time required to get to a destination by using auto, plane, or transit, respectively. The variable `Age` represents the age of the individual being surveyed, and the variable `Chosen` contains the individual's choice of travel mode.

```
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

;

In this example, `AutoTime`, `PlanTime`, and `TranTime` are alternative-specific variables, whereas `Age` is a characteristic of the individual. You use a generalized logit model to investigate the relationship between the choice of transportation and `Age`, and you use a conditional logit model to investigate how travel time affects the choice. To study how the choice depends on both the travel time and age of the individual, you need to use a mixed model that incorporates both types of variables.

A survey of the literature reveals a confusion in the terminology for the nominal response models. The term “multinomial logit model” is often used to describe the generalized logit model. The mixed logit is sometimes referred to as the multinomial logit model in which the generalized logit and the conditional logit models are special cases.

The following sections describe discrete choice models, illustrate how to use SAS/STAT software to fit these models, and discuss cross-alternative effects.

Modeling Discrete Choice Data

Consider an individual choosing among m alternatives in a choice set. Let Π_{jk} denote the probability that individual j chooses alternative k , let \mathbf{X}_j represent the characteristics of individual j , and let \mathbf{Z}_{jk} be the characteristics of the k th alternative for individual j . For example, \mathbf{X}_j may be an age and each \mathbf{Z}_{jk} a travel time.

The generalized logit model focuses on the individual as the unit of analysis and uses individual characteristics as explanatory variables. The explanatory variables, being characteristics of an individual, are constant over the alternatives. For example, for each of the m travel modes, $\mathbf{X}_j = (1 \text{ age})'$, and for the first subject, $\mathbf{X}_1 = (1 \ 32)'$. The probability that individual j chooses alternative k is

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

β_1, \dots, β_m are m vectors of unknown regression parameters (each of which is different, even though \mathbf{X}_j is constant across alternatives). Since $\sum_{k=1}^m \Pi_{jk} = 1$, the m sets of parameters are not unique. By setting the last set of coefficients to null (that is, $\beta_m = \mathbf{0}$), the coefficients β_k represent the effects of the \mathbf{X} variables on the probability of choosing the k th alternative over the last alternative. In fitting such a model, you estimate $m - 1$ sets of regression coefficients.

In the conditional logit model, the explanatory variables \mathbf{Z} assume different values for each alternative and the impact of a unit of \mathbf{Z} is assumed to be constant across alternatives. For example, for each of the m travel modes, $\mathbf{Z}_{jk} = (\text{time})'$, and for the first subject, $\mathbf{Z}_{11} = (10)'$, $\mathbf{Z}_{12} = (4.5)'$, and $\mathbf{Z}_{13} = (10.5)'$. The probability that the individual j chooses alternative k is

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

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.

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

$$\Pi_{jk} = \frac{\exp(\beta'_k \mathbf{X}_j + \boldsymbol{\theta}' \mathbf{Z}_{jk})}{\sum_{l=1}^m \exp(\beta'_l \mathbf{X}_j + \boldsymbol{\theta}' \mathbf{Z}_{jl})}$$

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

Fitting Discrete Choice Models

The CATMOD procedure in SAS/STAT software directly fits the generalized logit model. SAS/STAT software does not yet have a procedure that is specially designed to fit the conditional or mixed logit models. However, with some preliminary data processing, you can use the PHREG procedure to fit these models.

The PHREG procedure fits the Cox proportional hazards model to survival data (see the SAS/STAT documentation). The partial likelihood of Breslow has the same form as the likelihood in a conditional logit model.

Let z_l denote the vector of explanatory variables for individual l . Let $t_1 < t_2 < \dots < t_k$ denote k distinct ordered event times. Let d_i denote the number of failures at t_i . Let s_i be the sum of the vectors z_l for those individuals that fail at t_i , and let \mathcal{R}_i denote the set of indices for those who are at risk just before t_i .

The Breslow (partial) likelihood is

$$L_B(\boldsymbol{\theta}) = \prod_{i=1}^k \frac{\exp(\boldsymbol{\theta}' s_i)}{[\sum_{l \in \mathcal{R}_i} \exp(\boldsymbol{\theta}' z_l)]^{d_i}}$$

In a stratified analysis, the partial likelihood is the product of the partial likelihood for each individual stratum. For example, in a study of the time to first infection from a surgery, the variables of a patient consist of **Time** (time from surgery to the first infection), **Status** (an indicator of whether the observation time is censored, with value 2 identifying a censored time), **Z1** and **Z2** (explanatory variables thought to be related to the time to infection), and **Grp** (a variable identifying the stratum to which the observation belongs). The specification in PROC PHREG for fitting the Cox model using the Breslow likelihood is as follows:

```
proc phreg;
  model time*status(2) = z1 z2 / ties=breslow;
  strata grp;
run;
```

To cast the likelihood of the conditional logit model in the form of the Breslow likelihood, consider m artificial observed times for each individual who chooses one of m alternatives. The k th alternative is chosen at time 1; the choices of all other alternatives (second choice, third choice, ...) are not observed and would have been chosen at some later time. So a choice variable is coded with an observed time value of 1 for the chosen alternative and a larger value, 2, for all unchosen (unobserved or censored alternatives). For each individual, there is exactly one event time (1) and $m - 1$ nonevent times, and the risk set just prior to this event time consists of all the m alternatives. For individual j with

alternative-specific characteristics \mathbf{Z}_{jl} , the Breslow likelihood is then

$$L_B(\boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta}'\mathbf{Z}_{jk})}{\sum_{l=1}^m \exp(\boldsymbol{\theta}'\mathbf{Z}_{jl})}$$

This is precisely the probability that individual j chooses alternative k in a conditional logit model. By stratifying on individuals, you get the likelihood of the conditional logit model. Note that the observed time values of 1 and 2 are chosen for convenience; however, the censored times have to be larger than the event time to form the correct risk set.

Before you invoke PROC PHREG to fit the conditional logit, you must arrange your data in such a way that there is a survival time for each individual-alternative. In the example of travel demand, let **Subject** identify the individuals, let **TravTime** represent the travel time for each mode of transportation, and let **Choice** have a value 1 if the alternative is chosen and 2 otherwise. The **Choice** variable is used as the artificial time variable as well as a censoring variable in PROC PHREG. The following SAS statements reshape the data set TRAVEL into data set CHOICE and display the first nine observations:

```
data choice(keep=subject mode travtime choice);
  array times[3] autotime plantime trantime;
  array allmodes[3] $ _temporary_ ('Auto' 'Plane' 'Transit');
  set travel;
  Subject = _n_;
  do i = 1 to 3;
    Mode = allmodes[i];
    TravTime = times[i];
    Choice = 2 - (chosen eq mode);
    output;
  end;
run;

proc print data=choice(obs=9);
run;
```

	Obs	Subject	Mode	Trav Time	Choice
	1	1	Auto	10.0	2
	2	1	Plane	4.5	1
	3	1	Transit	10.5	2
	4	2	Auto	5.5	1
	5	2	Plane	4.0	2
	6	2	Transit	7.5	2
	7	3	Auto	4.5	2
	8	3	Plane	6.0	2
	9	3	Transit	5.5	1

Notice that each observation in TRAVEL corresponds to a block of three observations in CHOICE, exactly one of which is chosen.

The following SAS statements invoke PROC PHREG to fit the conditional logit model. The Breslow likelihood is requested by specifying `ties=breslow`. `Choice` is the artificial time variable, and a value of 2 identifies censored times. `Subject` is used as a stratification variable.

