Logit Models for Multicategory Responses

Situation:

- One response variable $Y$ with $J$ levels.
- One or more explanatory or predictor variables. The predictor variables may be quantitative, qualitative or both.

Model: Logistic regression.

What if you have multiple response variables?

There are 3 basic ways in which logistic regression for multicategory responses is different from logistic regression for dichotomous or binary data.

1. Forming Logits.
2. The sampling model.
3. Connections with other models such as Poisson regression and loglinear models.

Additional (general) References:


Additional References on Fitting (Conditional) Multinomial Models using SAS:


Search
Forming Logits

When \( J = 2 \), \( Y \) is dichotomous and we can model logs of odds that an event occurs or does not occur. There is only 1 logit that we can form

\[
\text{logit}(\pi) = \log \left( \frac{\pi}{1 - \pi} \right)
\]

When \( J > 2 \), . . .

- We have a multicategory or “polytomous” or “polychotomous” response variable.
- There are \( J(J-1)/2 \) logits (odds) that we can form, but only \( (J-1) \) are non-redundant.
- There are different ways to form a set of \( (J-1) \) non-redundant logits.

How to “dichotomized” the response \( Y \)?

1. Nominal \( Y \) —
   - (a) “Baseline” logit models or “Multinomial” logistic regression.
   - (b) “Conditional” or “Multinomial” logit models.

2. Ordinal \( Y \) —
   - (a) Cumulative logits.
   - (b) Adjacent categories.
   - (c) Continuation ratios.

(These are the most common and generally the most useful ones).
Sampling Model.

With dichotomous $Y$, at each combination of levels of the explanatory variables, we assume data arise from a Binomial distribution.


Results of clinical study by Junshi Chen of Chinese Academy.

<table>
<thead>
<tr>
<th>Pre-cancerous Mouth Lesions</th>
<th>Worse or no change</th>
<th>Shrink (improve)</th>
</tr>
</thead>
<tbody>
<tr>
<td>green tea</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>placebo</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td></td>
</tr>
</tbody>
</table>

With $J > 2$, at each combination of levels of the explanatory variables, we assume a Multinomial distribution.

e.g.,

<table>
<thead>
<tr>
<th>Pre-cancerous Mouth Lesions</th>
<th>Worse</th>
<th>No change</th>
<th>Shrink</th>
</tr>
</thead>
<tbody>
<tr>
<td>green tea</td>
<td></td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>placebo</td>
<td></td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>
Connections with other Models.

1. Some are equivalent to Poisson regression or loglinear models.

2. Some can be derived from (equivalent to) discrete choice models (e.g., Luce, McFadden).

3. Those that are equivalent to conditional multinomial models are equivalent to proportional hazard models (models for survival data), which is equivalent to Poisson regression model.

4. Some can be derived from latent variable models.

5. Some multcategory logit models are very similar to IRT models in terms of their parametric form. The difference between them is that in the IRT models, the predictor is unobserved (latent), and in the model we discuss here, the predictor variable is observed.

6. Others.
Nominal Response

$Y$ has $J$ categories (order is irrelevant).

$\{\pi_1, \pi_2, \ldots, \pi_J\}$ are probabilities that response is in each category.

$\pi_1 + \pi_2 + \ldots + \pi_J = \sum_{j=1}^{J} \pi_j = 1$.

The probability distribution for the number of outcomes that occur in the $J$ categories for a sample of $n$ independent observations is **Multinomial**.

- The Binomial distribution is a special case of the Multinomial.
- The multinomial distribution depends on $n$ & $\{\pi_1, \pi_2, \ldots, \pi_J\}$.
- The multinomial distribution gives the probability for each way to classify the $n$ observations into the $J$ categories of the response variable.

For example, the possible ways to classify $n = 2$ observations into $J = 3$ categories is

<table>
<thead>
<tr>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Multicategory Logit Models for Nominal Responses.

Possibilities:

1. Baseline or Multinominal logistic regression model. Use characteristics of individuals as predictor variables.

   The parameters differ for each category of the response variable.

2. Conditional Logit model. Use characteristics of the categories of the response variable as the predictors.

   The model parameters are the same for each category of the response variable.

3. Conditional or Mixed logit model. Uses characteristics or attributes of the individuals and the categories as predictor variables.

There is not a standard terminology for these models.

- Agresti (90) regarding 2: “Originally referred to by McFadden as a conditional logit model, it is now usually called the multinomial logit model.”

- Long (97): calls 1 “multinominal logit” model and calls 2 “conditional logit” model.

- Powers & Xie (00) regarding 2 & 3, “However, it is often called a multinominal logit model, leading to a great deal of confusion.”
Baseline/Multinomial Category Logit Model

The models we consider here give a simultaneous representation (summary, description) of the odds of being in one category relative to being in another category for all pairs of categories.

We need a set of \((J - 1)\) non-redundant odds (logits). Given this, we can figure out the odds for any pair of categories.

This model is basically just an extension of the binary logistic regression model.

Consider the HSB data:

Response variable is High school program (HSP) type where

1. General
2. Academic
3. Vo/Tech

Explanatory variables maybe

- Mean of the five achievement test scores, which is numerical/continuous \((x_i)\).
- Socio-economic status, which will be either nominal \((\beta^s_i)\) or ordinal/numerical \((s_i)\).
- School type, which would be nominal (public, private).
We could fit a binary logit model to each pair of program types:

\[
\begin{align*}
\log \left( \frac{\text{general}}{\text{academic}} \right) &= \log \left( \frac{\pi_1(x_i)}{\pi_2(x_i)} \right) = \alpha_1 + \beta_1 x_i \\
\log \left( \frac{\text{academic}}{\text{vo/tech}} \right) &= \log \left( \frac{\pi_2(x_i)}{\pi_3(x_i)} \right) = \alpha_2 + \beta_2 x_i \\
\log \left( \frac{\text{general}}{\text{vo/tech}} \right) &= \log \left( \frac{\pi_1(x_i)}{\pi_3(x_i)} \right) = \alpha_3 + \beta_3 x_i
\end{align*}
\]

We can write one of the odds in terms of the other 2,

\[
\left( \frac{\pi_1(x_i)}{\pi_2(x_i)} \right) \left( \frac{\pi_2(x_i)}{\pi_3(x_i)} \right) = \frac{\pi_1(x_i)}{\pi_3(x_i)},
\]

Therefore, we can find the model parameters of one from the other 2,

\[
\begin{align*}
\log \left( \frac{\pi_1(x_i)}{\pi_2(x_i)} \right) + \log \left( \frac{\pi_2(x_i)}{\pi_3(x_i)} \right) &= \log \left( \frac{\pi_1(x_i)}{\pi_3(x_i)} \right) \\
(\alpha_1 + \beta_1 x_i) + (\alpha_2 + \beta_2 x_i) &= \alpha_3 + \beta_3 x_i
\end{align*}
\]

which means that in the *Population*

\[
\begin{align*}
\alpha_1 + \alpha_2 &= \alpha_3 \\
\beta_1 + \beta_2 &= \beta_3
\end{align*}
\]
With sample data,

- The estimates from separate binary logit models are *consistent* estimators of the parameters of the model.

- Estimates from fitting separate binary logit models will not yield the equality between the parameters that holds in the population.

\[
\hat{\alpha}_1 + \hat{\alpha}_2 \neq \hat{\alpha}_3 \\
\hat{\beta}_1 + \hat{\beta}_2 \neq \hat{\beta}_3
\]

Solution: simultaneous estimation

- Enforces the logical relationships among parameters.

- Uses the data more *efficiently*, which means that the standard errors of parameter estimates are smaller with simultaneous estimation.

Problem: A difficulty is that there are a large number of comparisons and some of them are redundant.

Solution: Choose one of the categories and treat it as a “baseline.” Depending on the study and response variable,

- There maybe a natural choice for the baseline category.

- The choice maybe arbitrary.
Baseline Category Logit Model

For convenience, we’ll use the last level of the response variable as the baseline (i.e., the $J$th level or category).

$$\log \left( \frac{\pi_{ij}}{\pi_{iJ}} \right) \quad \text{for} \quad j = 1, \ldots, J - 1$$

The baseline category logit model with one explanatory variable $x$ is

$$\log \left( \frac{\pi_{ij}}{\pi_{iJ}} \right) = \alpha_j + \beta_j x_i \quad \text{for} \quad j = 1, \ldots, J - 1$$

- For $J = 2$, this is just regular (binary) logistic regression.
  $$\text{logit}(\pi) = \alpha + \beta x$$

- For $J > 2$, $\alpha$ and $\beta$ can differ depending on which 2 categories are being compared.

- The odds for any pair of categories of $Y$ that can be formed are a function of the parameters of the model.

Example: the HSB data where

**Response** variable is High school program (HSP) type where

1. General
2. Academic
3. Vo/Tech

**Explanatory** variable is the mean of the five achievement test scores, which is numerical/continuous ($x_i$).
For this example, we have \((3 - 1) = 2\) non-redundant logits (odds):

\[
\log \left( \frac{\text{general}}{\text{vo/tech}} \right) = \log \left( \frac{\pi_1}{\pi_3} \right) = \alpha_1 + \beta_1 x
\]

\[
\log \left( \frac{\text{academic}}{\text{vo/tech}} \right) = \log \left( \frac{\pi_2}{\pi_3} \right) = \alpha_2 + \beta_2 x
\]

The logit for (1) general and (2) academic equals

\[
\log \left( \frac{\pi_1}{\pi_2} \right) = \log \left( \frac{\pi_1/\pi_3}{\pi_2/\pi_3} \right) = \log(\pi_1/\pi_3) - \log(\pi_2/\pi_3) = (\alpha_1 + \beta_1 x) - (\alpha_2 + \beta_2 x) = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2) x
\]

The differences \((\beta_1 - \beta_2)\) are known as “contrasts”.

**Caution:** You must be certain what the computer program that you use to estimate the model is doing.

- Programs that explicitly estimate the “baseline” logit model generally either set \(\beta_1 = 0\) or set \(\beta_j = 0\), and some set the sum \(\Sigma_j \beta_j = 0\).
- Programs that fit the “multinomial” logit model may set \(\beta_1 = 0\), \(\beta_j = 0\), or \(\Sigma_j \beta_j = 0\).
Again...the estimation should be simultaneous, because

- Simultaneous fitting is more efficient.
- The standard errors of parameter estimates are smaller when model is fit all at once.
- Want to impose the logical relationships among the parameters.

