What is a multinomial response?

- Let $Y$ be a categorical response with $J$ categories
- These $J$ categories may be
  1. Nominal – Example: race with 1=white, 2=african american, 3=..., etc.
  2. Ordinal – Example: satisfaction rating with 1=very poor, ... 5 = very pleased
- When $J = 2$, we have our ordinary logistic regression model
- We desire a model to estimate multinomial responses in a manner similar to the logistics models we have developed
- We also want to summarize all of the $\binom{J}{2}$ possible odds ratios using the $J - 1$ non-redundant ORs (as we described previously)
- Some texts and statisticians refer to the nominal models as polytomous logit models
Generalized Odds Ratio

- Recall from Lecture 8 (Contingency Table Extensions) -
- For the $2 \times 2$ table, a single measure can summarize the association.
- For the general $I \times J$ case, a single measure cannot summarize the association without loss of information.

Note: “Loss of information” can be obtained by collapsing the categories into a $2 \times 2$ structure.
### The MI Example

<table>
<thead>
<tr>
<th></th>
<th>Myocardial Infarction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fatal Attack</td>
</tr>
<tr>
<td>Placebo</td>
<td>18</td>
</tr>
<tr>
<td>Aspirin</td>
<td>5</td>
</tr>
</tbody>
</table>

We want to estimate the association of Aspirin Use on MI.
Collapsed Categories

We could collapse the Fatal Attack and Nonfatal Attack categories together to obtain:

<table>
<thead>
<tr>
<th></th>
<th>Myocardial Infarction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal Attack</td>
<td>Placebo: 189, 10845</td>
</tr>
<tr>
<td>or Nonfatal</td>
<td>Aspirin: 104, 10933</td>
</tr>
<tr>
<td>attack</td>
<td></td>
</tr>
</tbody>
</table>

Then, the OR of having a MI is

\[
OR_{MI} = \frac{189 \times 10933}{104 \times 10845} = 1.83
\]

Thus, the odds of a MI are 1.83 times higher when taking placebo when compared to aspirin.
Generalized Odds Ratio

- There are \( \binom{I}{2} \) pairs of rows
- and \( \binom{J}{2} \) pairs of columns
- that can produce \( \binom{I}{2} \binom{J}{2} \) estimates of the odds ratio
- We are going to consider three cases for the generalized odds ratio
Case 1: Arbitrary row and column

For rows $a$ and $b$ and columns $c$ and $d$, the odds ratio ($\pi_{ac}\pi_{bd}/\pi_{bc}\pi_{ad}$) is the most loosely defined set of generalizes ORs.

There are $\left(\binom{I}{2}\right)\left(\binom{J}{2}\right)$ of this type.

For our example, lets compare Fatal MI to No MI.

$$OR_{\text{fatal vs. No MI}} = \frac{18 \times 10933}{5 \times 10845} = 3.63$$

That is, the odds of a having a fatal MI vs No MI are 3.63 times higher for the Placebo group when compared to the group taking Aspirin.
Case 2: Local ORs

The local ORs are obtained by comparing adjacent rows and columns.

That is,

\[ OR_{ij} = \frac{\pi_{ij} \pi_{i+1,j+1}}{\pi_{i+1,j} \pi_{i,j+1}} \]

For our example, we could obtain 2 local ORs

1. Fatal MI vs. Non Fatal MI \((OR = (18 \cdot 99)/(5 \cdot 171) = 2.08)\)
2. Non Fatal MI vs. No MI \((OR = (171 \cdot 10933)/(99 \cdot 10845) = 1.74)\)

Note: There are \((I - 1)(J - 1)\) local odds ratio.
Case 3: Last Column (Reference) OR

For the $I \times J$ table with $I$ representing the last row and $J$ representing the last column, then

$$\alpha_{ij} = \frac{\pi_{ij} \pi_{IJ}}{\pi_{Ij} \pi_{iJ}}, \quad i = 1, 2, \ldots, I - 1, \quad j = 1, 2, \ldots, J - 1$$

represents the OR obtained by referencing the last row and last column. For our example,

