BIOS 6612
Lecture 9

Logistic Regression VIII
Logistic Regression for Categorical Outcomes

Review (Lecture 8)/ Current (Lecture 9)/ Preview (Lecture 10)

- Lecture 8: Logistic Regression VII
  - Multiple logistic regression
  - Grouped data in multiple logistic regression
  - Deviance
  - Models: null, model of interest, saturated

- Lecture 9: Logistic Regression VIII
  - Baseline category logit
  - Continuation ratios
  - Adjacent category logits
  - Cumulative logits
    - Proportional odds cumulative logits
  - Examples: Penn State and Agresti
    https://onlinecourses.science.psu.edu/stat504/node/171

- Lecture 10: General Linear Models

Baseline Category Logit Models: Notation

- Usually, used for a categorical outcome with NO clear ordering for the outcome Y

- Use logit models to simultaneously compare each category to a common baseline category

- Predominantly used for categorical (nominal) responses where
  - No clear order for the outcome

  - For example, outcome= alligators eating invertebrates, reptiles, fish, birds, and other (i.e. rocks)
    - i.e. J=5 below

- For subject i, p covariates, and outcome j where j=1,..,J

$$\pi_{ij} = Pr(Y_i = j|x_i)$$

- For j=1,…,J-1

$$log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \beta_{0j} + \beta_{1j}x_{1i} + \cdots + \beta_{pj}x_{pi}$$

- Interpretation: β_{1j} is the increase in the log-odds of being in category j vs J resulting from a one unit increase in the covariate x_{1i}, holding the other covariates constant

Baseline Category Logit Models: Notation

- For $j \neq k$ and both less than J

$$log\left(\frac{\pi_{ij}}{\pi_{ik}}\right) = log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) - log\left(\frac{\pi_{ik}}{\pi_{iJ}}\right)$$
$$= \left(\beta_{0j} + \beta_{1j}x_{1i} + \cdots + \beta_{pj}x_{pi}\right) - \left(\beta_{0k} + \beta_{1k}x_{1i} + \cdots + \beta_{pk}x_{pi}\right)$$

- Solving for J

$$\pi_{iJ} = \frac{1}{1 + \sum_{h=1}^{J-1} exp\left(\beta_{0h} + \beta_{1h}x_{1i} + \cdots + \beta_{ph}x_{pi}\right)}$$

- Then for j=1,…,J-1

$$\pi_{ij} = \frac{exp\left(\beta_{0j} + \beta_{1j}x_{1i} + \cdots + \beta_{pj}x_{pi}\right)}{1 + \sum_{h=1}^{J-1} exp\left(\beta_{0h} + \beta_{1h}x_{1i} + \cdots + \beta_{ph}x_{pi}\right)}$$

- The difference of the logits for individual (i1) and (i2)

$$log\left(\frac{\pi_{(\mathrm{i1})j}}{\pi_{(\mathrm{i2})j}}\right) = log\left(\frac{\pi_{(\mathrm{i1})j}}{\pi_J(\mathbf{X})}\right) - log\left(\frac{\pi_{(\mathrm{i2})j}}{\pi_J(\mathbf{X})}\right)$$
$$= \left(\beta_{0j} + \beta_{1j}x_{1(\mathrm{i1})} + \cdots + \beta_{pj}x_{p(\mathrm{i1})}\right) - \left(\beta_{0j} + \beta_{1j}x_{1(\mathrm{i2})} + \cdots + \beta_{pj(\mathrm{i2})}\right)$$

Baseline Category Logit Models: Notes

- Any of the categories can be chosen to be the baseline
  - The model will fit equally well
    - Same likelihood and the same fitted values
  - Only the values and interpretation of the coefficients will change
    - Similar idea as seen in BIOS 6611 to changing the reference group for a categorical covariate in a linear regression
  - A 'natural' baseline should be chosen
    - If the response is ordinal, usually the highest or the lowest category in ordinal scale is chosen
- Models fit for all categories
  - Naïve approach: fit a separate model to only those observations falling in response category j and J
    - These 2 strategies are NOT identical
  - The separate fitting strategy is less efficient
    - Can be shown that efficiency loss in the separate pairwise approach is not too great if the most prevalent response category is selected for the baseline

