Lecture 20: Logit Models for Multinomial Responses

Dipankar Bandyopadhyay, Ph.D.

BMTRY 711: Analysis of Categorical Data Spring 2011

Division of Biostatistics and Epidemiology

Medical University of South Carolina

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 $\begin{pmatrix} J \\ 2 \end{pmatrix}$ 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 2×2 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 2×2 structure.

The MI Example

Agresti Table 2.1 - Page 37			
	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10845
Aspirin	5	99	10933

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

	Myocardial Infarction	
	Fatal Attack or Nonfatal attack	No Attack
Placebo	189	10845
Aspirin	104	10933

Then, the OR of having a MI is

$$\begin{aligned} OR_{MI} &= \frac{189*10933}{104*10845} \\ &= 1.83 \end{aligned}$$

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 $\begin{pmatrix} I \\ 2 \end{pmatrix}\begin{pmatrix} J \\ 2 \end{pmatrix}$ of this type.

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

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

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. $\alpha_{11} = (18 * 10933)/(5 * 10933) = 3.62$
2. $\alpha_{12} = (171 * 10933)/(99 * 10845) = 1.74$

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:

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

- Let $\mathbf{x}_i$ be the covariates for individual i. If Y_i is binary $J = 2$, we usually use a logistic regression model

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

and we model the logit:

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

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:

$$\log\left[\frac{P[Y_i = 1|x_{i1}, ..., x_{iK}]}{P[Y_i = J|x_{i1}, ..., x_{iK}]}\right] = \beta_{10} + \beta_{11}x_{i1} + ... + \beta_{1K}x_{iK} = \beta_1'\mathbf{x}_i$$

$$\log\left[\frac{P[Y_i = 2|x_{i1}, ..., x_{iK}]}{P[Y_i = J|x_{i1}, ..., x_{iK}]}\right] = \beta_{20} + \beta_{21}x_{i1} + ... + \beta_{2K}x_{iK} = \beta_2'\mathbf{x}_i$$

...

$$\log\left[\frac{P[Y_i = j|x_{i1}, ..., x_{iK}]}{P[Y_i = J|x_{i1}, ..., x_{iK}]}\right] = \beta_{j0} + \beta_{j1}x_{i1} + ... + \beta_{jK}x_{iK} = \beta_j'\mathbf{x}_i$$

...

$$\log\left[\frac{P[Y_i = J-1|x_{i1}, ..., x_{iK}]}{P[Y_i = J|x_{i1}, ..., x_{iK}]}\right] = \beta_{J0} + \beta_{J1}x_{i1} + ... + \beta_{JK}x_{iK} = \beta_J'\mathbf{x}_i$$

- Note, each one of these logits can have a different set of parameters β_j.
- Basically, we can think of the j^{th} logit

$$\log\left[\frac{P[Y_i = j|x_{i1}, ..., x_{iK}]}{P[Y_i = J|x_{i1}, ..., x_{iK}]}\right] = \beta_j' \mathbf{x}_i,$$

  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

$$p_{ij} = P[Y_i = j | x_{i1}, ..., x_{iK}], \qquad j = 1, ..., J,$$

  in terms of the parameters and covariates.

- Recall, when $J = 2$, we write

$$p_{i1} = \frac{\exp[\beta' \mathbf{x}_i]}{1 + \exp[\beta' \mathbf{x}_i]}$$

  and

$$p_{i2} = \frac{1}{1 + \exp[\beta' \mathbf{x}_i]}$$

- We need to generalize this probability formulation when $J > 2$

- For now, consider the following definitions of p_{ij},

$$p_{ij} = \frac{\exp[\beta_j' \mathbf{x}_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}$$

when $j < J$,

and

$$p_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}$$

when $j = J$

- We know, $\sum_{j \in J} p_{ij} = 1$

- Using the proposed definitions

$$\begin{aligned}
\sum_{j=1}^{J} p_{ij} &= \sum_{j=1}^{J-1} p_{ij} + p_{iJ} \\
&= \sum_{j=1}^{J-1} \left(\frac{\exp[\beta_j' \mathbf{x}_i]}{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]} \right) + \frac{1}{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]} \\
&= \frac{\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]} + \frac{1}{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]} \\
&= \frac{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}{1+\sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]} \\
&= 1
\end{aligned}$$

- So, our proposed definitions are consistent with a proper probability distribution
- Now, we shall derive the probabilities.