1. $\alpha_{11} = \frac{(18 \times 10933)}{(5 \times 10933)} = 3.62$
2. $\alpha_{12} = \frac{(171 \times 10933)}{(99 \times 10845)} = 1.74$
Summary of Generalized Methods

- Here, we have focused on an arbitrary $I \times J$ table
- Just as logistic regression extended the OR for a binary outcome with several predictors
- Multinomial logistic regression will extend the OR estimation for the three cases presented previously to multiple predictors.
• In general, suppose the response for individual $i$ is discrete with $J$ levels:

$$Y_i = \begin{cases} 
1 & \text{if with prob. } p_{i1} \\
2 & \text{if with prob. } p_{i2} \\
. & \\
. & \\
J & \text{if with prob. } p_{iJ}
\end{cases}$$

• Let $x_i$ be the covariates for individual $i$. If $Y_i$ is binary $J = 2$, we usually use a logistic regression model

$$P[Y_i = 1|x_i, \ldots, x_{iK}] = \frac{e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK}}}$$

and we model the logit:

$$\log \left[ \frac{P[Y_i = 1|x_i, \ldots, x_{iK}]}{P[Y_i = 2|x_i, \ldots, x_{iK}]} \right] = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK}$$

Usually, we think of assigning $Y_i = 2$, the value ‘0’.
Polytomous (or Multinomial) Logistic regression

- When \( J = 2 \), we form \( J - 1 = 1 \), non-redundant logits.
- When \( J > 2 \), we often use Polytomous (or Multinomial) Logistic regression, forming \( J - 1 \) non-redundant logits:

\[
\log \left( \frac{P[Y_i = 1|x_{i1}, \ldots, x_{iK}]}{P[Y_i = J|x_{i1}, \ldots, x_{iK}]} \right) = \beta_{10} + \beta_{11} x_{i1} + \ldots + \beta_{1K} x_{iK} = \beta'_1 x_i \\
\log \left( \frac{P[Y_i = 2|x_{i1}, \ldots, x_{iK}]}{P[Y_i = J|x_{i1}, \ldots, x_{iK}]} \right) = \beta_{20} + \beta_{21} x_{i1} + \ldots + \beta_{2K} x_{iK} = \beta'_2 x_i \\
\vdots
\log \left( \frac{P[Y_i = j|x_{i1}, \ldots, x_{iK}]}{P[Y_i = J|x_{i1}, \ldots, x_{iK}]} \right) = \beta_{j0} + \beta_{j1} x_{i1} + \ldots + \beta_{jK} x_{iK} = \beta'_j x_i \\
\vdots
\log \left( \frac{P[Y_i = J - 1|x_{i1}, \ldots, x_{iK}]}{P[Y_i = J|x_{i1}, \ldots, x_{iK}]} \right) = \beta_{J0} + \beta_{J1} x_{i1} + \ldots + \beta_{JK} x_{iK} = \beta'_J x_i
\]
• Note, each one of these logits can have a different set of parameters $\beta_j$.

• Basically, we can think of the $j^{th}$ logit

$$\log \left( \frac{P[Y_i = j|x_{i1}, \ldots, x_{iK}]}{P[Y_i = J|x_{i1}, \ldots, x_{iK}]} \right) = \beta_j' x_i,$$

as a usual logistic regression model when restricting yourself to categories $j$ and $J$.

• Here, we have formulated the “last column (reference)” definition of the generalized OR.
Now, we want to write the probabilities

\[ p_{ij} = P[Y_i = j|x_{i1}, \ldots, x_{iK}], \quad j = 1, \ldots, J, \]

in terms of the parameters and covariates.