Baseline Category Logit Models: Fitting the Model

- In SAS
  - PROC LOGISTIC with the link=GLOGIT
  - Can also use PROC CATMOD and GENMOD

  - Agresti appendix A.1 SAS code for examples

- In R
  - vglm() from the VGAM package
  - multinom() from the nnet package
  - mlogit() from the globaltest package from BIOCONDUCTOR

  - R (and S-PLUS) Manual to Accompany Agresti's Categorical Data Analysis (2002) 2nd edition. Laura A. Thompson, 2009

- **Note: poor model fit could be due to overdispersion**

  - Overdispersion occurs when the actual variance covariance matrix for Y exceeds that specified by the assumed distributed

    - Poisson distribution vs Negative Binomial

Baseline Category Logit Models: Example

- Agresti alligator example
  - A study of the primary food choices of alligators in four Florida lakes
  - Researchers classified the stomach contents of 219 captured alligators into five categories:

Fish (the most common primary food choice)

Invertebrate (snails, insects, crayfish, etc.)

Reptile (turtles, alligators)

Bird

Other (amphibians, plants, household pets, stones, and other debris).

			Primary Food Choice				
Lake	*Sex*	*Size*	*Fish*	*Inv.*	*Rept.*	*Bird*	*Other*
Hancock	M	small	7	1	0	0	5
		large	4	0	0	1	2
	F	small	16	3	2	2	3
		large	3	0	1	2	3
Oklawaha	M	small	2	2	0	0	1
		large	13	7	6	0	0
	F	small	3	9	1	0	2
		large	0	1	0	1	0
Trafford	M	small	3	7	1	0	1
		large	8	6	6	3	5
	F	small	2	4	1	1	4
		large	0	1	0	0	0
George	M	small	13	10	0	2	2
		large	9	0	0	1	2
	F	small	3	9	1	0	1
		large	8	1	0	0	1

Baseline Category Logit Models: Example

- Baseline category: fish
  - The primary food choice of alligators appears to be fish
  - 4 logit equations will then describe the log-odds that alligators select other primary food types instead of fish

- For the 5 categories by order of most common to least
  - fish (baseline)
  - invertebrates
  - reptiles
  - birds
  - other

- Covariates: 4 lakes (reference=Hancock), indicator variables for size (reference=small) and gender (reference= small)

  - The baseline category logit for j=bird, invertebrates, other, reptile

$$log\left(\frac{\pi_j}{\pi_{i(fish)}}\right) = \beta_{0j} + \beta_{Oj} Oklawaha_i + \beta_{Tj} Trafford_i + \beta_{Gj} George_i + \beta_{Lj} Large_i + \beta_{Fj} Female_i$$

Baseline Category Logit Models: SAS Code & Output

```
proc logist data=gator;
  freq count;
  class lake size sex / order=data param=ref ref=first;
  model food(ref='fish') = lake size sex / link=glogit
     aggregate scale=none;
  run;
```

Analysis of Maximum Likelihood Estimates

Parameter		food	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		bird	1	-2.4633	0.7739	10.1310	0.0015
Intercept		invert	1	-2.0744	0.6116	11.5025	0.0007
Intercept		other	1	-0.9167	0.4782	3.6755	0.0552
Intercept		reptile	1	-2.9141	0.8856	10.8275	0.0010
lake	Oklawaha	bird	1	-1.1256	1.1924	0.8912	0.3452
lake	Oklawaha	invert	1	2.6937	0.6692	16.2000	<.0001
lake	Oklawaha	other	1	-0.7405	0.7422	0.9956	0.3184
lake	Oklawaha	reptile	1	1.4008	0.8105	2.9872	0.0839
lake	Trafford	bird	1	0.6617	0.8461	0.6117	0.4341
lake	Trafford	invert	1	2.9363	0.6874	18.2469	<.0001
lake	Trafford	other	1	0.7912	0.5879	1.8109	0.1784
lake	Trafford	reptile	1	1.9316	0.8253	5.4775	0.0193
lake	George	bird	1	-0.5753	0.7952	0.5233	0.4694
lake	George	invert	1	1.7805	0.6232	8.1623	0.0043
lake	George	other	1	-0.7666	0.5686	1.8179	0.1776
lake	George	reptile	1	-1.1287	1.1925	0.8959	0.3439
size	large	bird	1	0.7302	0.6523	1.2533	0.2629
size	large	invert	1	-1.3363	0.4112	10.5606	0.0012
size	large	other	1	-0.2906	0.4599	0.3992	0.5275
size	large	reptile	1	0.5570	0.6466	0.7421	0.3890
sex	female	bird	1	0.6064	0.6888	0.7750	0.3787
sex	female	invert	1	0.4630	0.3955	1.3701	0.2418
sex	female	other	1	0.2526	0.4663	0.2933	0.5881
sex	female	reptile	1	0.6275	0.6852	0.8387	0.3598

