Lecture 21: Logit Models for Multinomial Responses Continued

Dipankar Bandyopadhyay, Ph.D.

BMTRY 711: Analysis of Categorical Data Spring 2011

Division of Biostatistics and Epidemiology

Medical University of South Carolina

Ordinal Regression Models

- In the previous lecture, we examined a multinomial logistic model defined for a nominal, multicategory response

- For each of the $J - 1$ levels of Y, we considered a log-odds model referencing level J

- This **baseline category** model estimated $p \times (J - 1)$ parameters to sufficiently explain all associations in the data

- In this lecture, we are going to consider simplifications of this model that are possible when Y is ordinal

- In formulating a regression model, we would like to take this ordering into account.

- We will focus on the most common model, the 'proportional odds model'

- Ordinal outcomes are common in
  1. Social sciences
  2. Market research
  3. Opinion polls
- Often a result of discretization of a **latent variable**
- A latent variable is a psychometric variable that is unobservable but is measured, typically, by a "scale"
- For example, the Hamilton Depression Rating Scale measures depression on a scale ranging from approximately 0 to 30 (depending on number of items used)
- Scores less than 7 indicate remission, 7 - 12 moderate depression

- The purpose of the regression analysis is to explore the association of a group of covariates on the outcome

- When the outcome is polychotomous, grouping (or dichotomizing) the outcome may not be possible

- However, if the outcome is ordinal, a first line approach to the analysis may be to group the outcome into binary categories

- Such as, Depressed v. Not Depressed; "good" v. "poor" rating; etc.

- However, just in the $(I \times J)$ contingency tables, collapsing the outcome resulted in a loss of power

Example–Arthritis Clinical Trial

- This is the same arthritis clinical trial comparing the drug auranofin and placebo therapy for the treatment of rheumatoid arthritis (Bombardier, et al., 1986).

- The response of interest is the self-assessment of arthritis, before, I said it was classified as (0) poor or (1) good. Actually, I had dichotomized the data. The self-assessment was actually a 5-level ordinal variable: (1) very good, (2) good, (3) fair, (4) poor, (5) very poor,
  (I dichotomized ≤ 3 versus > 3.)

- Individuals were randomized into one of the two treatment groups after baseline self-assessment of arthritis (with the same 5 levels as the response).

- The dataset contains 293 patients who were observed at both baseline and 13 weeks. The data from few cases are shown below:

Subset of cases from the arthritis clinical trial

| | | | | | Self assessment[b] |
CASE	SEX	AGE	TREATMENT[a]	BASELINE	13 WK.
1	M	54	A	4	1
2	M	64	P	4	5
3	M	48	A	3	3
4	F	41	A	3	2
5	M	55	P	3	2
6	M	64	A	2	2
7	M	64	P	3	4
8	F	55	P	1	2
9	M	39	P	2	5
10	F	60	A	4	3

[a] A = Auranofin, P = Placebo

[b] 1=very good, 2=good, 3=fair, 4=poor, 5=very poor.

- We are again interested in a pretest-posttest analysis, in which we relate the individual's discrete response

$$Y_i = \begin{cases} 1 \text{ if very good at 13 weeks} \\ 2 \text{ if good at 13 weeks} \\ 3 \text{ if fair at 13 weeks} \\ 4 \text{ if poor at 13 weeks} \\ 5 \text{ if very poor at 13 weeks} \end{cases}.$$

- 1. BASELINE self-assessment:

$$X_i = \begin{cases} 1 \text{ if very good at baseline} \\ 2 \text{ if good at baseline} \\ 3 \text{ if fair at baseline} \\ 4 \text{ if poor at baseline} \\ 5 \text{ if very poor at baseline} \end{cases}.$$

- 2. AGE IN YEARS,
- 3. GENDER (1 if male, 0 if female)
- 4. TREATMENT (1 if auranofin, 0 if placebo)

Example–Arthritis Clinical Trial

- The outcome is

$$
Y_i = \begin{cases}
1 \text{ if very good at 13 weeks} \\
2 \text{ if good at 13 weeks} \\
3 \text{ if fair at 13 weeks} \\
4 \text{ if poor at 13 weeks} \\
5 \text{ if very poor at 13 weeks}
\end{cases}.
$$