In SAS, this model can be estimated using either

- CATMOD
- GENMOD

Example: estimated model for High School and Beyond

\[
\text{general/votech: } \hat{\log}(\pi_1/\pi_3) = -2.8996 + .0599x \\
\text{academic/votech: } \hat{\log}(\pi_2/\pi_3) = -7.9388 + .1699x
\]

and for comparing general and academic

\[
\hat{\log}(\pi_1/\pi_2) = \hat{\log}(\pi_1/\pi_3) - \hat{\log}(\pi_2/\pi_3) \\
= -2.8996 + .0599x - (-7.9388 + .1699x) \\
= 5.039 - .110x
\]

If we use either general or academic instead of votech as the baseline category, we get the exact same results.
The results using votech as the baseline were obtained using SAS/CATMOD:

```sas
data hsb;
  set sasdata.hsb;
  achieve=(RDG+WRTG+MATH+SCI+CIV)/5;

proc catmod;
  response logits;
  direct achieve;
  model hsp = achieve ;
  title 'Baseline/multinomial logit model: achieve';
```

Edited output from CATMOD:

The CATMOD Procedure

Data Summary

<table>
<thead>
<tr>
<th>Response</th>
<th>HSP</th>
<th>Response Levels</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight Variable</td>
<td>None</td>
<td>Populations</td>
<td>490</td>
</tr>
<tr>
<td>Data Set</td>
<td>HSB</td>
<td>Total Frequency</td>
<td>600</td>
</tr>
<tr>
<td>Frequency Missing</td>
<td>0</td>
<td>Observations</td>
<td>600</td>
</tr>
</tbody>
</table>
### Population Profiles

<table>
<thead>
<tr>
<th>Sample</th>
<th>achieve</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32.94</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>52.76</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>33.74</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>52.78</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>34.98</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>52.8</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>35.3</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>52.82</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>35.36</td>
<td>1</td>
</tr>
<tr>
<td>488</td>
<td>52.64</td>
<td>1</td>
</tr>
<tr>
<td>489</td>
<td>52.66</td>
<td>1</td>
</tr>
<tr>
<td>490</td>
<td>52.7</td>
<td>1</td>
</tr>
</tbody>
</table>

### Response Profiles

<table>
<thead>
<tr>
<th>Response</th>
<th>HSP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

(General) (Academic) (Vo/Tech)
Maximum Likelihood Analysis

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Sub Iteration</th>
<th>-2 Log Likelihood</th>
<th>Convergence Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1318.3347</td>
<td>1.0000</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1087.9513</td>
<td>0.1748</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1083.8083</td>
<td>0.003808</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1083.7834</td>
<td>0.0000230</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1083.7834</td>
<td>1.2814E-9</td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th>Iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>1</td>
<td>-1.8247</td>
<td>-6.7704</td>
<td>0.0349</td>
<td>0.1457</td>
</tr>
<tr>
<td>2</td>
<td>-2.8551</td>
<td>-7.8374</td>
<td>0.0590</td>
<td>0.1677</td>
</tr>
<tr>
<td>3</td>
<td>-2.8992</td>
<td>-7.9380</td>
<td>0.0599</td>
<td>0.1698</td>
</tr>
<tr>
<td>4</td>
<td>-2.8996</td>
<td>-7.9388</td>
<td>0.0599</td>
<td>0.1699</td>
</tr>
</tbody>
</table>

Maximum likelihood computations converged.
### Maximum Likelihood Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2</td>
<td>92.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>achieve</td>
<td>2</td>
<td>112.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>976</td>
<td>920.51</td>
<td>0.8971</td>
</tr>
</tbody>
</table>

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
<th>Number</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td>1</td>
<td>-2.8996</td>
<td>0.8156</td>
<td>12.64</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>-7.9388</td>
<td>0.8438</td>
<td>88.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>achieve</td>
<td></td>
<td>1</td>
<td>0.0599</td>
<td>0.0168</td>
<td>12.77</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>0.1699</td>
<td>0.0168</td>
<td>102.72</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>$e^\beta$</th>
<th>ASE</th>
<th>Wald</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.8996</td>
<td>.8156</td>
<td></td>
<td>12.62</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>-7.9385</td>
<td>.8438</td>
<td></td>
<td>88.51</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Achieve</td>
<td>.0599</td>
<td>1.06</td>
<td>.0169</td>
<td>12.77</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>.1699</td>
<td>1.19</td>
<td>.0168</td>
<td>102.72</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td>$e^{\beta}$</td>
<td>ASE</td>
<td>Wald</td>
<td>$p$</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
<td>------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>Intercept</td>
<td>(general)</td>
<td>-2.8996</td>
<td>.8156</td>
<td>12.62</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>(academic)</td>
<td>-7.9385</td>
<td>.8438</td>
<td>88.51</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Achieve</td>
<td>(general)</td>
<td>.0599</td>
<td>1.06</td>
<td>.0169</td>
<td>12.77</td>
</tr>
<tr>
<td></td>
<td>(academic)</td>
<td>.1699</td>
<td>1.19</td>
<td>.0168</td>
<td>102.72</td>
</tr>
</tbody>
</table>

Notes regarding interpretation:

- For comparing General to Academic,
  \[ \exp(\hat{\beta}_1 - \hat{\beta}_2) = \exp (.0599 - .1699) = \exp (-.110) = 1.12. \]

- For a 10 point change in achievement, yields odds ratios
  
  General to Votech = \( \exp (10(.0599)) = 1.82. \)
  
  Academic to Votech = \( \exp (10(.1699)) = 5.47. \)
  
  General to Academic = \( \exp (10(-.110)) = .33. \)
  
  (or Academic to General = \( 1/.33 = 3.00 \).)
Trick to use SAS/GENMOD: re-arrange the data.

Consider the data as a 2–way, (Student × Program type) table:

<table>
<thead>
<tr>
<th>Student</th>
<th>Program Type</th>
<th>general</th>
<th>academic</th>
<th>vo/tech</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>600</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The saturated loglinear model for this table is

$$\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_{ij}^{SP}$$

Associated with each row/student is a numerical variable, “achieve”. Consider “Student” as being ordinal and fit a nominal by ordinal loglinear model where the achieve test scores $x_i$ are the category scores:

$$\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \beta_j^* x_i$$

We can convert the nominal by ordinal loglinear model into a logit model. For example, comparing General (1) and Vo/Tech (3):

$$\log \left( \frac{\mu_{i1}}{\mu_{i3}} \right) = \log(\mu_{i1}) - \log(\mu_{i3})$$

$$= (\lambda_1^P - \lambda_3^P) + (\beta_1^* - \beta_3^*) x_i$$

$$= \alpha_1 + \beta_1 x_i$$

... our baseline/multinominal model.
Using SAS/GENMOD...

data hsp2;
input student hsp count achieve;
datalines;
  1  1  1  41.32
  1  2  0  41.32
  1  3  0  41.32
  ... ... ...
  600 1  0  43.44
  600 2  0  43.44
  600 3  1  43.44

proc genmod;
class student hsp;
model count = student hsp hsp*achieve / link=log
dist=Poi;

"Student" ensures that the sum of each row of the fitted values
equals 1 (fixed by design) — the $\lambda_i^S$'s or "nuisance" parameters.

"HSP" ensures that the program type margin is fit perfectly — the
$\lambda_j^P$'s which gives us the $\alpha_j$'s in the logit model.

"HSP*achieve" — the $\beta_j^*$ which gives the parameter estimates for
the $\beta_j$'s in the logit model.
Given that SAS/GENMOD sets $\lambda_3^P = 0$ and $\beta_3^* = 0$, you get the correct ASE errors for the $\alpha_j$’s and $\beta_j$’s:

Since

$$\alpha_j = (\lambda_j^P - \lambda_3^P) = \lambda_j^P$$

The ASE of $\alpha_j$ simply equals the ASE of $\lambda_j^P$.

Since

$$\beta_j = (\beta_j^* - \beta_3^*) = \beta_j^*$$

The ASE of $\beta_j$ simply equals ASE of $\beta_j^*$.

Using either CATMOD or GENMOD, you can easily add more explanatory variables. For example,

**GENMOD:**

- SES as a nominal variable:
  
  ```
  proc genmod;
  class student hsp ses;
  model count = student hsp hsp*achieve hsp*ses
  / link=log dist=Poi;
  ```

- SES as a numerical variable (e.g., SES=1,2,3)
  
  ```
  proc genmod;
  class student hsp;
  model count = student hsp hsp*achieve hsp*ses
  / link=log dist=Poi;
  ```
CATMOD:

- SAS as a nominal variable:

```sas
proc catmod;
  response logits;
  direct achieve;
  model hsp = achieve ses ;
  title 'Achieve numerical and SES qualitative';
```

- SAS as a numerical (ordinal) variable:

```sas
proc catmod;
  response logits;
  direct achieve ses ;
  model hsp = achieve ses;
  title 'Both achieve and ses as numerical variable';
```

Other programs that I’ve used to fit multinomial models.


- FORTRAN program that I wrote.

Other programs that I know of (but haven’t used).

- SYSTAT
- SPSS
To illustrate the need for simultaneous estimation ... the binary logistic regression model was fit separately to 2 of the 3 possible logits,

\[
\log \left( \frac{\pi_1}{\pi_3} \right) = \alpha_1 + \beta_1 x \\
\log \left( \frac{\pi_2}{\pi_3} \right) = \alpha_2 + \beta_2 x
\]

yields

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Simultaneous Fit</th>
<th>Separate Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>ASE</td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(general)</td>
<td>-2.8996</td>
<td>.8156</td>
</tr>
<tr>
<td>(academic)</td>
<td>-7.9385</td>
<td>.8438</td>
</tr>
<tr>
<td>Achieve</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(general)</td>
<td>.0599</td>
<td>.0169</td>
</tr>
<tr>
<td>(academic)</td>
<td>.1699</td>
<td>.0168</td>
</tr>
</tbody>
</table>
How well did the (simultaneous) model fit?

Multinomial Logistic Regression
HSP by Achievement

Symbols = data (grouped)
Lines = fitted odds

Observed and Fitted Odds

Mean Achievement

Academic/General
General/VoTech
Multinomial Logistic Regression

Symbol=data (grouped)

Line-fitted (based on model fit to un-grouped data)
Computing Probabilities from the Baseline Logit Model.

Just as in logistic regression for $J = 2$, we can talk about (and interpret) baseline category logit model in terms of probabilities.

The probability of a response being in category $j$ is

$$
\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_{h=1}^{J} \exp(\alpha_h + \beta_h x)}
$$

Note:

- The denominator $\sum_{h=1}^{J} \exp(\alpha_h + \beta_h x)$ ensures that $\sum_{j=1}^{J} \pi_j = 1$.
- $\alpha_J = 0$ and $\beta_J = 0$ (baseline), which is an identification constraint.