Baseline Category Logit Models: SAS Output

Odds Ratio Estimates

Effect	food	Point Estimate	95% Wald Confidence Limits	
lake Oklawaha vs Hancock	bird	0.324	0.031	3.358
lake Oklawaha vs Hancock	invert	14.786	3.983	54.893
lake Oklawaha vs Hancock	other	0.477	0.111	2.042
lake Oklawaha vs Hancock	reptile	4.058	0.829	19.872
lake Trafford vs Hancock	bird	1.938	0.369	10.176
lake Trafford vs Hancock	invert	18.846	4.899	72.500
lake Trafford vs Hancock	other	2.206	0.697	6.983
lake Trafford vs Hancock	reptile	6.900	1.369	34.784
lake George vs Hancock	bird	0.563	0.118	2.673
lake George vs Hancock	invert	5.933	1.749	20.125
lake George vs Hancock	other	0.465	0.152	1.416
lake George vs Hancock	reptile	0.323	0.031	3.349
size large vs small	bird	2.076	0.578	7.454
size large vs small	invert	0.263	0.117	0.588
size large vs small	other	0.748	0.304	1.842
size large vs small	reptile	1.745	0.492	6.198
sex female vs male	bird	1.834	0.475	7.075
sex female vs male	invert	1.589	0.732	3.449
sex female vs male	other	1.287	0.516	3.211
sex female vs male	reptile	1.873	0.489	7.175

Baseline Category Logit Models: Partial Interpretation

$$log\frac{\hat{\pi}_{Blsg}}{\hat{\pi}_{Flsg}} = -2.4633 + -1.1256 Oklawaha + 0.6617 Trafford - 0.5753 George + 0.7302 large + 0.6064 female$$

- The intercepts give the estimated log-odds for the reference group lake = Hancock, size = small, sex = male.
  - For example, the estimated log-odds of birds versus fish in this group is −2.4633

- The lake effect is characterized by three dummy coefficients in each of the four logit equations.

- The estimated coefficient for the Lake Oklawaha dummy in the bird-versus-fish equation is −1.1256 with st. error 1.1924.

  - This means that alligators in Lake Oklawaha are less likely to choose birds over fish than their colleagues in Lake Hancock are.

Continuation Ratios

- Assume some sort of ordering for the outcome Y

- The continuation ratio logit models the conditional probabilities

$$Pr(Y_i = j|Y_i \geq j, x_i)$$
or
$$Pr(Y_i = j|Y_i > j, x_i)$$
or
$$Pr(Y_i = j|Y_i \leq j, x_i)$$

  - For j=1,..,J-1

$$Pr(Y_i = j|Y_i \geq j, x_i) = \frac{\pi_{ij}}{\pi_{ij} + \cdots + \pi_{iJ}}$$
or
$$Pr(Y_i = j|Y_i > j, x_i) = \frac{\pi_{ij}}{\pi_{i(j+1)} + \cdots + \pi_{iJ}}$$
or
$$Pr(Y_i = j|Y_i \leq j) = \frac{\pi_{ij}}{\pi_{i1} + \cdots + \pi_{ij}}$$

- The continuation-ratio logit model form is useful when a sequential mechanism determines the response outcome
  - Such as survival through various age periods

Adjacent Category Logits

- Assume some sort of ordering for the outcome Y
  - Adjacent category logit model incorporates the ordering of the response in to the analysis

- Consider comparing categories j and j+1 conditional on the response being in one of these 2 categories

$$Pr(Y_i = j|Y_i \in \{j, j+1\}, x_i) = \frac{Pr(Y_i = j|x_i)}{Pr(Y_i = j|x_i) + Pr(Y_i = j+1|x_i)}$$