```
proc phreg data=choice;
  model choice*choice(2) = travtime / ties=breslow;
  strata subject;
  title 'Conditional Logit Model Using PHREG';
run;
```

Conditional Logit Model Using PHREG

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

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

To study the relationship between the choice of transportation and the age of people making the choice, the analysis is based on the generalized logit model. You can use PROC CATMOD directly to fit the generalized logit model (see the *SAS/STAT User's Guide*). In the following invocation of PROC CATMOD, `Chosen` is the response variable and `Age` is the explanatory variable:

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

Response Profiles

Response	Chosen
1	Auto
2	Plane
3	Transit

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

Note that 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.

Let \mathbf{X}_j be a $(p+1)$ -vector representing the characteristics of individual j . The generalized logit model can be cast in the framework of a conditional model by defining the global parameter vector $\boldsymbol{\theta}$ and the alternative-specific regressor variables \mathbf{Z}_{jk} as follows:

$$\boldsymbol{\theta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{m-1} \end{bmatrix} \quad \mathbf{Z}_{j1} = \begin{bmatrix} \mathbf{X}_j \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad \mathbf{Z}_{j2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{X}_j \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad \dots \quad \mathbf{Z}_{j,m-1} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{X}_j \end{bmatrix} \quad \mathbf{Z}_{jm} = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

where the $\mathbf{0}$ is a $(p+1)$ -vector of zeros. The probability that individual j chooses alternative k for the generalized logit model is put in the form that corresponds to a conditional logit model as follows:

$$\begin{aligned} \Pi_{jk} &= \frac{\exp(\beta'_k \mathbf{X}_j)}{\sum_{l=1}^m \exp(\beta'_l \mathbf{X}_j)} \\ &= \frac{\exp(\boldsymbol{\theta}' \mathbf{Z}_{jk})}{\sum_{l=1}^m \exp(\boldsymbol{\theta}' \mathbf{Z}_{jl})} \end{aligned}$$

Here, the vector \mathbf{X}_j representing the characteristics of individual j includes the element 1 for the intercept parameter (provided that the intercept parameters are to be included in the model).

By casting the generalized logit model into a conditional logit model, you can then use PROC PHREG to analyze the generalized logit model. In the example of travel demand, the alternative-specific variables `Auto`, `Plane`, `AgeAuto`, and `AgePlane` are created from the individual characteristic variable `Age`. The following SAS statements reshape the data set `TRAVEL` into data set `CHOICE2` and display the first nine observations:

```
data choice2;
  array times[3] autotime plantime trantime;
  array allmodes[3] $ _temporary_ ('Auto' 'Plane' 'Transit');
  set travel;
  Subject = _n_;
  do i = 1 to 3;
    Mode = allmodes[i];
    TravTime = times[i];
    Choice = 2 - (chosen eq mode);
    Auto = (i eq 1);
    Plane = (i eq 2);
    AgeAuto = auto * age;
    AgePlane = plane * age;
    output;
  end;
  keep subject mode travtime choice auto plane ageauto ageplane;
run;

proc print data=choice2(obs=9);
run;
```

Obs	Subject	Mode	Trav Time	Choice	Auto	Plane	Age Auto	Age Plane
1	1	Auto	10.0	2	1	0	32	0
2	1	Plane	4.5	1	0	1	0	32
3	1	Transit	10.5	2	0	0	0	0
4	2	Auto	5.5	1	1	0	13	0
5	2	Plane	4.0	2	0	1	0	13
6	2	Transit	7.5	2	0	0	0	0
7	3	Auto	4.5	2	1	0	41	0
8	3	Plane	6.0	2	0	1	0	41
9	3	Transit	5.5	1	0	0	0	0

The following SAS statements invoke PROC PHREG to fit the generalized logit model:

```
proc phreg data=choice2;
  model choice*choice(2) = auto plane ageauto ageplane /
    ties=breslow;
  strata subject;
  title 'Generalized Logit Model Using PHREG';
run;
```

 Generalized Logit Model Using PHREG

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Auto	1	3.04494	2.42682	1.5743	0.2096	21.009
Plane	1	2.72121	2.29289	1.4085	0.2353	15.199
AgeAuto	1	-0.07097	0.06517	1.1859	0.2762	0.931
AgePlane	1	-0.05000	0.05958	0.7045	0.4013	0.951

By transforming individual characteristics into alternative-specific variables, the mixed logit model can be analyzed as a conditional logit model.

Analyzing the travel demand data for the effects of both travel time and age of individual requires the same data set as the generalized logit model, only now the `TravTime` variable will be used as well. The following SAS statements use PROC PHREG to fit the mixed logit model:

```
proc phreg data=choice2;
  model choice*choice(2) = auto plane ageauto ageplane travtime /
    ties=breslow;
  strata subject;
  title 'Mixed Logit Model Using PHREG';
run;
```

Mixed Logit Model Using PHREG

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

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

A special case of the mixed logit model is the conditional logit model with alternative-specific constants. Each alternative in the model can be represented by its own intercept, which captures the unmeasured desirability of the alternative.