- Suppose we dichotomize the outcome at 1 vs > 1 :

$$
U_{i1} = \begin{cases}
1 \text{ if very good at 13 weeks} \\
0 \text{ if good, fair, poor, very poor at 13 weeks}
\end{cases}.
$$

  and let

$$
F_{i1} = P(U_{i1} = 1|\mathbf{x}_i) = \text{prob very good}
$$

- Since U_{i1} is dichotomous, we could formulate a logistic regression model for it:

$$
\text{logit}(F_{i1}) = \log\left(\frac{F_{i1}}{1 - F_{i1}}\right) = \alpha_1 + \boldsymbol{\beta}'\mathbf{x}_i.
$$

- Next, we could dichotomize the outcome at ≤ 2 vs > 2 :

$$U_{i2} = \left\{ \begin{array}{l} \text{1 if very good or good at 13 weeks} \\ \text{0 if fair, poor, very poor at 13 weeks} \end{array} \right. .$$

and let

$$F_{i2} = P(U_{i2} = 1 | \mathbf{x}_i) = \text{prob very good or good}$$

- Since U_{i2} is dichotomous, we could formulate a logistic regression model for it:

$$\text{logit}(F_{i2}) = \alpha_2 + \boldsymbol{\beta}' \mathbf{x}_i.$$

Note, here, we have assumed the intercepts for $\text{logit}(F_{i1})$ and $\text{logit}(F_{i2})$ are different, but we have assumed the $\boldsymbol{\beta}$'s are the same.

- Going up the ordinal scale, we can form two more dichotomous variables:

$$U_{i3} = \begin{cases} 1 \text{ if very good,good, or fair, at 13 weeks} \\ 0 \text{ if poor, very poor at 13 weeks} \end{cases}.$$

$$U_{i4} = \begin{cases} 1 \text{ if very good, good, fair, or poor at 13 weeks} \\ 0 \text{ if very poor at 13 weeks} \end{cases}.$$

with

$$F_{i3} = P(U_{i3} = 1|\mathbf{x}_i) \qquad \text{and} \qquad \text{logit}(F_{i3}) = \alpha_3 + \boldsymbol{\beta}'\mathbf{x}_i$$

and

$$F_{i4} = P(U_{i4} = 1|\mathbf{x}_i) \qquad \text{and} \qquad \text{logit}(F_{i4}) = \alpha_4 + \boldsymbol{\beta}'\mathbf{x}_i.$$

- In general, the model is

$$
\begin{aligned}
\text{logit}(F_{ij}) &= log\left(\frac{F_{ij}}{1-F_{ij}}\right) \\
&= \alpha_j + \boldsymbol{\beta}'\mathbf{x}_i
\end{aligned}
$$

where $j = 1, \ldots, J - 1$ and $\boldsymbol{\beta}$ is a $p \times 1$ vector of covariates

- This is the cumulative logistic model:
  1. You dichotomize the ordinal variables going up (or down) the ordinal scale
  2. You form a logistic model for each dichotomous variable, in which the intercepts (say, α_j's are different, but the slopes ($\boldsymbol{\beta}$'s) are the same.

Cumulative probabilities

- In general,

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

where the multinomial probabilities are

$$p_{ij} = P[Y_{ij} = 1 | \mathbf{x}_i]$$

- We had defined the cumulative random variables U_{ij} :

$$U_{ij} = \left\{ \begin{array}{l} 1 \text{ if } Y_i \leq j \\ 0 \text{ if } Y_i > j \end{array} \right. .$$

- We also can define the cumulative probabilities as

$$F_{ij} = P[U_{ij} = 1 | \mathbf{x}_i] = P[Y_i \leq j | \mathbf{x}_i] = p_{i1} + ... + p_{ij}$$