Recall, when \( J = 2 \), we write

\[ p_{i1} = \frac{\exp[\beta' x_i]}{1 + \exp[\beta' x_i]} \]

and

\[ p_{i2} = \frac{1}{1 + \exp[\beta' x_i]} \]

We need to generalize this probability formulation when \( J > 2 \).
• For now, consider the following definitions of $p_{ij}$,

\[
p_{ij} = \frac{\exp[\beta_j' x_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]}
\]

when $j < J$,

and

\[
p_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]}
\]

when $j = J$

• We know, $\sum_{j \in J} p_{ij} = 1$
• Using the proposed definitions

\[ \sum_{j=1}^{J} p_{ij} = \sum_{j=1}^{J-1} p_{ij} + p_{iJ} \]

\[ = \sum_{j=1}^{J-1} \left( \frac{\exp[\beta_j' x_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \right) + \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \]

\[ = \frac{\sum_{j=1}^{J-1} \exp[\beta_j' x_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} + \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \]

\[ = \frac{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \]

\[ = 1 \]

• So, our proposed definitions are consistent with a proper probability distribution

• Now, we shall derive the probabilities.
Proof

- Now, consider

\[
\log \left[ \frac{p_{ij}}{p_{iJ}} \right] = \beta_j' x_i
\]

exponentiating both sides, we get

\[
\frac{p_{ij}}{p_{iJ}} = \exp[\beta_j' x_i],
\]

which is the odds for category \( j \) versus category \( J \).

- Multiplying both sides by \( p_{iJ} \), we obtain

\[
p_{ij} = p_{iJ} \exp[\beta_j' x_i],
\]

- Now, suppose we sum both sides over \( j = 1, \ldots, J - 1 \), we get

\[
\sum_{j=1}^{J-1} p_{ij} = p_{iJ} \sum_{j=1}^{J-1} \exp[\beta_j' x_i],
\]
• Note, though

\[ p_{iJ} + \sum_{j=1}^{J-1} p_{ij} = \sum_{j=1}^{J} p_{ij} = 1, \]

i.e.,

\[ \sum_{j=1}^{J-1} p_{ij} = 1 - p_{iJ} \]

so

\[ 1 - p_{iJ} = p_{iJ} \sum_{j=1}^{J-1} \exp[\beta'_j x_i], \]
• Then,

\[ 1 = p_{iJ} + p_{iJ} \sum_{j=1}^{J-1} \exp[\beta_j' x_i] \]

\[ = p_{iJ}(1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]), \]

• Or, finally

\[ p_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \]

and, since

\[ p_{iJ} = p_{iJ} \exp[\beta_j' x_i], \]

substituting in \( p_{iJ} \), we obtain

\[ p_{ij} = \frac{\exp[\beta_j' x_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' x_i]} \]
Interpretation

• It was shown above that the log-odds for category $j$ versus $J$ for covariates $(x_{i1}, \ldots, x_{iK})$ is

$$\log \left[ \frac{p_{ij}}{p_{iJ}} \right] = \beta_{j0} + \beta_{j1}x_{i1} + \ldots + \beta_{jk}x_{ik} + \ldots + \beta_{jK}x_{iK},$$

• We want to know the interpretation of the $\beta_{jk}$'s

• Now, suppose we have two individuals, $i$ and $i'$ with the same values of all the covariates, except that

$$x_{i'k} = x_{ik} + 1.$$  

i.e., all covariates are the same, but the $k^{th}$ covariates are one unit apart.