Proof

- Now, consider

$$\log\left[\frac{p_{ij}}{p_{iJ}}\right] = \beta_j' \mathbf{x}_i$$

  exponentiating both sides, we get

$$\frac{p_{ij}}{p_{iJ}} = \exp[\beta_j' \mathbf{x}_i],$$

  which is the odds for category j versus category J.

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

$$p_{ij} = p_{iJ} \exp[\beta_j' \mathbf{x}_i],$$

- Now, suppose we sum both sides over $j = 1, ..., J-1$, we get

$$\sum_{j=1}^{J-1} p_{ij} = p_{iJ} \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i],$$

- 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

$$1 - p_{iJ} = p_{iJ} \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i],$$

- Then,

$$\begin{aligned} 1 &= p_{iJ} + p_{iJ} \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i] \\ &= p_{iJ}(1 + \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]), \end{aligned}$$

- Or, finally

$$p_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}$$

and, since

$$p_{ij} = p_{iJ} \exp[\beta_j' \mathbf{x}_i],$$

substituting in p_{iJ}, we obtain

$$p_{ij} = \frac{\exp[\beta_j' \mathbf{x}_i]}{1 + \sum_{j=1}^{J-1} \exp[\beta_j' \mathbf{x}_i]}$$

Interpretation

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

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

- We want to know the interpretation of the β_{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

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

  and for subject i' is

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

- 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, β_{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, β_{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} :

$$\begin{aligned}\log\left[\frac{p_{i'j'}/p_{i'j}}{p_{ij'}/p_{ij}}\right] &= \log\left[\frac{p_{i'j'}/p_{i'J}}{p_{ij'}/p_{iJ}}\right] - \log\left[\frac{p_{i'j}/p_{i'J}}{p_{ij}/p_{iJ}}\right] \\ &= [\beta_{j'k} - \beta_{jk}]\end{aligned}$$

- 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:

```
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
                                referecne
 freq count;
run;
```

Selected Results

Response Profile

Ordered Value	mi	Total Frequency
1	1	23
2	2	270
3	3	21778

Logits modeled use mi=3 as the reference category.

Analysis of Maximum Likelihood Estimates

Parameter	mi	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	-7.6896	0.4472	295.6688	<.0001
Intercept	2	1	-4.7044	0.1010	2171.3642	<.0001
x	1	1	1.2885	0.5056	6.4947	0.0108
x	2	1	0.5546	0.1270	19.0675	<.0001

- In terms of the model,

$$\begin{aligned} \log\left[\frac{p_{i1}}{p_{i3}}\right] &= \beta_{10} + \beta_{11}x_{i1} \\ &= -7.6896 + 1.2885x_{i1} \end{aligned}$$

and

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

where $x_{i1} = 1$ if treated with placebo, 0 else.

Agresti Table 2.1 - Page 37			
	Myocardial Infarction		
	Fatal Attack	Nonfatal Attack	No Attack
Placebo	18	171	10845
Aspirin	5	99	10933

- Recall, we previously calculated the last category ORs to be
  1. $\alpha_{11} = (18 * 10933)/(5 * 10933) = 3.62$
  2. $\alpha_{12} = (171 * 10933)/(99 * 10845) = 1.74$

```
                Odds Ratio Estimates
                     Point          95% Wald
Effect     mi    Estimate      Confidence Limits

x          1        3.627       1.347       9.771
x          2        1.741       1.358       2.233
```

- 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

$$\begin{aligned} log(OR) &= \beta_{11} - \beta_{12} \\ &= 1.2885 - 0.5546 \\ &= 0.7345 \\ OR &= exp(0.7345) \\ &= 2.08 \end{aligned}$$

Maximum Likelihood Using the Multinomial

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

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

  $j = 1, ..., J$

- 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 ≤ 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