- Note, we only need the first $(J - 1)$ 'cumulative probabilities' $(F_{i1}, ..., F_{i,J-1})$ since the last one always equals 1,

$$F_{iJ} = P[Y_i \leq J | \mathbf{x}_i] = p_{i1} + ... + p_{iJ} = 1$$

- The 'cumulative' logit is defined as:

$$\text{logit}(F_{ij}) = \log \left(\frac{F_{ij}}{1 - F_{ij}} \right)$$

- These cumulative logits are related to covariates in the following logistic regression model,

$$\text{logit}(F_{ij}) = \alpha_j + \mathbf{x}'_i\beta,$$

for $j = 1, ..., J - 1$

- This model also implies that the cumulative logits j and j', $\text{logit}(F_{ij})$ and $\text{logit}(F_{ij'})$, have the same slopes $\boldsymbol{\beta}$, but the intercepts α_j differ

- In other words, the coefficients $\boldsymbol{\beta}$ of the covariate vector $\mathbf{x}_i$ are the same for all cumulative probabilities, and does not depend on j.

- The ordering of the data is taken into account with this common $\boldsymbol{\beta}$ assumption.

- The proportional odds model can also be derived by discretizing an underlying continuous logistic random variable (and, of course, any continuous variable has an ordering).

Interpretation of β

- Suppose we have two covariate $\mathbf{x}_i = (x_{i1}, x_{i2})$, to give the model,

$$\text{logit}(F_{ij}) = \alpha_k + x_{i1}\beta_1 + x_{i2}\beta_2$$

- What is the interpretation of β_1?

- Just as in ordinary logistic regression, β_1 has the interpretation as the log-odds ratio for a cumulative probability for a one unit increase in x_{i1} while keeping the other covariates constant, i.e.,

$$\beta_1 = \log\left(\frac{F_{ij}(x_{i1} = c+1)/[1 - F_{ij}(x_{i1} = c+1)]}{F_{ij}(x_{i1} = c)/[1 - F_{ij}(x_{i1} = c]}\right),$$

which is often called the 'cumulative' log(OR):

- It is actually the log-odds ratio for

$$(Y_i \leq j) \quad \text{versus} \quad (Y_i > j)$$

for a one unit change in the covariate x_{i1}.

- Further, for two values of x_{i1}, say c_1 and c_2,

$$\beta_1(c_1 - c_2) = \log \left(\frac{F_{ij}(x_{i1} = c_1)/[1 - F_{ij}(x_{i1} = c_1)]}{F_{ij}(x_{i1} = c_2)/[1 - F_{ij}(x_{i1} = c_2]} \right),$$

- The cumulative log-odds ratio is proportional to the distance between the two values of the covariate x_{i1}, which is one reason this is called the 'proportional odds'.

- Since the log-odds ratio does not depend on the intercept α_j (as is the case in ordinary logistic regression), the log-odds ratios will be the same, for any cumulative probability:

$$
\begin{aligned}
\beta_1 &= \log\left(\frac{F_{ij}(x_{i1}=c+1)/[1-F_{ij}(x_{i1}=c+1)]}{F_{ij}(x_{i1}=c)/[1-F_{ij}(x_{i1}=c]}\right) \\
&= \log\left(\frac{F_{ij'}(x_{i1}=c+1)/[1-F_{ij'}(x_{i1}=c+1)]}{F_{ij'}(x_{i1}=c)/[1-F_{ij'}(x_{i1}=c]}\right)
\end{aligned}
$$

- Then, the odds ratio for

$$(Y_i \le j) \quad \text{versus} \quad (Y_i > j)$$

for a one unit increase in a covariate does not depend on which cumulative probability (j) you are looking at.

- This model says that if you have a discrete, ordinal random variable, and you want to dichotomize it (above and below a given j), and use ordinary logistic regression, your odds ratio will not change, regardless of where you dichotomize it. Only the intercept will be different.

- In the above example, suppose you are looking at the response versus treatment odds ratio, then, when comparing the new treatment versus placebo, the cumulative odds ratios are all equal:

$$OR(\text{very good vs.} < \text{very good}) \quad =$$
$$OR(\geq \text{good vs.} < \text{good}) \quad =$$
$$OR(\geq \text{fair vs.} < \text{fair}) \quad =$$
$$OR(\geq \text{poor vs. very poor})$$

- When we look at the output, we will see that, unlike the above polytomuous logit, we will get only one set of β's, although we will get $J - 1$ intercepts.

$$\text{logit}(F_{ij}) = \alpha_j + \mathbf{x}'_i \beta,$$

Non-proportional Odds

- The proportional odds model says that if you have a discrete, ordinal random variable, and you want to dichotomize it (above and below a given j), and use ordinary logistic regression, your odds ratio will not change, regardless of where you dichotomize it.