• Then, the log-odds for subject $i$ is

$$\log \left[ \frac{p_{ij}}{p_{iJ}} \right] = \beta_{j0} + \beta_{j1}x_{i1} + \ldots + \beta_{jk}x_{ik} + \ldots + \beta_{jK}x_{iK},$$

and for subject $i'$ is

$$\log \left[ \frac{p_{i'j}}{p_{i'J}} \right] = \beta_{j0} + \beta_{j1}x_{i1} + \ldots + \beta_{jk}(x_{ik} + 1) + \ldots + \beta_{jK}x_{iK},$$
• Then, subtracting

\[ \log \left[ \frac{p_{ij}}{p_{iJ}} \right] \]

from

\[ \log \left[ \frac{p_{i'j}}{p_{i'J}} \right], \]

we obtain

\[ \log \left[ \frac{p_{i'j}/p_{i'J}}{p_{ij}/p_{iJ}} \right] = \beta_{jk}, \]

i.e.,

\[ \beta_{jk} \]

is the ‘log-odds ratio’ for response \( j \) versus \( J \) for a one unit increase in covariate \( x_{ik} \).
• We have just looked at response \( j \) versus \( J \)
• Using the MI example, \( \beta_{11} \) would be the log-odds of having a fatal MI instead of no MI for subjects on placebo when compared to subjects on aspirin.
• Similarly, \( \beta_{12} \) is the log-odds of having a non-fatal MI instead of a fatal MI.
• Previously, we stated that this model sufficiently describes all possible \((I - 1) \times (J - 1)\) ORs.
• Therefore, we should be able estimate the odd ratio for an arbitrary response \( j \) versus \( j' \).
Now, suppose we want the ‘log-odds ratio’ for response $j'$ versus $j$ for a one unit increase in covariate $x_{ik}$:

$$
\log \left[ \frac{p_{i'j'} / p_{i'j}}{p_{ij} / p_{ij}} \right] = \log \left[ \frac{p_{i'j'} / p_{i'j}}{p_{ij} / p_{ij}} \right] - \log \left[ \frac{p_{i'j} / p_{i'j}}{p_{ij} / p_{ij}} \right] \\
= [\beta_{j'k} - \beta_{jk}]
$$

Then

$$[\beta_{j'k} - \beta_{jk}]$$

is the ‘log-odds ratio’ for response $j'$ versus $j$ for a one unit increase in covariate $x_{ik}$. 
**Estimation Using Proc Logistic**

To estimate the ORs for the MI data using PROC LOGISTIC, we can use the following:

```sas
data mi;
  input x mi count;
  cards;
  1 1 18
  1 2 171
  1 3 10845
  0 1 5
  0 2 99
  0 3 10933
;run;
proc logistic;
  model mi = x /link=glogit; <--- glogit = generalized logit which is our last category reference
  freq count;
run;
```
### Selected Results

#### Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>mi</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>270</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>21778</td>
</tr>
</tbody>
</table>

Logits modeled use mi=3 as the reference category.

#### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>mi</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1</td>
<td>-7.6896</td>
<td>0.4472</td>
<td>295.6688</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>1</td>
<td>-4.7044</td>
<td>0.1010</td>
<td>2171.3642</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x</td>
<td>1</td>
<td>1</td>
<td>1.2885</td>
<td>0.5056</td>
<td>6.4947</td>
<td>0.0108</td>
</tr>
<tr>
<td>x</td>
<td>2</td>
<td>1</td>
<td>0.5546</td>
<td>0.1270</td>
<td>19.0675</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
• In terms of the model,

$$\log \left[ \frac{p_{i1}}{p_{i3}} \right] = \beta_{10} + \beta_{11}x_{i1}$$

$$= -7.6896 + 1.2885x_{i1}$$

and

$$\log \left[ \frac{p_{i2}}{p_{i3}} \right] = -4.7044 + 0.5546x_{i1}$$

where $x_{i1} = 1$ if treated with placebo, 0 else.
Recall, we previously calculated the last category ORs to be

1. $\alpha_{11} = \frac{(18 \times 10^933)}{(5 \times 10^933)} = 3.62$
2. $\alpha_{12} = \frac{(171 \times 10^933)}{(99 \times 10^845)} = 1.74$