Example: High school and beyond

$$
\hat{\pi}_{\text{votech}} = \frac{1}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}
$$

$$
\hat{\pi}_{\text{general}} = \frac{\exp(-2.90 + .06x)}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}
$$

$$
\hat{\pi}_{\text{academic}} = \frac{\exp(-7.94 + .17x)}{1 + \exp(-2.90 + .06x) + \exp(-7.94 + .17x)}
$$

which when plotted versus mean achievement scores...
Multinomial Logistic Regression

HSP by Achievement
Fitted Probabilities

![Graph showing fitted probabilities for different achievement levels.](image)
There are 2 kinds of tests we’ll talk about here:

1. Test whether an explanatory variable is related to the response variable.

2. Test whether the parameters for two (or more) categories of the response variable are the same.

Both of these tests can be done using either Wald or likelihood ratio (LR) tests. We’ll talk about LR tests here; see Long (1997) for the Wald tests.

Test whether an explanatory/predictor variable is not related to the response; that is,

\[ H_0 : \beta_{k1} = \ldots = \beta_{kJ} = 0 \]

for the \( k \)th explanatory variable.

Example of LR test: Consider HSB example but now include SES as a nominal variable and then as an ordinal variable.

From the CATMOD output,

<table>
<thead>
<tr>
<th>Model</th>
<th>(-2\text{Log(like)})</th>
<th>(\Delta df)</th>
<th>(\Delta G^2)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>achieve, nominal SES</td>
<td>1064.6659</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>achieve, ordinal SES</td>
<td>1068.2397</td>
<td>2</td>
<td>3.57</td>
<td>.16</td>
</tr>
<tr>
<td>achieve</td>
<td>1083.7834</td>
<td>2</td>
<td>15.54</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
Test whether 2 response categories have the same parameter estimates (i.e., can they be combined?).

If two response categories, \( j \) and \( j' \), are indistinguishable with respect to the variables in the model, then

\[
H_0 : (\beta_{1j} - \beta_{1j'}) = \ldots = (\beta_{Kj} - \beta_{Kj'}) = 0
\]

for the \( K \) explanatory variables.

Why don’t we have to consider the \( \alpha \)’s?

There are two LR tests that can be used:

**I** Fit the model with no restrictions on the parameters, and then fit the model restricting the parameters to be equal.

**II** Fit a binary logistic regression model to the two response categories in question.

Example: Consider the model with just mean achievement as the explanatory variable.

**Method I:**

<table>
<thead>
<tr>
<th>Multinomial/baseline model</th>
<th>(-2L_{\text{like}})</th>
<th>(\Delta df)</th>
<th>(\Delta G^2)</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No restrictions</td>
<td>1083.7834</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>( \hat{\beta}_1 = \hat{\beta}_3 )</td>
<td>1097.0522</td>
<td>1</td>
<td>13.27</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
Notes regarding Method I:

- This can be done easily using GENMOD.
- The trick is to create a new variable that is used to impose the equality restriction.

```
data hsblong;
  input student hsp count achieve;
  * Create a new dummy variable for equating parameters for votech (hsp=3) and general (hsp=1);
  xhsp=0;
  if hsp=2 then xhsp=1;
datalines;
  1  1  1  41.32000
  1  2  0  41.32000
  1  3  0  41.32000
  2  1  1  45.02000
  2  2  0  45.02000
  2  3  0  45.02000
  3  1  1  34.98000
  ...
  600 1  0  43.44000
  600 2  0  43.44000
  600 3  1  43.44000
```
You can use this method to check whether a sub-set of or specific parameters are equal.

You can use this trick to see if the parameters for more than two response categories are the same.
Method II: Using the binary logistic regression model to test

\[ H_0 : (\beta_{1j} - \beta_{2j'}) = \ldots = (\beta_{Kj} - \beta_{Kj'}) = 0 \]

for the \( K \) explanatory variables.

1. Create a new dataset that only contains the observations from response categories \( j \) and \( j' \).

2. Fit the binary logistic regression model to the new dataset.

3. Compute the likelihood ratio statistic that all the slope coefficients (\( \beta_k \)'s) are simultaneously equal to 0 — not the intercept term..

Example: We have

\[ LR=13.49 \text{ with } df = 1, \ p < .001. \]

Notes regarding Methods I and II:

- In this case, both methods give similar results (Method I: LR= 13.27).

- Method I is more flexible in terms of the range of possible tests that can be performed.

- The Method II is much easier. Just how easy this is,
data hsb;
  set sasdata.hsb;
  achieve=(RDG+WRTG+MATH+SCI+CIV)/5;
  if hsp=2 then delete;

proc logistic descending;
  model hsp = achieve;

Edited Output:

The LOGISTIC Procedure

Model Information

Data Set WORK.HSB
Response Variable HSP
Number of Response Levels 2
Number of Observations 292
Model binary logit
Optimization Technique Fisher’s scoring
Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>HSP</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>147</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>145</td>
</tr>
</tbody>
</table>

Probability modeled is HSP=3.

Model Convergence Status

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

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>406.784</td>
<td>395.293</td>
</tr>
<tr>
<td>SC</td>
<td>410.461</td>
<td>402.647</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>404.784</td>
<td>391.293</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>13.4910</td>
<td>1</td>
<td>0.0002</td>
</tr>
<tr>
<td>Score</td>
<td>13.2434</td>
<td>1</td>
<td>0.0003</td>
</tr>
<tr>
<td>Wald</td>
<td>12.7559</td>
<td>1</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
Baseline/Multinomial Logit model and Grouped Data (Loglinear Model Connection)

When the explanatory/predictor variables are all categorical, the baseline category logit model has an equivalent loglinear model.

Example: Data from Fienberg (1985). In 1963, 2400 men who were rejected for military service because they failed the Armed Forces Qualification Test were interviewed. The data from 2294 of them were as 4–way cross-classification is given below.

The response variable is $E$ (respondent’s education) = grammar school, some HS, HS graduate.

The 3 explanatory variables are

$R$ (Race) = White, Black.

$A$ (age) = under 22, 22 or older.

$F$ (father’s education) = 1 (grammar school), 2 (some HS), 3 (HS graduate), 4 (not available).
Father’s education =

1. Grammar school
2. Some high school
3. High school graduate
4. Not available
First, some loglinear modelling (and then the relationship between baseline category logit model and loglinear models).

All models include $\lambda_{ijk}^{RAF}$ because Race, Age and Father’s education are explanatory variables.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>$G^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(RAF,E)</td>
<td>30</td>
<td>254.8</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,EF)</td>
<td>24</td>
<td>162.6</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,EA)</td>
<td>28</td>
<td>242.7</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,ER)</td>
<td>28</td>
<td>152.8</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,EF,EA)</td>
<td>22</td>
<td>151.5</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,EF,ER)</td>
<td>22</td>
<td>46.7</td>
<td>.002</td>
</tr>
<tr>
<td>(RAF,EA,ER)</td>
<td>26</td>
<td>142.5</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>(RAF,EF,EA,ER)</td>
<td>20</td>
<td>36.9</td>
<td>.01</td>
</tr>
<tr>
<td>(RAF,EFA,ER)</td>
<td>14</td>
<td>27.9</td>
<td>.01</td>
</tr>
<tr>
<td>(RAF,EFR,EA)*</td>
<td>14</td>
<td>18.1</td>
<td>.20</td>
</tr>
<tr>
<td>(RAF,EAR,EA)</td>
<td>18</td>
<td>33.2</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>(RAF,EFA,EFR)</td>
<td>8</td>
<td>9.7</td>
<td>.29</td>
</tr>
</tbody>
</table>

So a good fitting model is

$$
\log(\mu_{ijkl}) = \lambda + \lambda_i^R + \lambda_j^A + \lambda_k^F + \lambda_{ij}^{RA} + \lambda_{jk}^{RF} + \lambda_{jk}^{AF} + \lambda_{ijk}^{RAF} + \lambda_{il}^E + \lambda_{il}^{RE} + \lambda_{kl}^{FE} + \lambda_{ikl}^{FRE} + \lambda_{jl}^{AE}
$$
Loglinear Model (RAF,EFR,EA):

\[
\log(\mu_{ijkl}) = \lambda + \lambda^R_i + \lambda^A_j + \lambda^F_k + \lambda^{RA}_{ij} + \lambda^{RF}_{jk} + \lambda^{AF}_{jk} + \lambda^{RAF}_{ijk} + \lambda^E_l + \lambda^{RE}_{il} + \lambda^{FE}_{kl} + \lambda^{FRE}_{ikl} + \lambda^{AE}_{jl}
\]

Based on this model, the log odds of grammar school (i.e., \(l = 1\)) to high school graduate (i.e., \(l = 3\)) is

\[
\log(\pi_1/\pi_3) = \log(\mu_{ijk1}/\mu_{ijk3}) = \log(\mu_{ijk1}) - \log(\mu_{ijk3}) = (\lambda^E_1 - \lambda^E_3) + (\lambda^{RE}_{i1} - \lambda^{RE}_{i3}) + (\lambda^{AE}_{j1} - \lambda^{AE}_{j3}) + \lambda^{FRE}_{ik}\]

where

\[
\alpha = (\lambda^E_1 - \lambda^E_3) \\
\beta^R_i = (\lambda^{RE}_{i1} - \lambda^{RE}_{i3}) \\
\beta^A_j = (\lambda^{AE}_{j1} - \lambda^{AE}_{j3}) \\
\beta^F_k = (\lambda^{FE}_{k1} - \lambda^{FE}_{k3}) \\
\beta^{RF}_{ik} = (\lambda^{FRE}_{ik1} - \lambda^{FRE}_{ik3})
\]

The logit model for any other pair of categories of the respondent’s education is found in an analogous way (i.e., simply take the difference between appropriate \(\lambda\)’s from the loglinear model).
Estimated parameters of the baseline categories logit model with HS graduate at the baseline:

<table>
<thead>
<tr>
<th></th>
<th>Grammar vs HS grad</th>
<th>Some HS vs HS grad</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>exp((\hat{\beta}))</td>
</tr>
<tr>
<td>(\hat{\alpha})</td>
<td>.73</td>
<td>.90</td>
</tr>
<tr>
<td>(\beta^R_i)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>white</td>
<td>.98</td>
<td>.12</td>
</tr>
<tr>
<td>black</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>(\beta^A_j)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(&lt; 22)</td>
<td>.20</td>
<td>1.22</td>
</tr>
<tr>
<td>(\geq 22)</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>(\beta^F_k)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>-.71</td>
<td>-.63</td>
</tr>
<tr>
<td>2</td>
<td>-1.10</td>
<td>-.47</td>
</tr>
<tr>
<td>3</td>
<td>-1.25</td>
<td>-.39</td>
</tr>
<tr>
<td>4</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>(\beta^{RF}_{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>white 1</td>
<td>.49</td>
<td>1.63</td>
</tr>
<tr>
<td>2</td>
<td>-.23</td>
<td>.79</td>
</tr>
<tr>
<td>3</td>
<td>-1.06</td>
<td>.35</td>
</tr>
<tr>
<td>4</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>black 1</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>3</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>4</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

\(F\) = Father’s education = 1 grammar, 2 some HS, 3 HS graduate, 4 not available.
Some Interpretation:

- **Age**: The odds that someone younger than 22 has had some high school (versus graduating) are 1.55 times larger than the odds for a person older than 22. An older person has had more time to complete HS.