```
proc phreg data=choice2;
  model choice*choice(2) = auto plane travtime / ties=breslow;
  strata subject;
  title 'Conditional Logit Model with Alternative Specific Constants';
run;
```

Conditional Logit Model with Alternative Specific Constants

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Auto	1	-0.11966	0.70820	0.0285	0.8658	0.887
Plane	1	-1.63145	1.24251	1.7241	0.1892	0.196
TravTime	1	-0.48665	0.20725	5.5139	0.0189	0.615

With transit as the reference mode, the intercept for auto, which is negative, may reflect the inconvenience of having to drive over traveling by bus/train, and the intercept for plane may reflect the high expense of traveling by plane over bus/train.

Cross-Alternative Effects

Discrete choice models are often derived from the principle of maximum random utility. It is assumed that an unobserved utility V_k is associated with the k th alternative, and the response function Y is determined by

$$Y = k \Leftrightarrow V_k = \max\{V_l, 1 \leq l \leq m\}$$

Both the generalized logit and the conditional logit models are based on the assumption that V_1, \dots, V_m are independently distributed and each follows an extreme maxima value distribution (Hoffman and Duncan 1988). An important property of such models is Independence from Irrelevant Alternatives (IIA); that is, the ratio of the choice probabilities for any two alternatives for a particular observation is not influenced systematically by any other alternatives. IIA can be tested by fitting a model that contains all the cross-alternative effects and examining the significance of these effects. The cross-alternative effects pick up a variety of IIA violations and other sources of error in the model. (See pages 452, 459, 468, and 679 for other discussions of IIA.)

In the example of travel demand, there may be separate effects for the three travel modes and travel times. In addition, there may be cross-alternative effects for travel times. Not all the effects are estimable, only two of the three intercepts and three of the six cross-alternative effects can be estimated.

The following SAS statements create the design variables for all the cross-alternative effects and display the first nine observations:

```

* Number of alternatives in each choice set;
%let m = 3;

data choice3;
  drop i j k autotime plantime trantime;

  * Values of the variable CHOSEN;
  array allmodes[&m] $
    _temporary_ ('Auto' 'Plane' 'Transit');

  * Travel times for the alternatives;
  array times[&m] autotime plantime trantime;

  * New variables that will contain the design;;
  array inters[&m]
    Auto      /*intercept for auto      */
    Plane     /*intercept for plane     */
    Transit;  /*intercept for transit     */

  array cross[%eval(&m * &m)]
    TimeAuto /*time of auto alternative      */
    PlanAuto /*cross-effect of plane on auto  */
    TranAuto /*cross-effect of transit on auto*/
    AutoPlan /*cross-effect of auto on plane  */
    TimePlan /*time of plane alternative      */
    TranPlan /*cross-effect of transit on plane*/
    AutoTran /*cross-effect of auto on transit */
    PlanTran /*cross-effect of plane on transit*/
    TimeTran; /*time of transit alternative     */

  set travel;

  subject = _n_;

  * Create &m observations for each choice set;
  do i = 1 to &m;
    Mode = allmodes[i]; /* this alternative */
    Travtime = times[i]; /* travel time */
    Choice = 2 - (chosen eq mode); /* 1 - chosen */
    do j = 1 to &m;
      inters[j] = (i eq j); /* mode indicator */
      do k = 1 to &m;
        * (j=k) - time, otherwise, cross-effect;
        cross[&m*(j-1)+k]=times[k]*inters[j];
      end;
    end;
  output;
  end;
run;

```