We see that PROC LOGISTIC has estimated the same values

Note $e^{1.2885} = 3.627$ which is what is presented above in the “Odds Ratio Estimates”
The odds of having a fatal attack instead of a nonfatal attack is estimated to be

$$OR = \frac{18 \cdot 99}{5 \cdot 171} = 2.08$$

Or from our logistic regression model

$$\log(OR) = \beta_{11} - \beta_{12}$$

$$= 1.2885 - 0.5546$$

$$= 0.7345$$

$$OR = \exp(0.7345)$$

$$= 2.08$$
To write down the multinomial likelihood, we form $J$ indicator random variables ($J - 1$ of which are non-redundant).

$$Y_{ij} = \begin{cases} 
1 & \text{if } Y_i = j \\
0 & \text{otherwise}
\end{cases},$$

$j = 1, ..., J$

Maximum likelihood can be used to estimate the parameters of these models, i.e., maximize

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} p_{ij}^{y_{ij}},$$

as a function of $\beta = [\beta'_1, \beta'_2, ..., \beta'_J]'$
• Then, we obtain the MLE and use the inverse information to estimate its variance.
• Can obtain the MLE in SAS Proc Catmod or Proc Logistic.
• CATMOD is a general modeling PROC that can be used to fit data that can be grouped into a contingency table (i.e., discrete with relatively few levels).
• You can use likelihood ratio (or change in Deviance), Wald or score statistics for hypothesis testing.
• You can also use the Deviance as a goodness-of-fit statistic if the data are grouped multinomial, meaning you have $n_j$ subjects with the same covariate values (and thus the same multinomial distribution).
Example–Primary Food Choice of Alligators

- We are interested in examining the relationship of
  1. Lake:

\[ X_i = \begin{cases} 
1 & \text{if Hancock} \\
2 & \text{if Oklawaha} \\
3 & \text{if Trafford} \\
4 & \text{if George} 
\end{cases} \]

2. GENDER (1 if male, 0 if female)
3. SIZE (1 if \( \leq 2.3 \), 0 if \( > 2.3 \))

- On the choice of food

\[ Y_i = \begin{cases} 
1 & \text{if fish} \\
2 & \text{if invertebrate} \\
3 & \text{if reptile} \\
4 & \text{if bird} \\
5 & \text{if other} 
\end{cases} \]

- Presented in Agresti Page 269

- We want to work to reproduce some of Table 7.2 on pg. 269
data one;
    input lake gender size food count;
cards;
1 1 1 1 7
1 1 1 2 1
1 1 1 3 0
1 1 1 4 0
1 1 1 5 5
1 1 0 1 4
1 1 0 2 0
1 1 0 3 0
1 1 0 4 1
1 1 0 5 2
1 0 1 1 16
... (more data here)
4 0 0 4 0
4 0 0 5 1
proc logistic;
model food(ref='1') = /
    LINK=GLOGIT
    aggregate=(lake size gender) scale=1;
freq count;
run;

• Note, we are creating the $J$ multinomials by looking at unique combinations of lake, size and gender
• This happens to be the way we entered the data
• But, as you can see above, you do not have to estimate a parameter for each to aggregate on them
### Selected Results

#### Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Total food</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>94</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>32</td>
</tr>
</tbody>
</table>

Logits modeled use food=1 as the reference category.

#### Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>116.7611</td>
<td>60</td>
<td>1.9460</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pearson</td>
<td>106.4922</td>
<td>60</td>
<td>1.7749</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Number of unique profiles: 16
## Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>food</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2</td>
<td>1</td>
<td>-0.4324</td>
<td>0.1644</td>
<td>6.9173</td>
<td>0.0085</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>1</td>
<td>-1.5989</td>
<td>0.2515</td>
<td>40.4037</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1</td>
<td>-1.9783</td>
<td>0.2959</td>
<td>44.6984</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>5</td>
<td>1</td>
<td>-1.0776</td>
<td>0.2047</td>
<td>27.7197</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Examine effect of gender