- \(1/.35 = 2.86 \implies\) Given that the father graduated from high school, the odds that a black completes some high school (versus graduates) are 2.86 times larger than the odds that a white completes some high school (versus graduates).
Final few remarks regarding baseline category model

- With the baseline category logit model, there is a single global fit statistic, which is valid if sample size is large enough (eg. for grouped data).

- This model can be used when the categories of the response variable are ordered, but it may not be the best model for the case of ordinal responses.

- The explanatory variable has the same value regardless of which 2 categories/levels of the response variable that are being compared.

- The model and interpretation can be very complex because for each way of forming odds, there are different parameters.

- The multinomial logit model described here can also be derived as a choice model based on random utilities.

- Bock’s nominal response (IRT) model for polytomous items

\[
P(Y = j | \theta) = \frac{\exp(\alpha_j + \beta_j \theta)}{\sum_{h=1}^{J} \exp(\alpha_h + \beta_h \theta)}
\]

where \( \theta \) is an unobserved explanatory variable.
Conditional Logit Model

In Psychology, this is either Bradley & Terry (1952) or the Luce (1959) choice model. In business/economics, this is McFadden’s (1974) conditional logit model.

Situation: Individuals are given a set of possible choices, which differ on certain attributes. We would like to model/predict the probability of choices using the attributes of the choices as explanatory/predictor variables.

Examples:

- Subjects are given 8 chocolate candies and asked which one they like the best (SAS Logistic Regression examples, 1995; Kuhfeld; 2001). The explanatory variables are
  - Type of chocolate: milk or dark
  - Texture: hard or soft
  - Include nuts: nuts or no nuts

- Individuals must choose which of 5 brands of a product that they prefer (SAS Logistic Regression examples, 1995; Kuhfeld; 2001). The explanatory variable is the price of the product. The company presents different combinations of prices for the different brands to see how much of an effect this has on choice behavior.

- The classic example: choice of mode of transportation (eg, train, bus, car). Characteristics or attributes of these include time waiting, how long it takes to get to work, and cost.
The conditional logit model:

- The coefficients of the explanatory variables are the same over the categories (choices) of the response variable.
- The values of the explanatory variables differ over the outcomes (and possibly over individuals).

\[ \pi_j(x_{ij}) = \frac{\exp[\alpha + \beta x_{ij}]}{\sum_{j \in C_i} \exp[\alpha + \beta x_{ij}]} \]

where

\( x_{ij} \) is the value of the explanatory variable for individual \( i \) and response choice \( j \).

The summation in the denominator is over response options/choices that individual \( i \) is given.

Properties of this model:

- The odds that individual \( i \) chooses option \( j \) versus \( k \) is a function of the difference between \( x_{ij} \) and \( x_{ik} \):

\[ \log \left( \frac{\pi_j(x_{ij})}{\pi_k(x_{ik})} \right) = \beta (x_{ij} - x_{ik}) \]

- The odds of choosing \( j \) versus \( k \) does not depend on any of the other options in the choice set or the other options’ values on the attribute variables.

Property of “Independence from Irrelevant Alternatives”.

43
• The multinomial/baseline model can be written in the same form as the conditional logit model (see Agresti (90), p 316-317).

• This model can incorporate attributes or characteristics of the decision maker/individual.

• It can be written as a proportional hazard model.

Examples:

1. Three examples that only include attributes of the response alternatives.

2. An example that includes both attributes of the response alternatives and characteristics of the individual (“mixed model”).
Example 1: chocolates

The model that was fit is

$$\pi_j(c_j, t_j, n_j) = \frac{\exp[\alpha + \beta_1 c_j + \beta_2 t_j + \beta_3 n_j]}{\sum_{h=1}^{8} (\exp[\alpha + \beta_1 c_h + \beta_2 t_h + \beta_3 n_h])}$$

where

- Type of chocolate is dummy coded:
  
  $$c_j = \begin{cases} 
  1 & \text{if milk} \\
  0 & \text{if dark} 
  \end{cases}$$

- Texture is dummy coded:
  
  $$t_j = \begin{cases} 
  1 & \text{if hard} \\
  0 & \text{if soft} 
  \end{cases}$$

- Nuts is dummy coded:
  
  $$n_j = \begin{cases} 
  1 & \text{if no nuts} \\
  0 & \text{if nuts} 
  \end{cases}$$
Or in terms of Odds:
\[
\frac{\pi_j(c_j, t_j, n_j)}{\pi_k(c_k, t_k, n_k)} = \exp[\beta_1(c_j - c_k)] \exp[\beta_2(t_j - t_k)] \exp[\beta_3(n_j - n_k)]
\]

<table>
<thead>
<tr>
<th>parameter</th>
<th>df</th>
<th>value</th>
<th>ASE</th>
<th>Wald</th>
<th>p</th>
<th>exp β</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>1</td>
<td>-2.88</td>
<td>1.03</td>
<td>7.78</td>
<td>.01</td>
<td>—</td>
</tr>
<tr>
<td>Type of chocolate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>milk</td>
<td>1</td>
<td>-1.38</td>
<td>.79</td>
<td>3.07</td>
<td>.08</td>
<td>.25   or (1/.25) = 4.00</td>
</tr>
<tr>
<td>dark</td>
<td>0</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Texture</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hard</td>
<td>1</td>
<td>2.20</td>
<td>1.05</td>
<td>4.35</td>
<td>.04</td>
<td>9.00</td>
</tr>
<tr>
<td>soft</td>
<td>0</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no nuts</td>
<td>1</td>
<td>-.85</td>
<td>.69</td>
<td>1.51</td>
<td>.22</td>
<td>.43   or (1/.43) = 2.33</td>
</tr>
<tr>
<td>nuts</td>
<td>0</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Use exp β for interpretation.
The predicted probabilities.

<table>
<thead>
<tr>
<th>Obs</th>
<th>drk</th>
<th>sft</th>
<th>nts</th>
<th>phat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dark</td>
<td>hard</td>
<td>nuts</td>
<td>0.50400</td>
</tr>
<tr>
<td>2</td>
<td>dark</td>
<td>hard</td>
<td>no n</td>
<td>0.21600</td>
</tr>
<tr>
<td>3</td>
<td>milk</td>
<td>hard</td>
<td>nuts</td>
<td>0.12600</td>
</tr>
<tr>
<td>4</td>
<td>dark</td>
<td>soft</td>
<td>nuts</td>
<td>0.05600</td>
</tr>
<tr>
<td>5</td>
<td>milk</td>
<td>hard</td>
<td>no n</td>
<td>0.05400</td>
</tr>
<tr>
<td>6</td>
<td>dark</td>
<td>soft</td>
<td>no n</td>
<td>0.02400</td>
</tr>
<tr>
<td>7</td>
<td>milk</td>
<td>soft</td>
<td>nuts</td>
<td>0.01400</td>
</tr>
<tr>
<td>8</td>
<td>milk</td>
<td>soft</td>
<td>no n</td>
<td>0.00600</td>
</tr>
</tbody>
</table>
Estimation of the model:

1. SAS Logistic Regression Examples (1995) and Kuhfeld (2001; http://www.sas.com/service/techsup/tnote/tnote_stat.html) describes how this can be done using proc PHREG (proportional hazard regression), which is related to Poisson regression. The “trick” here is to dummy code for “time” so that the non-selected category is 1 and the chosen is 0.

2. The model can be fit as a Poisson regression model using GENMOD by appropriately arranging the data.

The Data file:

data chocs;
  title 'Chocolate Candy Data';
  input subj choose dark soft nuts @@;
  t=2-choose;
  if dark=1 then drk='dark'; else drk='milk'; * Needed for printing out table;
  if soft=1 then sft='soft'; else sft='hard'; * of predicted probabilities;
  if nuts=1 then nts='nuts'; else nts='no nuts';
  datalines;
  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
  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
  ;
Using SAS/GENMOD:

```
proc genmod data=chocs;
  class subj dark soft nuts;
  model choose = dark soft nuts /link=log dist=poi obstats;
  output out=fitted pred=phat;
  title 'Conditional logit model using GENMOD';

data subset;
  merge chocs fitted;
  if subj>1 then delete;

proc sort;
  by descending phat;

proc print;
  var drk sft nts phat;
  title 'Predicted probabilities for different chocolates';
```
Output from GENMOD

Model Information

Data Set WORK.CHOC5
Distribution Poisson
Link Function Log
Dependent Variable choose
Observations Used 80

Class Level Information

Class Levels Values
subj 10 1 2 3 4 5 6 7 8 9 10
dark 2 0 1
soft 2 0 1
nuts 2 0 1

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>76</td>
<td>28.7270</td>
<td>0.3780</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>76</td>
<td>28.7270</td>
<td>0.3780</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>76</td>
<td>66.7195</td>
<td>0.8779</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>76</td>
<td>66.7195</td>
<td>0.8779</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-</td>
<td>-24.3635</td>
<td></td>
</tr>
</tbody>
</table>

Algorithm converged.
Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.8824</td>
<td>1.0334</td>
<td>-4.9078</td>
<td>7.78</td>
<td>0.0053</td>
</tr>
<tr>
<td>dark</td>
<td>0</td>
<td>-1.3863</td>
<td>0.7906</td>
<td>-2.9358</td>
<td>3.07</td>
<td>0.0795</td>
</tr>
<tr>
<td>dark</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td>0</td>
<td>2.1972</td>
<td>1.0541</td>
<td>0.1312</td>
<td>4.35</td>
<td>0.0371</td>
</tr>
<tr>
<td>soft</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nuts</td>
<td>0</td>
<td>-0.8473</td>
<td>0.6901</td>
<td>-2.1998</td>
<td>1.51</td>
<td>0.2195</td>
</tr>
<tr>
<td>nuts</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Using PROC PHREG.

But first, what’s a proportional hazard regression?

- It’s typically used for modeling survival data; that is, modeling the time until death (or other event of interest).
- It’s equivalent to a Poisson regression for the number of deaths and to a negative exponential for survival times.
- For more details see Agresti (1990).