- On the other hand, we could have a 'non-proportional' odds model, in which the proportionality constant (log-odds ratio) depends on the response level j

$$\text{logit}(F_{ij}) = \alpha_k + \mathbf{x}'_i \beta_j$$

- Here, the log-odds ratio depends on j :

$$\beta_{1j}(c_1 - c_2) = \log\left(\frac{F_{ij}(x_{i1} = c_1)/[1 - F_{ij}(x_{i1} = c_1)]}{F_{ij}(x_{i1} = c_2)/[1 - F_{ij}(x_{i1} = c_2]}\right).$$

- Unfortunately, you can't fit this model easily in the computer.

Score Stat for Proportional Odds

- SAS gives the score test for all the $(K \times 1)$ vectors β_j's being equal,

$$H_0 : \beta_1 = \beta_2 = ... = \beta_{J-1} = \beta$$

Under the null, there is one $K \times 1$ vector β, and under the alternative, there are $(J-1)$, $K \times 1$ vectors $\beta_1, \beta_2, ..., \beta_{J-1}$, so the score statistic will have

$$
\begin{aligned}
df \quad &= \quad \text{\# parameters in full model - \# parameters in reduced model} \\
&= \quad (J-1)K - K = (J-2)K
\end{aligned}
$$

MLE's

- To write down the likelihood, note, we can write the original multinomial probabilities in terms of the cumulative probabilities via:

$$
\begin{aligned}
p_{ij} &= (p_{i1} + ... + p_{ij}) - (p_{i1} + ... + p_{i,j-1}) \\
&= F_{ij} - F_{i,j-1}
\end{aligned}
$$

- The likelihood is the product over the multinomial likelihoods (of sample size 1) for individual:

$$
L_i(\alpha, \boldsymbol{\beta}) = \prod_{j=1}^{J} [p_{ij}(\alpha, \beta)]^{y_{ij}},
$$

- The overall likelihood is

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

- Then, we obtain the MLE and use the inverse information to estimate its variance.

- Can obtain the MLE in SAS Proc Logistic.

- 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).

- You can also use Pearson's chi-square as a goodness-of-fit statistic.

Example–Arthritis Clinical Trial

- The outcome is

$$Y_i = \begin{cases} 1 \text{ if very good at 13 weeks} \\ 2 \text{ if good at 13 weeks} \\ 3 \text{ if fair at 13 weeks} \\ 4 \text{ if poor at 13 weeks} \\ 5 \text{ if very poor at 13 weeks} \end{cases}.$$

- There are 4 cumulative probabilities created by default in SAS Proc Logistic (going from lowest to highest):

$$F_{i1} = p_{i1} = \text{prob very good}$$

$$F_{i2} = p_{i1} + p_{i2} = \text{prob very good or good}$$

$$F_{i3} = p_{i1} + p_{i2} + p_{i3} = \text{prob very good, good, or fair}$$

$$F_{i4} = p_{i1} + p_{i2} + p_{i3} + p_{i4} = \text{prob very good, good, fair, or poor}$$

- The model is

$$\text{logit}(F_{ij}) = \alpha_j + \beta_1 x_i + \beta_{\text{SEX}}\text{SEX}_i +$$
$$\beta_{\text{AGE}}\text{AGE}_i + \beta_{\text{TRT}}\text{TRT}_i$$

where the covariates are age in years at baseline (AGE_i), sex $(\text{SEX}_i,$ 1=male, 0=female), treatment $(\text{TRT}_i,$ 1 = auranofin, 0 = placebo), and x_i is baseline response (treated as continuous, 1-5)

- The main question is still whether the treatment increases the odds of a more favorable response, after controlling for baseline response; secondary questions are whether the response differs by age and sex.