- Since the data are so sparse, we want to see if we can collapse some of the data
- Biologically and ecologically, size of gator and location seem to be important predictors
- Gender, on the other hand, may not be an important characteristic
- To test for the significance of gender, consider the following two models
  1. Gender only compared to intercept only
  2. Gender, Lake and Size compared to only lake and size
- We will calculate change in deviance to assess fit
Model 1: Gender Only

```plaintext
proc logistic;
  class lake size gender;
model food(ref='1') = gender/
    LINK=GLOGIT
    aggregate=(lake size gender) scale=1;
  freq count;
run;
```

- Note: we are still aggregating over lake size and gender
Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>114.6571</td>
<td>56</td>
<td>2.0474</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pearson</td>
<td>101.2480</td>
<td>56</td>
<td>1.8080</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

- The change in deviance is

\[ \Delta D^2 = 116.76116^* - 114.6571 = 2.104 \]

on \( 60 - 56 = 4df \)

- \( p \)-value = 0.72

- So we have evidence that \( \beta_{\text{gender}} = 0 \)

- \( * \) 116.8 is the deviance for the model with intercept only presented earlier

- \( df = 4 \) is because we would estimate 1 gender effect for the 5 – 1 levels of food choice
Model 2

- Similarly, fitting these two models

```r
proc logistic;
   class lake size gender;
model food(ref='1') = lake size gender/
   LINK=GLOGIT
       aggregate=(lake size gender) scale=1;
   freq count;
run;

proc logistic;
   class lake size gender;
model food(ref='1') = lake size /
   LINK=GLOGIT
       aggregate=(lake size gender) scale=1;
   freq count;
run;
```

- Can also assess the significance of gender
### Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model with Lake Size and Gender</td>
<td>50.2637</td>
<td>40</td>
<td>1.2566</td>
<td>0.1282</td>
</tr>
<tr>
<td>Deviance</td>
<td>52.4785</td>
<td>44</td>
<td>1.1927</td>
<td>0.1784</td>
</tr>
</tbody>
</table>

With

\[ \Delta D^2 = 52.4785 - 50.2637 = 2.2148 \]

*p-value* = 0.70

Both models indicate that Gender is not a significant predictor.
If this were an ordinary regression model, we would just “throw away” gender and estimate the following model:

```latex
proc logistic;
  class lake size;
  model food(ref='1') = lake size /
    LINK=GLOGIT
    aggregate=(lake size gender) scale=1;
  freq count;
run;
```

However, what we want to collapse the tables on Gender to increase our cell sizes.

I’ll show two approaches to this.
The Hard Way

- The hard way is to collapse the data across Gender
- Essentially, we want the marginal “table” that results from summing across gender
- Since we still have lake, size and food choice, our “table” has 4 dimensions
Using PROC SQL

- A very useful tool for collapsing data over variables is PROC SQL
- The following code will collapse (or sum the counts) the data over lake, size and food choice

```sql
proc sql;
  create table nogender as
    select lake, size, food, sum(count) as count
    from one
    group by lake, size, food;
run;
```
proc print data=nogender;
run;

<table>
<thead>
<tr>
<th>Obs</th>
<th>lake</th>
<th>size</th>
<th>food</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>8</td>
</tr>
</tbody>
</table>

... 

- Recall, there were previously 4 males and 3 females (or 7) gators living in lake 1, eating fish (food=1) that were > 2.3 (size =0)
- Note, we have summed out the effects of gender
Fitting without Gender

```
proc logistic data=nogender;
  class lake size;
  model food(ref='1') = lake size / 
    link = glogit 
    aggregate scale=1;
  freq count;
run;
```

- Note, I have changed the dataset and modified the aggregate option
- Since the factors to aggregate on are not specified, it uses the covariates in the model
Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>17.0798</td>
<td>12</td>
<td>1.4233</td>
<td>0.1466</td>
</tr>
<tr>
<td>Pearson</td>
<td>15.0429</td>
<td>12</td>
<td>1.2536</td>
<td>0.2391</td>
</tr>
</tbody>
</table>