Using SAS PROC PHREG: input

```sas
proc phreg data=chocs outest=betas;
  strata subj;
  model t*choose(0)=dark soft nuts;
  title 'Conditional Logit model fit using PROC PHREG';
run;
```

Output:

The PHREG Procedure

Model Information

<table>
<thead>
<tr>
<th>Data Set</th>
<th>WORK.CHOCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>t</td>
</tr>
<tr>
<td>Censoring Variable</td>
<td>choose</td>
</tr>
<tr>
<td>Censoring Value(s)</td>
<td>0</td>
</tr>
<tr>
<td>Ties Handling</td>
<td>BRESLOW</td>
</tr>
</tbody>
</table>
### Summary of the Number of Event and Censored Values

<table>
<thead>
<tr>
<th>Stratum</th>
<th>subj</th>
<th>Total</th>
<th>Event</th>
<th>Censored</th>
<th>Percent Censored</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>8</td>
<td>1</td>
<td>7</td>
<td>87.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Total</strong></td>
<td><strong>80</strong></td>
<td><strong>10</strong></td>
<td><strong>70</strong></td>
</tr>
</tbody>
</table>

### Convergence Status

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

### Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Without Covariates</th>
<th>With Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 LOG L</td>
<td>41.589</td>
<td>28.727</td>
</tr>
<tr>
<td>AIC</td>
<td>41.589</td>
<td>34.727</td>
</tr>
<tr>
<td>SBC</td>
<td>41.589</td>
<td>35.635</td>
</tr>
</tbody>
</table>

### Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>12.8618</td>
<td>3</td>
<td>0.0049</td>
</tr>
<tr>
<td>Score</td>
<td>11.6000</td>
<td>3</td>
<td>0.0089</td>
</tr>
<tr>
<td>Wald</td>
<td>8.9275</td>
<td>3</td>
<td>0.0303</td>
</tr>
</tbody>
</table>

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>dark</td>
<td>1</td>
<td>1.38629</td>
<td>0.79057</td>
<td>3.0749</td>
<td>0.0795</td>
<td>4.000</td>
</tr>
<tr>
<td>soft</td>
<td>1</td>
<td>-2.19722</td>
<td>1.05409</td>
<td>4.3450</td>
<td>0.0371</td>
<td>0.111</td>
</tr>
<tr>
<td>nuts</td>
<td>1</td>
<td>0.84730</td>
<td>0.69007</td>
<td>1.5076</td>
<td>0.2195</td>
<td>2.333</td>
</tr>
</tbody>
</table>
Example 2: Five brands that differ in terms of price where price is manipulated. For each of the 8 combinations of brand and price included in the study, 100 individuals made their choice.

In all models that we fit, we assume (i.e., fit a parameter) for brand preference.

The two models that are fit:

1. The effect of price does not depend on brand.
2. The effect of price depends on the brand (i.e. the strength of brand loyalty depends on price).

Complex model: $G^2 = 2782.0879$

Simpler model: $G^2 = 2782.4901$

LR statistic for testing whether effect of price depends on brand:

$$G^2 = 2782.4901 - 2782.0879 = .4022, \quad df = 3, \quad p = .94$$

So let’s look at simpler model...
\[
\pi_j(b_{1j}, b_{2j}, b_{3j}, b_{4j}, p_j) = \frac{\exp[\alpha + \beta_1 b_{1j} + \beta_2 b_{2j} + \beta_3 b_{3j} + \beta_4 b_{4j} + \beta_5 p_j]}{\sum_{h=1}^{5} \exp[\alpha + \beta_1 b_{1h} + \beta_2 b_{2h} + \beta_3 b_{3h} + \beta_4 b_{4h} + \beta_5 p_h]}
\]

where

- Brands are dummy coded. Eg,

\[
b_{1j} = \begin{cases} 
1 & \text{if brand is 1} \\
0 & \text{otherwise}
\end{cases}
\]

Note: for the 5th brand, \(b_{1j} = b_{2j} = b_{3j} = b_{4j} = 0\).

- Price is a numerical variable, \(p_j\).

Or in terms of odds:

\[
\frac{\pi_j(b_{1j}, b_{2j}, b_{3j}, b_{4j}, p_j)}{\pi_k(b_{1k}, b_{2k}, b_{3k}, b_{4k}, p_k)} = \frac{\exp[\beta_1 (b_{1j} - b_{1k})] \exp[\beta_2 (b_{2j} - b_{2k})]}{\exp[\beta_3 (b_{3j} - b_{3k})] \exp[\beta_4 (b_{4j} - b_{4k})] \exp[\beta_5 (p_j - p_k)]}
\]

Estimated parameters for the common price model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>(p)</th>
<th>(\exp \hat{\beta})</th>
</tr>
</thead>
<tbody>
<tr>
<td>brand1</td>
<td>(\beta_1)</td>
<td>1</td>
<td>0.66727</td>
<td>0.12305</td>
<td>29.4065</td>
<td>&lt; .0001</td>
<td>1.95</td>
</tr>
<tr>
<td>brand2</td>
<td>(\beta_2)</td>
<td>1</td>
<td>0.38503</td>
<td>0.12962</td>
<td>8.8235</td>
<td>0.0030</td>
<td>1.47</td>
</tr>
<tr>
<td>brand3</td>
<td>(\beta_3)</td>
<td>1</td>
<td>-0.15955</td>
<td>0.14725</td>
<td>1.1740</td>
<td>0.2786</td>
<td>.85</td>
</tr>
<tr>
<td>brand4</td>
<td>(\beta_4)</td>
<td>1</td>
<td>0.98964</td>
<td>0.11720</td>
<td>71.2993</td>
<td>&lt; .0001</td>
<td>2.69</td>
</tr>
<tr>
<td>brand5</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>1.00</td>
</tr>
<tr>
<td>price</td>
<td>(\beta_5)</td>
<td>1</td>
<td>0.14966</td>
<td>0.04406</td>
<td>11.5379</td>
<td>0.0007</td>
<td>1.16</td>
</tr>
</tbody>
</table>
• Which brand is the most preferred?
• Which brand is least preferred?
• What is the effect of price?

How would you interpret \( \exp[.1497] = 1.16 \)?

Estimating the common price effect and the price \( \times \) brand interaction model using SAS:

**GENMOD**: as a Poisson regression model.

**PHREG**: as a proportional hazard model.

First the “raw” data:

```plaintext
data brands;
  title 'Brand Choice Data';
  input p1-p5 f1-f5;
  datalines;
```

55
Format of data needed for input to GENMOD:

data brands2;
  input combo brand price choice @@;
datalines;
  1 1 5.99 12 1 2 5.99 0 1 3 5.99 0 1 4 5.99 0 1 5 4.99 0 
  1 1 5.99 0 1 2 5.99 19 1 3 5.99 0 1 4 5.99 0 1 5 4.99 0 
  1 1 5.99 0 1 2 5.99 0 1 3 5.99 22 1 4 5.99 0 1 5 4.99 0 
  1 1 5.99 0 1 2 5.99 0 1 3 5.99 0 1 4 5.99 33 1 5 4.99 0 
  1 1 5.99 0 1 2 5.99 0 1 3 5.99 0 1 4 5.99 0 1 5 4.99 14 
  2 1 5.99 0 2 2 5.99 0 2 3 3.99 0 2 4 3.99 0 2 5 4.99 0 
  2 1 5.99 0 2 2 5.99 26 2 3 3.99 0 2 4 3.99 0 2 5 4.99 0 
  2 1 5.99 0 2 2 5.99 0 2 3 3.99 8 2 4 3.99 0 2 5 4.99 0 
  2 1 5.99 0 2 2 5.99 0 2 3 3.99 0 2 4 3.99 27 2 5 4.99 0 
  2 1 5.99 0 2 2 5.99 0 2 3 3.99 0 2 4 3.99 0 2 5 4.99 5 
  etc.

PROC GENMOD commands:

proc genmod;
  class combo brand ;
  model choice = combo brand price /link=log dist=poi;
  title 'Brands Model 1 ';

proc genmod;
  class combo brand ;
  model choice = combo brand brand*price /link=log dist=poi;
  title 'Brands Model 2 ';
run;
Some Edited output from Brands Model 1 (common price effect):

The GENMOD Procedure

Model Information

Data Set WORK.BRANDS2
Distribution Poisson
Link Function Log
Dependent Variable choice
Observations Used 200

Class Level Information

Class Levels Values
combo 8 1 2 3 4 5 6 7 8
brand 5 1 2 3 4 5

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>187</td>
<td>2782.4901</td>
<td>14.8796</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>187</td>
<td>2782.4901</td>
<td>14.8796</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>187</td>
<td>4235.1363</td>
<td>22.6478</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>187</td>
<td>4235.1363</td>
<td>22.6478</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>383.9789</td>
<td></td>
</tr>
</tbody>
</table>
Algorithm converged.

Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.3039</td>
<td>0.2249</td>
<td>-0.1368 - 0.7446</td>
<td>1.83</td>
</tr>
<tr>
<td>combo</td>
<td>1</td>
<td>-0.2616</td>
<td>0.1610</td>
<td>-0.5772 - 0.0539</td>
<td>2.64</td>
</tr>
<tr>
<td>combo</td>
<td>2</td>
<td>-0.1370</td>
<td>0.1478</td>
<td>-0.4268 - 0.1527</td>
<td>0.86</td>
</tr>
<tr>
<td>combo</td>
<td>3</td>
<td>-0.1136</td>
<td>0.1460</td>
<td>-0.3997 - 0.1725</td>
<td>0.61</td>
</tr>
<tr>
<td>combo</td>
<td>4</td>
<td>-0.1817</td>
<td>0.1520</td>
<td>-0.4797 - 0.1163</td>
<td>1.43</td>
</tr>
<tr>
<td>combo</td>
<td>5</td>
<td>-0.0951</td>
<td>0.1447</td>
<td>-0.3787 - 0.1885</td>
<td>0.43</td>
</tr>
<tr>
<td>combo</td>
<td>6</td>
<td>-0.1644</td>
<td>0.1503</td>
<td>-0.4590 - 0.1302</td>
<td>1.20</td>
</tr>
<tr>
<td>combo</td>
<td>7</td>
<td>-0.1417</td>
<td>0.1483</td>
<td>-0.4322 - 0.1489</td>
<td>0.91</td>
</tr>
<tr>
<td>combo</td>
<td>8</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 - 0.0000</td>
<td></td>
</tr>
<tr>
<td>brand</td>
<td>1</td>
<td>0.6673</td>
<td>0.1230</td>
<td>0.4261 - 0.9084</td>
<td>29.41</td>
</tr>
<tr>
<td>brand</td>
<td>2</td>
<td>0.3850</td>
<td>0.1296</td>
<td>0.1310 - 0.6391</td>
<td>8.82</td>
</tr>
<tr>
<td>brand</td>
<td>3</td>
<td>-0.1595</td>
<td>0.1472</td>
<td>-0.4481 - 0.1291</td>
<td>1.17</td>
</tr>
<tr>
<td>brand</td>
<td>4</td>
<td>0.9896</td>
<td>0.1172</td>
<td>0.7599 - 1.2194</td>
<td>71.30</td>
</tr>
<tr>
<td>brand</td>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 - 0.0000</td>
<td></td>
</tr>
<tr>
<td>price</td>
<td>1</td>
<td>0.1497</td>
<td>0.0441</td>
<td>0.0633 - 0.2360</td>
<td>11.54</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

- Intercept and combo are nuisance parameters.
- Brand 1 through Brand 5 are $\beta_1 - \beta_4$.
- Price is $\beta_5$. 
Now using PHREG.