- If you use the 'descending' option in Proc Logistic, you get the 4 cumulative probabilities going from highest to lowest:

$$F_{i1} = p_{i5} = \text{prob very poor}$$

$$F_{i2} = p_{i5} + p_{i4} = \text{prob very poor or poor}$$

$$F_{i3} = p_{i1} + p_{i2} + p_{i3} = \text{prob very poor, poor, or fair}$$

$$F_{i4} = p_{i1} + p_{i2} + p_{i3} + p_{i4} = \text{prob very poor, poor, fair, or good}$$

SAS Proc Logistic

```
The following ascii is in the current directory,  and called art2.dat
  1       54      1       4       1
  0       41      0       3       2
  1       48      1       3       2
  1       40      0       3       2
  1       29      1       3       2

  .        .      .       .       .

  .        .      .       .       .

  .        .      .       .       .
  1       39      1       3       3
  0       35      1       3       3
  0       35      1       3       3
  0       65      0       3       3
  1       55      0       4       3
  0       42      1       5       4
  1       37      0       3       3
  1       52      0       3       3
  1       60      0       3       4
  1       63      1       4       4
```

```
/* SAS STATEMENTS */

 DATA ARTH;
  infile 'art2.dat';
 input SEX AGE TRT x y;
;

proc logistic;
 model y = SEX  AGE TRT  x;
run;
```

```
Data Set                        WORK.ARTH
Response Variable               y
Number of Response Levels       5
Model                           cumulative logit


    Response Profile
Ordered
  Value         Y      Count


     1          1        38
     2          2        93
     3          3       103
     4          4        49
     5          5        10
Probabilities modeled are cumulated over the lower Ordered Values.


Score Test for the Proportional Odds Assumption
Chi-Square = 12.8763 with 12 DF (p=0.3781)
```

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
INTERCP1	1	0.9850	0.6395	2.3727	0.1235
INTERCP2	1	2.9290	0.6531	20.1114	0.0001
INTERCP3	1	4.7706	0.6904	47.7450	0.0001
INTERCP4	1	6.9144	0.7677	81.1098	0.0001
SEX	1	0.2648	0.2416	1.2018	0.2730
AGE	1	-0.0165	0.00978	2.8470	0.0915
TRT	1	0.6926	0.2181	10.0890	0.0015
X	1	-0.9190	0.1271	52.2807	0.0001

Conditional Odds Ratio and 95% Confidence Limits

	Odds		
Variable	Ratio	Lower	Upper
INTERCP1	2.678	0.765	9.378
INTERCP2	18.710	5.201	67.301
INTERCP3	117.992	30.491	456.600
INTERCP4	999.000	223.549	999.000
SEX	1.303	0.812	2.092
AGE	0.984	0.965	1.003
TRT	1.999	1.304	3.065
X	0.399	0.311	0.512

- We see that the assumption of parallel lines (proportional odds) is not violated since the test for proportional odds is not rejected:

```
Chi-Square = 12.8763 with 12 DF (p=0.3781)
```

- We interpret the results to mean that

1. Treatment $(p = 0.0015)$ does significantly improve the response. Since the treatment effect is approximately .69, being on auranofin tends to increase the odds of response level j or lower (which means a better response), by $\exp(.69) \approx 2.0$.

Comparison to earlier results

- When we dichotomized Y earlier, we estimated $\beta_{tx} = 0.7005$ with $\exp(.7) = 2.015$.
- The estimated standard error was 0.3136 compared to the proportional odds estimate of 0.2181
- I.e., dichotomizing the outcome resulted in a loss of power for $H_0 : \beta_{tx} = 0$
- but the parameter estimate is nearly identical (as expected under the proportional odds model – i.e., same model regardless of cut point selection)

- 2. Individuals with a better baseline status tend to have a better response at thirteen weeks $(p = 0.0001)$. Since the baseline effect is approximately -.92, a one unit increase in the baseline response (say, from fair to poor), tends to decrease the odds of response level j or lower (the better response), by $\exp(-.92) \approx .4$

- 3. Older individuals seem to have a worse outcome than younger individuals $(p = 0.0915)$, although not significant at the .05 level),

- 4. SEX $(p = 0.2730)$ is not significant.

One more example

The data are reproduced from Lindsey (1995) and show the severity of pneumoconiosis as related to the number of years working at a coal factory.