Number of unique profiles: 8

- Note, number of unique profiles is now 8 (4 lakes times 2 sizes)
- Previously, number of unique profiles equalled 16 (4 lakes, 2 sizes, 2 genders)
- Now, lets consider the easy approach
• Lets go back to “data one” . . . the one with gender unaggregated

• To collapse over gender, all we need to do is aggregate over just lake and size

  proc logistic data=one;
  class lake /param=ref;
  model food(ref='1') = lake size /
      LINK=GLOGIT
      aggregate= (lake size) scale=1;
  freq count;
run;

• This will produce exactly the same model as before, except that I have changed the
dummy variable coding to be reference coding

• I also took size out of the class statement so that we would be estimating the same
  model as Agresti

• Goodness of Fit statistics are unaffected by variable coding convention
Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>17.0798</td>
<td>12</td>
<td>1.4233</td>
<td>0.1466</td>
</tr>
<tr>
<td>Pearson</td>
<td>15.0429</td>
<td>12</td>
<td>1.2536</td>
<td>0.2391</td>
</tr>
</tbody>
</table>

Number of unique profiles: 8

Now, we will examine the parameter estimates.
## Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>food</th>
<th>DF</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>2</td>
<td>1</td>
<td>-1.5490</td>
<td>0.4249</td>
<td>13.2890</td>
<td>0.0003</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>1</td>
<td>-3.3139</td>
<td>1.0528</td>
<td>9.9081</td>
<td>0.0016</td>
</tr>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1</td>
<td>-2.0931</td>
<td>0.6622</td>
<td>9.9894</td>
<td>0.0016</td>
</tr>
<tr>
<td>Intercept</td>
<td>5</td>
<td>1</td>
<td>-1.9043</td>
<td>0.5258</td>
<td>13.1150</td>
<td>0.0003</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>2</td>
<td>-1.6583</td>
<td>0.6129</td>
<td>7.3216</td>
<td>0.0068</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>3</td>
<td>1.2422</td>
<td>1.1852</td>
<td>1.0985</td>
<td>0.2946</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>4</td>
<td>0.6951</td>
<td>0.7813</td>
<td>0.7916</td>
<td>0.3736</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>5</td>
<td>0.8262</td>
<td>0.5575</td>
<td>2.1959</td>
<td>0.1384</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>2</td>
<td>0.9372</td>
<td>0.4719</td>
<td>3.9443</td>
<td>0.0470</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>3</td>
<td>2.4583</td>
<td>1.1179</td>
<td>4.8360</td>
<td>0.0279</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>4</td>
<td>-0.6532</td>
<td>1.2021</td>
<td>0.2953</td>
<td>0.5869</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>5</td>
<td>0.00565</td>
<td>0.7766</td>
<td>0.0001</td>
<td>0.9942</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>2</td>
<td>1.1220</td>
<td>0.4905</td>
<td>5.2321</td>
<td>0.0222</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>3</td>
<td>2.9347</td>
<td>1.1161</td>
<td>6.9131</td>
<td>0.0086</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>4</td>
<td>1.0878</td>
<td>0.8417</td>
<td>1.6703</td>
<td>0.1962</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>5</td>
<td>1.5164</td>
<td>0.6214</td>
<td>5.9541</td>
<td>0.0147</td>
</tr>
<tr>
<td>size</td>
<td>2</td>
<td>1</td>
<td>1.4582</td>
<td>0.3959</td>
<td>13.5634</td>
<td>0.0002</td>
</tr>
<tr>
<td>size</td>
<td>3</td>
<td>1</td>
<td>-0.3513</td>
<td>0.5800</td>
<td>0.3668</td>
<td>0.5448</td>
</tr>
<tr>
<td>size</td>
<td>4</td>
<td>1</td>
<td>-0.6307</td>
<td>0.6425</td>
<td>0.9635</td>
<td>0.3263</td>
</tr>
<tr>
<td>size</td>
<td>5</td>
<td>1</td>
<td>0.3316</td>
<td>0.4483</td>
<td>0.5471</td>
<td>0.4595</td>
</tr>
</tbody>
</table>
Thus, the estimated model for estimating the log(odds) of an alligator eating invertebrate animals instead of fish would be

$$log(\pi_{inv}/\pi_{fish}) = -1.5490 + 1.4582 \text{ Size} -1.6583 \text{ lake 1}$$
$$0.9372 \text{ lake 2} + 1.1220 \text{ lake 3}$$