The following program is basically from the SAS Logistic Regression (1995) book, which is also pretty much the same as the program in Kuhfeld (2001).

This section puts the data in the format needed for PROC PHREG:

```
data brands3;
  set brands;
  drop p1-p5 f1-f5;

* Define arrays for variables of the original data set;
array p[5] p1-p5; /* Array for prices*/
array f[5] f1-f5; /* Array for frequencies*/

* Define arrays for design matrices in new data set;
array pb[5] price1-price5; /* Array for prices*/
array brand[5] brand1-brand5; /* Array for brands*/;

* Initialize brand and brand by price design matrices;
do j=1 to 5; /* 5=number of choice options*/
  brand[j]=0;
  pb[j]=0;
end;

* Count the total number of choices;
nobs = sum(of f1-f5);
```
* Store choice set number to stratify;

ch_set=_n_;

* Create design matrix;
do j=1 to 5;
   price = p[j];
   brand[j]=1;
   pb[j] = price;

* Output number of times each brand choosen;
freq = f[j];
choose=1;
t = 1;  /* choice occurs at time 1 */
output;

* Output number of times each brand was not choosen;
freq = nobs-f[j];
choose =0;
t = 2;  /* NON choice occurs at time 2 */
output;

* Set up for next alternative;
brand[j] = 0;
pb[j] = 0;
end;
run;
The data looks like:

```
  p  p  p  p  p  b  b  b  b  b
r  r  r  r  r  r  r  r  r  r  r
h  p  h
i  i  i  i  i  a  a  a  a  a  a
n  n  n  n  n  o  s  i  r  o
b  e  e  e  e  e  d  d  d  d  d
b  e  c  e  e  s
s  1  2  3  4  5  1  2  3  4  5  
```

```
1  5.99  0.00  0.00  0.00  0.00  1  0  0  0  0  1  100  1  5.99  12  1  1
2  5.99  0.00  0.00  0.00  0.00  1  0  0  0  0  1  100  1  5.99  88  0  2
3  0.00  5.99  0.00  0.00  0.00  0  1  0  0  0  2  100  1  5.99  19  1  1
4  0.00  5.99  0.00  0.00  0.00  0  1  0  0  0  2  100  1  5.99  81  0  2
5  0.00  0.00  5.99  0.00  0.00  0  1  0  0  3  100  1  5.99  22  1  1
6  0.00  0.00  5.99  0.00  0.00  0  1  0  0  3  100  1  5.99  78  0  2
7  0.00  0.00  0.00  5.99  0.00  0  0  1  0  4  100  1  5.99  33  1  1
8  0.00  0.00  0.00  5.99  0.00  0  0  1  0  4  100  1  5.99  67  0  2
9  0.00  0.00  0.00  0.00  4.99  0  0  0  0  1  5  100  1  4.99  14  1  1
10  0.00  0.00  0.00  0.00  4.99  0  0  0  0  1  5  100  1  4.99  86  0  2
11  5.99  0.00  0.00  0.00  0.00  1  0  0  0  0  1  100  2  5.99  34  1  1
12  5.99  0.00  0.00  0.00  0.00  1  0  0  0  0  1  100  2  5.99  66  0  2
13  0.00  5.99  0.00  0.00  0.00  0  1  0  0  0  2  100  2  5.99  26  1  1
14  0.00  5.99  0.00  0.00  0.00  0  1  0  0  0  2  100  2  5.99  74  0  2
15  0.00  0.00  3.99  0.00  0.00  0  0  1  0  0  3  100  2  3.99  8  1  1
```

The PHREG commands to fit the two models:

```
proc phreg data=brands3;
strata ch_set;
model t*choose(0)=brand1 brand2 brand3 brand4 brand5 price;
freq freq;
title 'PHREG: Discrete choice with common price effect';
```
proc phreg data=brands3;
  strata ch_set;
  model t*choose(0)=brand1-brand5 price1-price5;
  freq freq;
  title 'PHREG: Discrete choice with brand by price effect';
run;

A little edited output:

The PHREG Procedure

Model Information

  Data Set WORK.BRANDS3
  Dependent Variable t
  Censoring Variable choose
  Censoring Value(s) 0
  Frequency Variable freq
  Ties Handling BRESLOW
### Summary of the Number of Event and Censored Values

<table>
<thead>
<tr>
<th>Stratum</th>
<th>ch_set</th>
<th>Total</th>
<th>Event</th>
<th>Censored</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>500</td>
<td>100</td>
<td>400</td>
<td>80.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>4000</strong></td>
<td><strong>800</strong></td>
<td><strong>3200</strong></td>
<td><strong>80.00</strong></td>
</tr>
</tbody>
</table>

### Convergence Status

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

### Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Without Covariates</th>
<th>With Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 LOG L</td>
<td>9943.373</td>
<td>9793.486</td>
</tr>
<tr>
<td>AIC</td>
<td>9943.373</td>
<td>9803.486</td>
</tr>
<tr>
<td>SBC</td>
<td>9943.373</td>
<td>9826.909</td>
</tr>
</tbody>
</table>

The PHREG Procedure

Analysis of Maximum Likelihood Estimates
<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>brand1</td>
<td>1</td>
<td>0.66727</td>
<td>0.12305</td>
<td>29.4065</td>
<td>&lt;.0001</td>
<td>1.949</td>
</tr>
<tr>
<td>brand2</td>
<td>1</td>
<td>0.38503</td>
<td>0.12962</td>
<td>8.8235</td>
<td>0.0030</td>
<td>1.470</td>
</tr>
<tr>
<td>brand3</td>
<td>1</td>
<td>-0.15955</td>
<td>0.14725</td>
<td>1.1740</td>
<td>0.2786</td>
<td>0.853</td>
</tr>
<tr>
<td>brand4</td>
<td>1</td>
<td>0.98964</td>
<td>0.11720</td>
<td>71.2993</td>
<td>&lt;.0001</td>
<td>2.690</td>
</tr>
<tr>
<td>brand5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>price</td>
<td>1</td>
<td>0.14966</td>
<td>0.04406</td>
<td>11.5379</td>
<td>0.0007</td>
<td>1.161</td>
</tr>
</tbody>
</table>

The Response variable is mode of transportation:  
$j = 1$ for train, 2 for bus, and 3 for car. 

Explanatory Variables are:  

$t_{ij} = \text{time waiting in Terminal.}$ 

$v_{ij} = \text{time spent in the Vehicle.}$ 

$c_{ij} = \text{Cost of time spent in vehicle.}$ 

$g_{ij} = \text{Generalized cost measure} = c_{ij} + v_{ij}(value_{ij})$ where value equals subjective value of respondent’s time for each mode of transportation. 

The multinomial logit model that appears to fit the data is 

$$
\pi_{ij} = \frac{\exp[\beta_1 t_{ij} + \beta_2 v_{ij} + \beta_3 c_{ij} + \beta_4 g_{ij}]}{\sum_{h=1}^{3} \exp[\beta_1 t_{ih} + \beta_2 v_{ih} + \beta_3 c_{ih} + \beta_4 g_{ih}]}
$$

The odds of choosing mode $j$ versus mode $k$ for individual $i$,  

$$
\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp[\beta_1 (t_{ij} - t_{ik})] \exp[\beta_2 (v_{ij} - v_{ik})] \exp[\beta_3 (c_{ij} - c_{ik})] \exp[\beta_4 (g_{ij} - g_{ik})]}{\exp[\beta_1 (t_{ik} - t_{ik})] \exp[\beta_2 (v_{ik} - v_{ik})] \exp[\beta_3 (c_{ik} - c_{ik})] \exp[\beta_4 (g_{ik} - g_{ik})]}
$$
The odds of choosing mode $j$ versus mode $k$ for individual $i$,

\[
\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})] \exp[\beta_2(v_{ij} - v_{ik})] \exp[\beta_3(c_{ij} - c_{ik})] \exp[\beta_4(g_{ij} - g_{ik})]
\]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Value</th>
<th>ASE</th>
<th>Wald</th>
<th>$p$-value</th>
<th>$e^\beta$</th>
<th>$1/e^\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>terminal, $t_{ij}$</td>
<td>$\beta_1$</td>
<td>-.002</td>
<td>.007</td>
<td>.98</td>
<td>.75</td>
<td>.99</td>
<td>1.002</td>
</tr>
<tr>
<td>vehicle, $v_{ij}$</td>
<td>$\beta_2$</td>
<td>-.435</td>
<td>.133</td>
<td>10.75</td>
<td>.001</td>
<td>.65</td>
<td>1.55</td>
</tr>
<tr>
<td>cost, $c_{ij}$</td>
<td>$\beta_3$</td>
<td>-.077</td>
<td>.019</td>
<td>15.93</td>
<td>&lt; .001</td>
<td>.03</td>
<td>1.08</td>
</tr>
<tr>
<td>generalized cost, $g_{ij}$</td>
<td>$\beta_4$</td>
<td>.431</td>
<td>.133</td>
<td>10.48</td>
<td>.001</td>
<td>1.54</td>
<td>.65</td>
</tr>
</tbody>
</table>

Odds of choosing a particular mode of transportation decreases as

- Time waiting in terminal increases.
- Time spent in vehicle increases.
- Cost increases.

Odds of choosing a particular model of transportation increases as

- Generalized cost (value of individual’s time) increases
The Mixed Model

The conditional multinomial model that incorporates attributes of the categories (choices) and of the decision maker.

This model is a combination of the multinomial and conditional multinomial modela.

Suppose

- Response variable $Y$ has $J$ categories/levels.
- Explanatory variables

  $x_i$ that is a measure of an attribute of individual $i$
  $w_j$ that is a measure of an attribute of alternative $j$.
  $z_{ij}$ that is a measure of an attribute of alternative $j$ for individual $i$.