	Pneumoconiosis		
Years	Normal	Mild	Severe
0.5 - 11	98	0	0
12-18	51	2	1
19-24	34	6	3
25-30	35	5	8
31-36	32	10	9
37-42	23	7	8
43-49	12	6	10
50-59	4	2	5

```
data lindsey;
 input years $rep $year count @@;
  if rep eq 'sev' then resp='asever';
   else if rep eq 'mild' then resp='bmild';
    else resp = 'normal';
    lyear = log(year);
cards;
1 norm 5.75 98 1 mild 5.75 0 1 sev 5.75 0
2 norm 15 51 2 mild 15 2 2 sev 15 1
3 norm 21.5 34 3 mild 21.5 6 3 sev 21.5 3
4 norm 27.5 35 4 mild 27.5 5 4 sev 27.5 8
5 norm 33.5 32 5 mild 33.5 10 5 sev 33.5 9
6 norm 39.5 23 6 mild 39.5 7 6 sev 39.5 8
7 norm 46 12 7 mild 46 6 7 sev 46 10
8 norm 51.5 4 8 mild 51.5 2 8 sev 51.5 5
;
run;
```

```
proc logistic;
 weight count;
 model resp = lyear / aggregate scale=1;
 run;

/* Selected Output */
              Model Information

Data Set                        WORK.LINDSEY
Response Variable               resp
Number of Response Levels       3
Number of Observations          22
Weight Variable                 count
Sum of Weights                  371
Model                           cumulative logit
Optimization Technique          Fisher's scoring
```

Selected Output

```
                    Response Profile

  Ordered                        Total              Total
   Value       resp           Frequency            Weight

      1        asever                 7          44.00000
      2        bmild                  7          38.00000
      3        normal                 8         289.00000
```

Probabilities modeled are cumulated over the lower Ordered Values.

NOTE: 2 observations having zero frequencies or weights were excluded since
      contribute to the analysis.

```
Score Test for the Proportional Odds Assumption

Chi-Square          DF       Pr > ChiSq

   0.1387            1           0.7096


        Deviance and Pearson Goodness-of-Fit Statistics

Criterion              Value          DF      Value/DF      Pr > ChiSq

Deviance               5.0007         13       0.3847         0.9752
Pearson                4.6806         13       0.3600         0.9816

Number of unique profiles: 8
```

For this data, we have good justification for the null hypothesis of proportional odds assumption and that our model fits the data well. However, we have some indication that our model is predicting greater variability than what was observed.

```
                   Analysis of Maximum Likelihood Estimates


                                        Standard           Wald
Parameter              DF    Estimate      Error     Chi-Square    Pr > ChiSq

Intercept asever        1    -10.5728     1.3463       61.6776       <.0001
Intercept bmild         1     -9.6672     1.3249       53.2392       <.0001
lyear                   1      2.5943     0.3813       46.2850       <.0001


            Odds Ratio Estimates


            Point           95% Wald
Effect     Estimate      Confidence Limits

lyear       13.387       6.340        28.268
```

Thus, our estimated logs are

$$odds \left(\frac{\text{Severe}}{\text{Mild or Normal}} \right) = \exp(-10.5728 + 2.5943 lyear)$$

and

$$odds \left(\frac{\text{Severe or Mild}}{\text{Normal}} \right) = \exp(-9.6672 + 2.5943 lyear)$$

Or, for a person working for 20 years

$$odds \left(\frac{\text{Severe}}{\text{Mild or Normal}} \right) = \exp(-10.5728 + 2.5943 \cdot \ln(20)) = 0.059$$

and

$$odds \left(\frac{\text{Severe or Mild}}{\text{Normal}} \right) = \exp(-9.6672 + 2.5943 \cdot \ln(20)) = 0.143$$

Therefore,

1. Approximately 6% (0.059/(1+0.059)) or 1 in 18 miners working for 20 years is expected to develop severe pneumoconiosis