Thus, in a given lake (or controlling for the effects of lake), the estimated odds that primary food choice was invertebrates instead of fish for small alligators ($\leq 2.3$) are $exp(1.4582) = 4.3$ times the estimated odds for large alligators.
## Odds Ratio Summary

### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>food</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>lake 1 vs 4</td>
<td>2</td>
<td>0.190</td>
<td>0.057</td>
</tr>
<tr>
<td>lake 1 vs 4</td>
<td>3</td>
<td>3.463</td>
<td>0.339</td>
</tr>
<tr>
<td>lake 1 vs 4</td>
<td>4</td>
<td>2.004</td>
<td>0.433</td>
</tr>
<tr>
<td>lake 1 vs 4</td>
<td>5</td>
<td>2.285</td>
<td>0.766</td>
</tr>
<tr>
<td>lake 2 vs 4</td>
<td>2</td>
<td>2.553</td>
<td>1.012</td>
</tr>
<tr>
<td>lake 2 vs 4</td>
<td>3</td>
<td>11.685</td>
<td>1.306</td>
</tr>
<tr>
<td>lake 2 vs 4</td>
<td>4</td>
<td>0.520</td>
<td>0.049</td>
</tr>
<tr>
<td>lake 2 vs 4</td>
<td>5</td>
<td>1.006</td>
<td>0.219</td>
</tr>
<tr>
<td>lake 3 vs 4</td>
<td>2</td>
<td>3.071</td>
<td>1.174</td>
</tr>
<tr>
<td>lake 3 vs 4</td>
<td>3</td>
<td>18.815</td>
<td>2.111</td>
</tr>
<tr>
<td>lake 3 vs 4</td>
<td>4</td>
<td>2.968</td>
<td>0.570</td>
</tr>
<tr>
<td>lake 3 vs 4</td>
<td>5</td>
<td>4.556</td>
<td>1.348</td>
</tr>
<tr>
<td>size</td>
<td>2</td>
<td>4.298</td>
<td>1.978</td>
</tr>
<tr>
<td>size</td>
<td>3</td>
<td>0.704</td>
<td>0.226</td>
</tr>
<tr>
<td>size</td>
<td>4</td>
<td>0.532</td>
<td>0.151</td>
</tr>
<tr>
<td>size</td>
<td>5</td>
<td>1.393</td>
<td>0.579</td>
</tr>
</tbody>
</table>
Summary

- We see that the odds of eating invertebrates instead of fish are higher for lakes 2 and 3 when compared to lake 4, but in lake 1, alligators of either size are less likely to eat invertebrates.
- This could be because there are more fish in lake 1 or the alligators in lake 1 somehow prefer the taste of fish to the invertebrates.
- We see that small alligators prefer invertebrates and “other” instead of fish, after controlling for lakes.
- Whereas the odds of a larger alligators preferring reptile or birds are higher than for small alligators.
- Here, food preference is likely a function of hunting ability... alligators that can catch (and swallow) birds are likely more experienced hunters and older (thus larger).
Recap

- When assessing nested goodness of fit, you need to consider your $n_j$ multinomials.
- Keep your aggregate function consistent so that your models are properly nested.
- Once you decide to eliminate a parameter, you may adjust your aggregate appropriately.
- This model is commonly called the **Baseline Category** model.
- It is used for **NOMINAL OUTCOMES**.
- We will examine a simplification of this model for ordinal outcomes in the next lecture.