The “Mixed” Model:

$$
\pi_j(x_i, w_j, z_{ij}) = \frac{\exp[\alpha_j + \beta_1 x_i + \beta_2 w_j + \beta_3 z_{ij}]}{\sum_{h=1}^{J} \exp[\alpha_h + \beta_1 x_i + \beta_2 w_h + \beta_3 z_{ih}]}
$$

The odds of individual $i$ choosing category $j$ versus category $k$,

$$
\frac{\pi_j(x_i, w_j, z_{ij})}{\pi_k(x_i, w_k, z_{ik})} = \exp[\alpha_j - \alpha_k] \exp[(\beta_1 j - \beta_1 k)x_i] \exp[\beta_2 (w_j - w_k)] \exp[\beta_3 (z_{ij} - z_{ik})]
$$
Transportation Example (continued)...


The Response variable is mode of transportation: 
\( j = 1 \) for train, 2 for bus, and 3 for car.

Explanatory Variables are:

\( t_{ij} \) = time waiting in Terminal.

\( v_{ij} \) = time spent in the Vehicle.

\( c_{ij} \) = Cost of time spent in vehicle.

\( g_{ij} \) = Generalized cost measure = \( c_{ij} + v_{ij}(\text{value}_{ij}) \) where value equals subjective value of respondent’s time for each mode of transportation.

\( h_i \) = Household income.

The mixed model that appears to fit the data is

\[
\pi_{ij} = \frac{\exp[\beta_1 t_{ij} + \beta_2 v_{ij} + \beta_3 c_{ij} + \beta_4 g_{ij} + \alpha_j + \beta_5 j h_i]}{\sum_{h=1}^{3} \exp[\beta_1 t_{ih} + \beta_2 v_{ih} + \beta_3 c_{ih} + \beta_4 g_{ih} + \alpha_h + \beta_5 h_i]}
\]

The odds of choosing mode \( j \) versus mode \( k \) for individual \( i \),

\[
\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp[\beta_1 (t_{ij} - t_{ik})] \exp[\beta_2 (v_{ij} - v_{ik})] \exp[\beta_3 (c_{ij} - c_{ik})] \exp[\beta_4 (g_{ij} - g_{ik})] \exp[(\alpha_j - \alpha_k)] \exp[(\beta_5 j - \beta_5 k) h_i]}{\exp[\beta_1 (t_{ih} - t_{ik})] \exp[\beta_2 (v_{ih} - v_{ik})] \exp[\beta_3 (c_{ih} - c_{ik})] \exp[\beta_4 (g_{ih} - g_{ik})] \exp[(\alpha_j - \alpha_k)] \exp[(\beta_5 h_i)]}
\]
The odds of choosing mode $j$ versus mode $k$ for individual $i$,

$$\frac{\pi_{ij}}{\pi_{ik}} = \exp[\beta_1(t_{ij} - t_{ik})] \exp[\beta_2(v_{ij} - v_{ik})] \exp[\beta_3(c_{ij} - c_{ik})] \exp[\beta_4(g_{ij} - g_{ik})] \exp[(\alpha_j - \alpha_k)] \exp[(\beta_{5j} - \beta_{5k})h_i]$$

Parameter Estimates:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Value</th>
<th>ASE</th>
<th>Wald</th>
<th>$p$-value</th>
<th>$e^\beta$</th>
<th>$1/e^\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal, $t_{ij}$</td>
<td>$\beta_1$</td>
<td>-.074</td>
<td>.017</td>
<td>19.01</td>
<td>&lt; .001</td>
<td>.93</td>
<td>1.08</td>
</tr>
<tr>
<td>Vehicle, $v_{ij}$</td>
<td>$\beta_2$</td>
<td>-.619</td>
<td>.152</td>
<td>16.54</td>
<td>&lt; .001</td>
<td>.54</td>
<td>1.86</td>
</tr>
<tr>
<td>Cost, $c_{ij}$</td>
<td>$\beta_3$</td>
<td>-.096</td>
<td>.022</td>
<td>19.02</td>
<td>&lt; .001</td>
<td>.91</td>
<td>1.10</td>
</tr>
<tr>
<td>Generalized cost, $g_{ij}$</td>
<td>$\beta_4$</td>
<td>.581</td>
<td>.150</td>
<td>15.08</td>
<td>&lt; .001</td>
<td>1.79</td>
<td>.56</td>
</tr>
<tr>
<td>Bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept,</td>
<td>$\alpha_1$</td>
<td>-2.108</td>
<td>.730</td>
<td>6.64</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income, $h_i$</td>
<td>$\beta_{51}$</td>
<td>.031</td>
<td>.021</td>
<td>1.97</td>
<td>.16</td>
<td>1.03</td>
<td>.97</td>
</tr>
<tr>
<td>Car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\alpha_2$</td>
<td>-6.147</td>
<td>1.029</td>
<td>35.70</td>
<td>&lt; .001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income, $h_i$</td>
<td>$\beta_{52}$</td>
<td>.048</td>
<td>.023</td>
<td>7.19</td>
<td>.01</td>
<td>1.05</td>
<td>.95</td>
</tr>
</tbody>
</table>

Effect of household income:

- The odds of choosing a bus versus a train given household income increases from $h_i$ to $h_i + 100$ unit is $\exp(100(.031)) = 22.2$ times larger.

- The odds of choosing a car versus a train given household income increases from $h_i$ to $h_i + 100$ unit is $\exp(100(.048)) = 121.5$ times larger.
• The odds of choosing a car versus a bus given household income increases from $h_i$ to $h_i + 100$ unit is\[ \exp(100(.048 - .031)) = \exp(1.7) = 5.5 \text{ times larger.} \]
Logit Models for ordinal responses

Situation: Polytomous response and categories are ordered.

The logit model for this situation

• Use the ordering of the categories in forming logits.
• Yield simpler models with simpler interpretations than nominal model.
• Is more powerful than nominal models.

Outline:

1. Cumulative logit model, or the “proportional odds” model.
2. Adjacent categories logit model.
3. Continuation ratio logits.
Cumulative Logit Model

**Forming logits** or how to dichotomize categories of $Y$ such that we incorporate the ordinal information.

Use Cumulative Probabilities:

$Y = 1, 2, \ldots, J$ and order is relevant.

$\{\pi_1, \pi_2, \ldots, \pi_J\}$.

$P(Y \leq j) = \pi_1 + \ldots + \pi_j = \sum_{h=1}^{j} \pi_h$ for $j = 1, \ldots, J - 1$.

“Cumulative logits”

$$
\log \left( \frac{P(Y \leq j)}{P(Y > j)} \right) = \log \left( \frac{P(Y \leq j)}{1 - P(Y \leq j)} \right) = \log \left( \frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J} \right) \quad \text{for} \quad j = 1, \ldots, J - 1
$$

The “**Proportional Odds Model**”

$$
\text{logit}[P(Y \leq j)] = \log \left( \frac{P(Y \leq j)}{P(Y > j)} \right) = \alpha_j + \beta x \quad \text{for} \quad j = 1, \ldots, J - 1
$$

- $\alpha_j$ (intercepts) can differ.

- $\beta$ (slope) is constant.

  - The effect of $x$ is the same for all $J - 1$ ways to collapse $Y$ into dichotomous outcomes.

  - A single parameter describes the effect of $x$ on $Y$ (versus $J - 1$ in the baseline model).
Interpretation in terms of odds ratios.

For a given level of $Y$ (say $Y = j$)

$$
\frac{P(Y \leq j|X = x_2)/P(Y > j|X = x_2)}{P(Y \leq j|X = x_1)/P(Y > j|X = x_1)} = \frac{P(Y \leq j|x_2)P(Y > j|x_1)}{P(Y \leq j|x_1)P(Y > j|x_2)}
= \frac{\exp(\alpha_j + \beta x_2)}{\exp(\alpha_j + \beta x_1)}
= \exp[\beta(x_2 - x_1)]
$$
or log odds ratio $= \beta(x_2 - x_1)$.

The odds ratio is proportional to the difference (distance) between $x_1$ and $x_2$ (this is sometimes referred to as “difference model”).

Since the proportionality $= \beta$ is constant, this model is called the “Proportional Odds Model”.

Note that the cumulative probabilities are given by

$$
P(Y \leq j) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)}
$$

Since $\beta$ is constant, curves of cumulative probabilities plotted against $x$ are parallel.

We can compute the probability of being in category $j$ by taking differences between the cumulative probabilities.

$$
P(Y = j) = P(Y \leq j) - P(Y \leq j - 1) \quad \text{for} \quad j = 2, \ldots, J
$$
and

$$
P(Y = 1) = P(Y \leq 1)
$$
Since $\beta$ is constant, these probabilities are guaranteed to be non-negative.

• In fitting this model to data, it must be simultaneous.

In SAS
  – LOGISTIC (maximum likelihood).
  – CATMOD (weighted least squares).

For larger samples with categorical explanatory variables, the results are almost the same.
Example: High School and Beyond

\[ X = \text{mean of 5 achievement test scores.} \]

\[ Y = \text{high school program type} \]
\[
\begin{cases}
  1 & \text{Academic} \\
  2 & \text{General} \\
  3 & \text{VoTech}
\end{cases}
\]

So the logit model is

\[
\begin{align*}
\text{Academic vs (Gen & VoTech):} & \quad \logit(Y \leq 1) = \alpha_1 + \beta x \\
\text{(Academic & Gen) vs VoTech:} & \quad \logit(Y \leq 2) = \alpha_2 + \beta x
\end{align*}
\]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>( e^\beta )</th>
<th>ASE</th>
<th>Wald</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>-6.8408</td>
<td>.6118</td>
<td>125.04</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-5.5138</td>
<td>.5866</td>
<td>88.37</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>( \beta )</td>
<td>.1330</td>
<td>1.142</td>
<td>.0118</td>
<td>127.64</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

For a 10 point increase in mean achievement, the odds ratio (for either case) equals

\[ \exp(10(.1330)) = 3.78 \]
Plot of the estimated cumulative probabilities...

Cumulative Probabilities of High School Program Type

![Graph showing cumulative probabilities for academic and general programs based on mean achievement test scores.](image)
Plot of estimated probabilities of response category.

Probabilities of High School Program Type
Cumulative Probability Logit Model

Estimated Probability

<-- Prob(Vo/Tech)
<-- Prob(Academic)
<-- Prob(General)

Mean Achievement Test Score
In the example, we would get the exact same results regarding interpretation if we had used

\[ Y = \begin{cases} 1 & \text{VoTech} \\ 2 & \text{General} \\ 3 & \text{Academic} \end{cases} \]

This reversal of the ordering of \( Y \) would

- Change the signs of the estimated parameters.
- Yield curves of cumulative probabilities that decrease (rather than increase).