```
proc print data=choice3(obs=9) label noobs;
  var subject mode travtime choice auto plane transit
      timeauto timeplan timetran autoplan autotran planauto
      plantran tranauto tranplan;
run;
```

subject	Mode	Travtime	Choice	Auto	Plane	Transit			
1	Auto	10.0	2	1	0	0			
1	Plane	4.5	1	0	1	0			
1	Transit	10.5	2	0	0	1			
2	Auto	5.5	1	1	0	0			
2	Plane	4.0	2	0	1	0			
2	Transit	7.5	2	0	0	1			
3	Auto	4.5	2	1	0	0			
3	Plane	6.0	2	0	1	0			
3	Transit	5.5	1	0	0	1			
	Time	Time	Time	Auto	Auto	Plan	Plan	Tran	Tran
	Auto	Plan	Tran	Plan	Tran	Auto	Tran	Auto	Plan
	10.0	0.0	0.0	0.0	0.0	4.5	0.0	10.5	0.0
	0.0	4.5	0.0	10.0	0.0	0.0	0.0	0.0	10.5
	0.0	0.0	10.5	0.0	10.0	0.0	4.5	0.0	0.0
	5.5	0.0	0.0	0.0	0.0	4.0	0.0	7.5	0.0
	0.0	4.0	0.0	5.5	0.0	0.0	0.0	0.0	7.5
	0.0	0.0	7.5	0.0	5.5	0.0	4.0	0.0	0.0
	4.5	0.0	0.0	0.0	0.0	6.0	0.0	5.5	0.0
	0.0	6.0	0.0	4.5	0.0	0.0	0.0	0.0	5.5
	0.0	0.0	5.5	0.0	4.5	0.0	6.0	0.0	0.0

PROC PHREG allows you to specify `test` statements for testing linear hypotheses of the parameters. The test is a Wald test, which is based on the asymptotic normality of the parameter estimators. The following SAS statements invoke PROC PHREG to fit the so called “Mother Logit” model that includes all the cross-alternative effects. The `TEST` statement, with label `IIA`, specifies the null hypothesis that cross-alternative effects `AutoPlan`, `PlanTran`, and `TranAuto` are 0. Since only three cross-alternative effects are estimable and these are the first cross-alternative effects specified in the model, they account for all the cross-alternative effects in the model.

```
proc phreg data=choice3;
  model choice*choice(2) = auto plane transit timeauto timeplan
      timetran autoplan plantran tranauto planauto tranplan
      autotran / ties=breslow;
  IIA: test autoplan,plantran,tranauto;
  strata subject;
  title 'Mother Logit Model';
run;
```

Mother Logit Model

The PHREG Procedure

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	46.142	24.781
AIC	46.142	40.781
SBC	46.142	49.137

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	21.3607	8	0.0062
Score	15.4059	8	0.0517
Wald	6.2404	8	0.6203

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Auto	1	-0.73812	3.05933	0.0582	0.8093	0.478
Plane	1	-3.62435	3.48049	1.0844	0.2977	0.027
Transit	0	0
TimeAuto	1	-2.23433	1.89921	1.3840	0.2394	0.107
TimePlan	1	-0.10112	0.68621	0.0217	0.8829	0.904
TimeTran	1	0.09785	0.70096	0.0195	0.8890	1.103
AutoPlan	1	0.44495	0.68616	0.4205	0.5167	1.560
PlanTran	1	-0.53234	0.63481	0.7032	0.4017	0.587
TranAuto	1	1.66295	1.51193	1.2097	0.2714	5.275
PlanAuto	0	0
TranPlan	0	0
AutoTran	0	0

Linear Hypotheses Testing Results

Label	Wald		Pr > ChiSq
	Chi-Square	DF	
IIA	1.6526	3	0.6475

The χ^2 statistic for the Wald test is 1.6526 with 3 degrees of freedom, indicating that the cross-alternative effects are not statistically significant ($p = .6475$). A generally more preferable way of testing the significance of the cross-alternative effects is to compare the likelihood of the “Mother logit” model with the likelihood of the reduced model with the cross- alternative effects removed. The following SAS statements invoke PROC PHREG to fit the reduced model:

```
proc phreg data=choice3;
  model choice*choice(2) = auto plane transit timeauto
    timeplan timetran / ties=breslow;
  strata subject;
  title 'Reduced Model without Cross-Alternative Effects';
run;
```

Reduced Model without Cross-Alternative Effects

The PHREG Procedure

Convergence Status

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

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	46.142	27.153
AIC	46.142	37.153
SBC	46.142	42.376

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	18.9886	5	0.0019
Score	14.4603	5	0.0129
Wald	7.3422	5	0.1964

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
Auto	1	1.71578	1.80467	0.9039	0.3417	5.561
Plane	1	-3.60073	3.30555	1.1866	0.2760	0.027
Transit	0	0
TimeAuto	1	-0.79543	0.36327	4.7946	0.0285	0.451
TimePlan	1	0.12162	0.58954	0.0426	0.8366	1.129
TimeTran	1	-0.42184	0.25733	2.6873	0.1012	0.656

The chi-squared statistic for the likelihood ratio test of IIA is $(27.153 - 24.781) = 2.372$, which is not statistically significant ($p = .4989$) when compared to a χ^2 distribution with 3 degrees of freedom. This is consistent with the previous result of the Wald test. (See pages 452, 459, 468, and 674 for other discussions of IIA.)

Final Comments

For some discrete choice problems, the number of available alternatives is not the same for each individual. For example, in a study of consumer brand choices of laundry detergents as prices change, data are pooled from different locations, not all of which offer a brand that contains potash. The varying choice sets across individuals can easily be accommodated in PROC PHREG. For individual j who chooses from a set of m_j alternatives, consider m_j artificial times in which the chosen alternative has an event time 1 and the unchosen alternatives have a censored time of 2. The analysis is carried out in the same fashion as illustrated in the previous section.

Unlike the example of travel demand in which data for each individual are provided, choice data are often given in aggregate form, with choice frequencies indicating the repetition of each choice. One way of dealing with aggregate data is to expand the data to the individual level and carry out the analysis as if you have nonaggregate data. This approach is generally not recommended, because it defeats the purpose of having a smaller aggregate data set. PROC PHREG provides a `FREQ` statement that allows you to specify a variable that identifies the frequency of occurrence of each observation. However, with the specification of a `FREQ` variable, the artificial event time is no longer the only event time in a given stratum, but has ties of the given frequency. With proper stratification, the Breslow likelihood is proportional to the likelihood of the conditional logit model. Thus PROC PHREG can be used to obtain parameter estimates and hypothesis testing results for the choice models.

The `ties=discrete` option should not be used instead of the `ties=breslow` option. This is especially detrimental with aggregate choice data because the likelihood that PROC PHREG is maximizing may no longer be the same as the likelihood of the conditional logit model. `ties=discrete` corresponds to the discrete logistic model for genuinely discrete time scale, which is also suitable for the analysis of case-control studies when there is more than one case in a matched set (Gail, Lubin, and Rubinstein 1981). For nonaggregate choice data, all `ties=` options give the same results; however, the resources required for the computation are not the same, with `ties=breslow` being the most efficient.

Once you have a basic understanding of how PROC PHREG works, you can use it to fit a variety of models for the discrete choice data. The major involvement in such a task lies in reorganizing the data to create the observations necessary to form the correct risk sets and the appropriate design variables. There are many options in PROC PHREG that can also be useful in the analysis of discrete choice data. For example, the `offset=` option allows you to restrict the coefficient of an explanatory variable to the value of 1; the `selection=` option allows you to specify one of four methods for selecting variables into the model; the `outest=` option allows you to specify the name of the SAS data set that contains the parameter estimates, based on which you can easily compute the predicted probabilities of the alternatives.

This article deals with estimating parameters of discrete choice models. There is active research in the field of marketing research to use design of experiments to study consumer choice behavior. If you are interested in this area, see Carson et al. (1994), Kuhfeld et al. (1994), and Lazari et al. (1994).