- The adjacent categories logit model has a separate set of regression coefficients for each set of adjacent categories

- For j=1,…,J-1

$$log\left(\frac{\pi_{ij}}{\pi_{i(j+1)}}\right) = \beta_{0j} + \beta_{1j}x_{1i} + \cdots + \beta_{pj}x_{pi}$$

- Similar idea to a baseline-category logit model
  - But here the denominator changes from one category to the next
  - Denominator is always the reference group for the baseline category logit

Cumulative Logit Model

- Assume an ordering for the outcome Y
  - Most popular model for ordinal data

- Let

$$\gamma_1(x_i) = Pr(Y_i \leq 1|x_i) = \pi_{i1}$$
$$\gamma_2(x_i) = Pr(Y_i \leq 2|x_i) = \pi_{i1} + \pi_{i2}$$

$$\gamma_j(x_i) = Pr(Y_i \leq j|x_i) = \pi_{i1} + \cdots + \pi_{ij}$$

$$\gamma_{J-1}(x_i) = Pr(Y_i \leq J-1|x_i) = \pi_{i1} + .. + \pi_{i(J-1)}$$

- Define the cumulative logit

$$log\left(\frac{\Pr(Y_i \leq j|x_i)}{\Pr(Y_i > j|x_i)}\right) = log\left(\frac{\Pr(Y_i \leq j|x_i)}{1-\Pr(Y_i \leq j|x_i)}\right) = log\left(\frac{\gamma_j(x_i)}{1-\gamma_j(x_i)}\right)$$
$$= log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i(j+1)} + \cdots + \pi_{iJ}}\right)$$

Cumulative Logit Model

- Then for the cumulative logit model

$$log\left(\frac{\Pr\left(Y_i \leq 1|x_i\right)}{\Pr\left(Y_i > 1|x_i\right)}\right) = log\left(\frac{\pi_{i1}}{\pi_{i2} + \cdots + \pi_{iJ}}\right) = \beta_{01} + \beta_{11}x_{1i} + \cdots + \beta_{p1}x_{pi}$$

$$log\left(\frac{\Pr\left(Y_i \leq 2|x_i\right)}{\Pr\left(Y_i > 2|x_i\right)}\right) = log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3} + \cdots + \pi_{iJ}}\right) = \beta_{02} + \beta_{12}x_{1i} + \cdots + \beta_{p2}x_{pi}$$

$$log\left(\frac{\Pr\left(Y_i \leq j|x_i\right)}{\Pr\left(Y_i > j|x_i\right)}\right) log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i(j+1)} + \cdots + \pi_{iJ}}\right) = \beta_{0j} + \beta_{1j}x_{1i} + \cdots + \beta_{pj}x_{pi}$$

$$log\left(\frac{\Pr\left(Y_i \leq J-1|x_i\right)}{\Pr\left(Y_i > J-1|x_i\right)}\right) = log\left(\frac{\pi_{i1} + \cdots + \pi_{i(J-1)}}{\pi_{iJ}}\right) = \beta_{0(J-1)} + \cdots + \beta_{p(J-1)}x_{pi}$$

- Unlike the adjacent-category logit model, this is not a linear reparameterization of the baseline-category model
  - The cumulative logits are not simple differences between the baseline-category logits
  - Therefore, the above model will not give a fit equivalent to that of the baseline-category model

Proportional Odds Cumulative Logit Model

- Then for the proportional odds cumulative logit model simplifies the model

$$log\left(\frac{\Pr(Y_i \leq 1|x_i)}{\Pr(Y_i > 1|x_i)}\right) = log\left(\frac{\pi_{i1}}{\pi_{i2} + \cdots + \pi_{iJ}}\right) = \alpha_1 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

$$log\left(\frac{\Pr(Y_i \leq 2|x_i)}{\Pr(Y_i > 2|x_i)}\right) = log\left(\frac{\pi_{i1} + \pi_{i2}}{\pi_{i3} + \cdots + \pi_{iJ}}\right) = \alpha_2 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

$$log\left(\frac{\Pr(Y_i \leq j|x_i)}{\Pr(Y_i > j|x_i)}\right) log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i(j+1)} + \cdots + \pi_{iJ}}\right) = \alpha_j + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

$$log\left(\frac{\Pr(Y_i \leq J-1|x_i)}{\Pr(Y_i > J-1|x_i)}\right) = log\left(\frac{\pi_{i1} + \cdots + \pi_{i(J-1)}}{\pi_{iJ}}\right) = \alpha_{J-1} + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