SAS code for the example presented:

```
PROC LOGISTIC;
   MODEL hsp = achieve;
```

Fitting the cumulative logit model is the default if the response variables has more than 2 categories.
Final Comments on Cumulative Logit Models

Nice things about proportional odds model:

- It takes into account the ordering of the categories of the response variable.
- $P(Y=1)$ is monotonically increasing as a function of $x$. (see figure of estimated probabilities).
- $P(Y=J)$ is monotonically decreasing as a function of $x$. (see figure of estimated probabilities).
- Curves of probabilities for intermediate categories are uni-modal with the modes (maximum) corresponding to the order of the categories.
- The conclusions regarding the relationship between $Y$ and $x$ are not affected by the response category.

The specific combination of categories examined does not lead to substantially difference conclusions regarding the relationship between responses and $x$.

If the proportional odds model does not fit well, then you can use the baseline (nominal) model and use the ordering of the responses in your interpretation of the model. For other possibilities, see Long (1997).

IRT connection: Samejima’s (1969) graded response model for polytomous items is the same as the proportional odds model except that $x$ is a latent continuous variable.
Adjacent–Categories Logit Models
for ordinal response

Rather than using all categories in forming logits, we can just use \( J - 1 \) pairs of them.

To incorporate the ordering of the response, we use adjacent categories:

\[
\log \left( \frac{\pi_{j+1}}{\pi_j} \right) \quad j = 1, \ldots, J - 1
\]

The logit model for one (continuous) explanatory variable \( x \) is

\[
\log \left( \frac{\pi_{j+1}}{\pi_j} \right) = \alpha_j + \beta_j x \quad j = 1, \ldots, J - 1
\]

This is similar to the baseline category model in that

- Both \( \alpha \) and \( \beta \) depend on the logit.
- When the explanatory variable is categorical, the logit model has an equivalent loglinear model.

Example: Data from Fienberg (1985) where the response variable is the education level of the rejectees from military service.

Explanatory variables were Race (white, black), Age \((< 22, \geq 22)\), and Father’s education (grammar, some HS, HS graduate, not available).

A good fitting loglinear model was \((\text{RAF,EFR,EA})\), which had \( df = 14, G^2 = 18.1 \).
Estimated parameters for the adjacent category logit model (columns 1 and 3) and for the baseline category model (columns 2 and 3).

<table>
<thead>
<tr>
<th>Adjacent Category Logits</th>
<th>Baseline Logits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grammar vs Some HS</td>
</tr>
<tr>
<td>$\hat{\alpha}$</td>
<td>-.17</td>
</tr>
<tr>
<td>$\hat{\beta}_i^R$</td>
<td>white</td>
</tr>
<tr>
<td></td>
<td>black</td>
</tr>
<tr>
<td>$\hat{\beta}_j^A$</td>
<td>&lt; 22</td>
</tr>
<tr>
<td></td>
<td>$\geq 22$</td>
</tr>
<tr>
<td>$\hat{\beta}_k^F$</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>$\hat{\beta}_{ik}^{RF}$</td>
<td>white</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>black</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

$F =$ Father’s education = 1 grammar, 2 some HS, 3 HS graduate, 4 not available.
A simpler logit model for adjacent categories:

\[
\log \left( \frac{\pi_{j+1}}{\pi_j} \right) = \alpha_j + \beta x \\
\]  
\[j = 1, \ldots, J - 1\]

This is similar to the cumulative logit model in that the effect of \(x\) on \(Y\) is constant across logit (in this case, pairs of categories).

For two categories (say \(Y=1\) and \(Y=4\)), the effect of \(x\) equals 

\[
\beta (4 - 1)
\]

If you have just 1 categorical variable,

- The more complex adjacent categories logit model with \(\beta_j\) and the corresponding loglinear model are saturated.
- The simpler model with \(\beta\) constant is not saturated.

The simpler model is equivalent to a loglinear with linear by linear term for the relationship between the explanatory and response variables.
Example of uniform association model.

General Social Survey (1994) data from before.

**Item 1:** A working mother can establish just as warm and secure of a relationship with her children as a mother who does not work.

**Item 2:** Working women should have paid maternity leave.

<table>
<thead>
<tr>
<th>Item 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>97</td>
<td>96</td>
<td>22</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>102</td>
<td>199</td>
<td>48</td>
<td>38</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>102</td>
<td>25</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>18</td>
<td>7</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

When using $u_i = i$ and $v_j = j$ as scores and fitting the independence loglinear model and the uniform association model

$$
\log(\mu_{ij}) = \lambda + \lambda_i^I + \lambda_j^II + \beta^* i \cdot j
$$

we got

<table>
<thead>
<tr>
<th>Model/Test</th>
<th>df</th>
<th>$G^2$</th>
<th>$p$</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>12</td>
<td>44.96</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Uniform Assoc</td>
<td>11</td>
<td>8.67</td>
<td>.65</td>
<td>$\hat{\beta}^* = .24, ASE = .0412$</td>
</tr>
</tbody>
</table>
Now suppose that we consider item 2 as the response variable and model adjacent category logits with the restriction that $\beta_j = \beta = \text{a constant}$.

\[
\log \left( \frac{\mu_{i,j+1}}{\mu_{i,j}} \right) = \lambda + \lambda_i^I + \lambda_j^{II} + \beta^* i(j + 1) \\
- (\lambda + \lambda_i^I + \lambda_j^{II} + \beta^* i j) \\
= (\lambda_{j+1}^{II} - \lambda_j^{II}) + \beta^* (ij + i - ij) \\
= \alpha_i^* + \beta i
\]

So the estimated local odds ratio equals (and the effect of response on item 1 on item 2 for adjacent categories)

\[
e^\hat{\beta} = e^{.24} = 1.28
\]
Continuation–Ratio Logits
for ordinal responses

In this approach, the order of the categories of the response variable
is incorporated by forming a series of \((J - 1)\) logits

\[
\log\left(\frac{\pi_1}{\pi_2}\right), \log\left(\frac{\pi_1 + \pi_2}{\pi_3}\right), \ldots, \log\left(\frac{\pi_1 + \ldots + \pi_{J-1}}{\pi_J}\right)
\]

or

\[
\log\left(\frac{\pi_1}{\pi_2 + \ldots + \pi_J}\right), \log\left(\frac{\pi_2}{\pi_3 + \ldots + \pi_J}\right), \ldots, \log\left(\frac{\pi_{J-1}}{\pi_J}\right)
\]

These are called “continuation–ratio logits”.

When the models have different parameters for each logit, e.g.,

\[
\alpha_j + \beta_j x
\]

- Just apply regular binary logistic regression to each one.
- The fitting can be separate.
- The sum of the separate \(df\) and \(G^2\) provide an overall global
goodness of fit test and measure (same as simultaneous fitting).

Response variable is $\mathbf{E}$ for respondent’s education and the logits used are

$$\frac{\text{some HS grammar}}{\text{some HS + grammar}}, \quad \text{and} \quad \frac{\text{HS grad}}{\text{some HS + grammar}}$$

Explanatory variables are

- $\mathbf{F}$ for Father’s education
- $\mathbf{A}$ for respondent’s age
- $\mathbf{R}$ for respondent’s race.

Logit models fit to each continuation ratio:

<table>
<thead>
<tr>
<th>Loglinear Model</th>
<th>log($\frac{\mu_{ijk1}}{\mu_{ijk2}}$) $df$</th>
<th>log($\frac{\mu_{ijk3}}{\mu_{ijk1}+\mu_{ijk2}}$) $df$</th>
<th>Combined Fit $df$</th>
<th>$G^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(FAR,E)</td>
<td>15 131.5</td>
<td>15 123.3</td>
<td>30 254.8</td>
<td></td>
</tr>
<tr>
<td>(FAR,EF)</td>
<td>12 97.9</td>
<td>12 64.7</td>
<td>24 162.6</td>
<td></td>
</tr>
<tr>
<td>(FAR,EA)</td>
<td>14 123.3</td>
<td>14 119.4</td>
<td>28 242.7</td>
<td></td>
</tr>
<tr>
<td>(FAR,ER)</td>
<td>14 49.0</td>
<td>14 103.8</td>
<td>28 152.8</td>
<td></td>
</tr>
<tr>
<td>(FAR,EF,EA)</td>
<td>11 91.9</td>
<td>11 60.3</td>
<td>22 152.2</td>
<td></td>
</tr>
<tr>
<td>(FAR,EF,ER)</td>
<td>11 16.1</td>
<td>11 35.6</td>
<td>22 51.7</td>
<td></td>
</tr>
<tr>
<td>(FAR,EA,ER)</td>
<td>13 43.7</td>
<td>13 98.7</td>
<td>26 142.4</td>
<td></td>
</tr>
<tr>
<td>(FAR,EF,EA,ER)</td>
<td>10 12.4</td>
<td>10 29.8</td>
<td>20 42.2</td>
<td></td>
</tr>
<tr>
<td>(FAR,EFA,ER)</td>
<td>7 9.3</td>
<td>7 23.2</td>
<td>14 32.5</td>
<td></td>
</tr>
<tr>
<td>(FAR,EFR,EA)</td>
<td>7 11.5</td>
<td>7 7.0</td>
<td>14 18.5</td>
<td></td>
</tr>
<tr>
<td>(FAR,ERA,EF)</td>
<td>9 8.6</td>
<td>9 29.7</td>
<td>18 38.3</td>
<td></td>
</tr>
<tr>
<td>(FAR,EFA,EFR)</td>
<td>4 8.5</td>
<td>4 1.2</td>
<td>8 9.7</td>
<td></td>
</tr>
</tbody>
</table>
Notes:

• The best fitting loglinear model, which we could use for baseline or adjacent categories model is

\[(\text{FAR,EFR,EA})\]

which has \(df = 14, G^2 = 18.5, p = .18\) for the combined fit, and for each separate logits
\(df = 7, G^2 = 11.5, p = .12\) and \(df = 7, G^2 = 7.0, p = .43\).

• We can find simpler models that fit for the logit comparing Some HS and Grammar School:

\[(\text{FAR,EF,EA,ER})\] with \(df = 10, G^2 = 12.4, p = .23\)
\[(\text{FAR,EF,ER})\] with \(df = 11, G^2 = 16.1, p = .14\).

The likelihood ratio statistic for \(H_o : \lambda_{jl}^{EA} = 0\) equals
\(G^2 = 16.1 - 12.4 = 3.7\) with \(df = 1\), has \(p = .05\).

The simpler of these two models states that given that the respondent did not complete high school, the odds of their completing some high school depends only on race and on their father’s education.

The odds that a respondent completes high school requires a more complex model.