2. Approximately 13% or roughly 1 in 8 miners working for 20 years is expected to develop severe or mild pneumoconiosis

The adjacent categories logit

- Recall, for individual i, we had the covariate vector $\mathbf{x}_i$,

- Suppose we look at categories j and $j + 1$, and we condition on the response being in one of these two categories

$$
\begin{aligned}
p_{ij}^* &= P[Y_{ij} = 1 | Y_{ij} + Y_{i,j+1} = 1, \mathbf{x}_i] \\[2mm]
&= \frac{P[Y_{ij}=1|\mathbf{x}_i]}{P[Y_{ij}=1|\mathbf{x}_i]+P[Y_{i,j+1}=1|\mathbf{x}_i]} \\[2mm]
&= \frac{p_{ij}}{p_{ij}+p_{i,j+1}}
\end{aligned}
$$

- Then, consider the logit of being in category j (given that the response is category j or $j + 1$).

$$
\begin{aligned}
\text{logit}(p_{ij}^*) &= \log\left(\frac{p_{ij}^*}{1 - p_{ij}^*}\right) \\
&= \log\left(\frac{p_{ij}/[p_{ij} + p_{i,j+1}]}{p_{i,j+1}/[p_{ij} + p_{i,j+1}]}\right) \\
&= \log\left(\frac{p_{ij}}{p_{i,j+1}}\right)
\end{aligned}
$$

- Suppose we model this logit with

$$
\begin{aligned}
\text{logit}(p_{ij}^*) &= \log\left(\frac{p_{ij}}{p_{i,j+1}}\right) \\
&= \alpha_j + \beta' \mathbf{x}_i,
\end{aligned}
$$

for $j = 1, ..., J - 1$.

- Note, β is the same for all j.

- What is the interpretation of an element of the vector β, (assuming it is a scalar)

- As was the case with ordinary logistic regression, β is the log- odds ratio for response j versus $j + 1$ when the covariate x is increased by one unit.

- The logistic model says that the log-odds ratio for going from category j to $j + 1$ is the same as going from category j' to $j' + 1$, i.e., adjacent categories have the same log-odds ratio.

- The ordering is taken into account, because categories d levels apart, i.e., $d = j' - j$, have log-odds ratio equal to

$$d\beta.$$

- For example, suppose we look at j and $j - 2$:

- For category $j - 1$ and j

$$\log\left(\frac{p_{i,j-1}}{p_{ij}}\right) = \alpha_{j-1} + \beta'\mathbf{x}_i,$$

- For category $j - 2$ and $j - 1$, ,

$$\log\left(\frac{p_{i,j-2}}{p_{i,j-1}}\right) = \alpha_{j-2} + \beta'\mathbf{x}_i,$$

- Then,

$$
\begin{aligned}
\log\left(\frac{p_{i,j-2}}{p_{ij}}\right) &= \log\left(\frac{p_{i,j-1}}{p_{ij}}\right) + \log\left(\frac{p_{i,j-2}}{p_{i,j-1}}\right) \\
&= \text{"after a little algebra"} \\
&= [\alpha_{j-1} + \beta'\mathbf{x}_i] + [\alpha_{j-2} + \beta'\mathbf{x}_i] \\
&= [\alpha_{j-1} + \alpha_{j-2}] + [2\beta']\mathbf{x}_i
\end{aligned}
$$

- Then, odds ratio for responses two levels apart is

$$[2\beta']$$

- In general, the adjacent categories logit is a special case of the polytomous logistic (so you can use a polytomous logistic regression package):

- Recall, the $J - 1$ logits for polytomous logistic regression uses the last level J as reference:

$$\log\left(\frac{p_{ij}}{p_{iJ}}\right) = [\alpha_j + ... + \alpha_{J-1} + (J - j)\beta' \mathbf{x}_i].$$

- In terms of interpretation and implementation, you do better to use the **baseline category model** or the **proportional odds model**

Pictures of the estimated response profiles

```
data estimated;
do lyear = 1.5 to 6.0 by 0.001;
 mod = "Severe v. Mild or Normal";
 prob = exp(-10.5728 + 2.5943* lyear)/ (1+exp(-10.5728 + 2.5943* lyear));
 output;
 mod="Severe or Mild v. Normal";
 prob = exp(-9.6672 + 2.5943* lyear)/ (1+exp(-9.6672 + 2.5943* lyear));
 output;
end;
run;


proc gplot data=estimated;
 plot prob * lyear =mod;
 run;
```

mod +++ Severe or Mild v. Normal +++ Severe v. Mild or Normal