Model 1: Intercept Only Model

```
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

Ordered Value	food	Total Frequency
1	1	94
2	2	61
3	3	19
4	4	13
5	5	32

Logits modeled use food=1 as the reference category.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	116.7611	60	1.9460	<.0001
Pearson	106.4922	60	1.7749	0.0002

Number of unique profiles: 16

Analysis of Maximum Likelihood Estimates

Parameter	food	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	2	1	-0.4324	0.1644	6.9173	0.0085
Intercept	3	1	-1.5989	0.2515	40.4037	<.0001
Intercept	4	1	-1.9783	0.2959	44.6984	<.0001
Intercept	5	1	-1.0776	0.2047	27.7197	<.0001

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

```
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

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	114.6571	56	2.0474	<.0001
Pearson	101.2480	56	1.8080	0.0002

- 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

```
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

Criterion	Value	DF	Value/DF	Pr > ChiSq
Model with Lake Size and Gender				
Deviance	50.2637	40	1.2566	0.1282
Model lake and Size				
Deviance	52.4785	44	1.1927	0.1784

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

```
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

```
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;
```

Obs	lake	size	food	count
1	1	0	1	7
2	1	0	2	0
3	1	0	3	1
4	1	0	4	3
5	1	0	5	5
6	1	1	1	23
7	1	1	2	4
8	1	1	3	2
9	1	1	4	2
10	1	1	5	8

...

- 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

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	17.0798	12	1.4233	0.1466
Pearson	15.0429	12	1.2536	0.2391

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

Criterion	Value	DF	Value/DF	Pr > ChiSq
Deviance	17.0798	12	1.4233	0.1466
Pearson	15.0429	12	1.2536	0.2391

Number of unique profiles: 8

Now, we will examine the parameter estimates.

Parameter Estimates

Parameter		food	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		2	1	-1.5490	0.4249	13.2890	0.0003
Intercept		3	1	-3.3139	1.0528	9.9081	0.0016
Intercept		4	1	-2.0931	0.6622	9.9894	0.0016
Intercept		5	1	-1.9043	0.5258	13.1150	0.0003
lake	1	2	1	-1.6583	0.6129	7.3216	0.0068
lake	1	3	1	1.2422	1.1852	1.0985	0.2946
lake	1	4	1	0.6951	0.7813	0.7916	0.3736
lake	1	5	1	0.8262	0.5575	2.1959	0.1384
lake	2	2	1	0.9372	0.4719	3.9443	0.0470
lake	2	3	1	2.4583	1.1179	4.8360	0.0279
lake	2	4	1	-0.6532	1.2021	0.2953	0.5869
lake	2	5	1	0.00565	0.7766	0.0001	0.9942
lake	3	2	1	1.1220	0.4905	5.2321	0.0222
lake	3	3	1	2.9347	1.1161	6.9131	0.0086
lake	3	4	1	1.0878	0.8417	1.6703	0.1962
lake	3	5	1	1.5164	0.6214	5.9541	0.0147
size		2	1	1.4582	0.3959	13.5634	0.0002
size		3	1	-0.3513	0.5800	0.3668	0.5448
size		4	1	-0.6307	0.6425	0.9635	0.3263
size		5	1	0.3316	0.4483	0.5471	0.4595

Thus, the estimated model for estimating the log(odds) of an alligator eating invertebrate animals instead of fish would be

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

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 (≤ 2.3) are $exp(1.4582) = 4.3$ times the estimated odds for large alligators.

Odds Ratio Summary

Odds Ratio Estimates

Effect	food	Point Estimate	95% Wald Confidence Limits	
lake 1 vs 4	2	0.190	0.057	0.633
lake 1 vs 4	3	3.463	0.339	35.343
lake 1 vs 4	4	2.004	0.433	9.266
lake 1 vs 4	5	2.285	0.766	6.814
lake 2 vs 4	2	2.553	1.012	6.437
lake 2 vs 4	3	11.685	1.306	104.508
lake 2 vs 4	4	0.520	0.049	5.490
lake 2 vs 4	5	1.006	0.219	4.608
lake 3 vs 4	2	3.071	1.174	8.032
lake 3 vs 4	3	18.815	2.111	167.717
lake 3 vs 4	4	2.968	0.570	15.447
lake 3 vs 4	5	4.556	1.348	15.400
size	2	4.298	1.978	9.339
size	3	0.704	0.226	2.194
size	4	0.532	0.151	1.875
size	5	1.393	0.579	3.354

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
- Where as 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 you 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.