- (r − 1) intercepts plus p slopes
  - total of r + p − 1 parameters to be estimated

- The intercepts can differ, but that slope for each variable stays the same across different equations!
  - The proportional-odds condition forces the lines corresponding to each cumulative logit to be parallel

Proportional Odds Cumulative Logit Model: Interpretation

$$log\left(\frac{\Pr(Y_i \le j|x_i)}{\Pr(Y_i > j|x_i)}\right) log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i(j+1)} + \cdots + \pi_{iJ}}\right) = \alpha_j + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

- In this model, intercept α_j is the log-odds of falling into or below category j when all x's=0
- Just as in ordinary logistic regression, β_k has the interpretation of being the log-odds ratio for a one unit increase in x_{ki} while keeping the other covariates constant
- **Notes:**
  - Proportional odds cumulative logit models are especially useful for testing whether covariates have an overall relationship with the response since it gives a 1 degree-of-freedom test for each covariate
  - Also useful if some response categories have a small number of cases, so estimating a separate set of regression parameters for that category might not be possible

Proportional Odds Cumulative Logit Model: Cheese Tasting Example

(McCullagh and Nelder, 1989). In this example, subjects were randomly assigned to taste one of four different cheeses. Response categories are 1 = strong dislike to 9 = excellent taste.

Cheese	Response category								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
A	0	0	1	7	8	8	19	8	1
B	6	9	12	11	7	6	1	0	0
C	1	1	6	8	23	7	5	1	0
D	0	0	0	1	3	7	14	16	11

```
proc logist data=cheese; freq count;
class cheese / order=data param=ref ref=first;
model response (order=data descending) = cheese / link=logit
    aggregate=(cheese) scale=none;
run;
```

- In PROC LOGISTIC, the order=data option tells SAS to arrange the response categories from lowest to highest in the order that they arise in the dataset
    - The option descending tells SAS to reverse the ordering of the categories, so that 9 becomes the lowest and 1 becomes the highest, and a positive β indicates that a higher value of X leads to greater liking

- Other procedures such as PROC GENMOD can be also used

Proportional Odds Cumulative Logit Model: Cheese Tasting Example

$$\log\frac{P(Y \le 1)}{P(Y > 1)} = \alpha_1 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\log\frac{P(Y \le 2)}{P(Y > 2)} = \alpha_2 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

$$\log\frac{P(Y \le 8)}{P(Y > 8)} = \alpha_8 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

where X1 = 1 for cheese B and zero otherwise
X2 = 1 for cheese C and zero otherwise
X3 = 1 for cheese D and zero otherwise

In this case, a positive coefficient β means that increasing the value of X tends to lower the response categories (i.e. produce greater dislike)

Proportional Odds Cumulative Logit Model: Cheese Tasting Example

The estimated slope for comparing cheese B to the reference cheese A is −3.3517

This indicates that cheese B does not taste as good as cheese A

Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	9	1	-3.1058	0.4044	58.9727	<.0001
Intercept	8	1	-1.5459	0.3042	25.8287	<.0001
Intercept	7	1	-0.0443	0.2598	0.0291	0.8646
Intercept	6	1	0.9077	0.2748	10.9125	0.0010
Intercept	5	1	2.2440	0.3262	47.3307	<.0001
Intercept	4	1	3.3126	0.3697	80.2992	<.0001
Intercept	3	1	4.4121	0.4247	107.9168	<.0001
Intercept	2	1	5.4673	0.5202	110.4514	<.0001
cheese	B	1	-3.3517	0.4235	62.6335	<.0001
cheese	C	1	-1.7098	0.3731	21.0072	<.0001
cheese	D	1	1.6128	0.3778	18.2265	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
cheese B vs A	0.035	0.015	0.080
cheese C vs A	0.181	0.087	0.376
cheese D vs A	5.017	2.393	10.520