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
• 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

<table>
<thead>
<tr>
<th>CASE</th>
<th>SEX</th>
<th>AGE</th>
<th>TREATMENT&lt;sup&gt;a&lt;/sup&gt;</th>
<th>BASELINE</th>
<th>13 WK.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>54</td>
<td>A</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>64</td>
<td>P</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>48</td>
<td>A</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>41</td>
<td>A</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>55</td>
<td>P</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>64</td>
<td>A</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>64</td>
<td>P</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>55</td>
<td>P</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>39</td>
<td>P</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>60</td>
<td>A</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup> A = Auranofin, P = Placebo

<sup>b</sup> 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:

\[
\text{logit}(F_{i1}) = \log \left( \frac{F_{i1}}{1 - F_{i1}} \right) = \alpha_1 + \beta' \mathbf{x}_i.
\]
• Next, we could dichotomize the outcome at $\leq 2$ vs $>2$:

\[
U_{i2} = \begin{cases} 
1 & \text{if very good or good at 13 weeks} \\
0 & \text{if fair, poor, very poor at 13 weeks} 
\end{cases}
\]

and let

\[
F_{i2} = P(U_{i2} = 1|x_i) = \text{prob very good or good}
\]

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

\[
\text{logit}(F_{i2}) = \alpha_2 + \beta'x_i.
\]

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

\[ F_{i3} = P(U_{i3} = 1 | x_i) \quad \text{and} \quad \logit(F_{i3}) = \alpha_3 + \beta' x_i \]

and

\[ F_{i4} = P(U_{i4} = 1 | x_i) \quad \text{and} \quad \logit(F_{i4}) = \alpha_4 + \beta' x_i. \]
• In general, the model is

\[
\text{logit}(F_{ij}) = \log \left( \frac{F_{ij}}{1 - F_{ij}} \right) = \alpha_j + \beta' x_i
\]

where \( j = 1, \ldots, J - 1 \) and \( \beta \) 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, \( \alpha_j \)'s are different, but the slopes (\( \beta \)'s) are the same.
Cumulative probabilities

- In general,

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

where the multinomial probabilities are

\[ p_{ij} = P[Y_{ij} = 1|x_i] \]
• We had defined the cumulative random variables $U_{ij}$:

$$U_{ij} = \begin{cases} 1 & \text{if } Y_i \leq j \\ 0 & \text{if } Y_i > j \end{cases}$$

• We also can define the cumulative probabilities as

$$F_{ij} = P[U_{ij} = 1|x_i] = P[Y_i \leq j|x_i] = p_{i1} + \ldots + p_{ij}$$

• Note, we only need the first $(J - 1)$ ‘cumulative probabilities’ $(F_{i1}, \ldots, F_{i,J-1})$ since the last one always equals 1,

$$F_{iJ} = P[Y_i \leq J|x_i] = p_{i1} + \ldots + p_{iJ} = 1$$

• 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,

\[
\text{logit}(F_{ij}) = \alpha_j + x_i'\beta,
\]

for \( j = 1, \ldots, J - 1 \)

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

• In other words, the coefficients \( \beta \) of the covariate vector \( x_i \) 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 \( \beta \) 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 $\beta$

- Suppose we have two covariates $x_i = (x_{i1}, x_{i2})$, to give the model,

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

- What is the interpretation of $\beta_1$?
- Just as in ordinary logistic regression, $\beta_1$ 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 \( \alpha_j \) (as is the case in ordinary logistic regression), the log-odds ratios will be the same, for any cumulative probability:

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

Then, the odds ratio for

\( (Y_i \leq j) \) versus \( (Y_i > j) \)

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:

\[
\begin{align*}
\text{OR(very good vs. < very good)} & = \\
\text{OR(≥ good vs. < good)} & = \\
\text{OR(≥ fair vs. < fair)} & = \\
\text{OR(≥ poor vs. very poor)} & = 
\end{align*}
\]

• When we look at the output, we will see that, unlike the above polytomous logit, we will get only one set of \( \beta \)'s, although we will get \( J - 1 \) intercepts.

\[
\text{logit}(F_{i,j}) = \alpha_j + \textbf{x}_i'\beta,
\]
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 \( j \):

\[
\text{logit}(F_{ij}) = \alpha_k + x_i' \beta_j
\]

- 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 \(\beta_j\)'s being equal,

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

Under the null, there is one \(K \times 1\) vector \(\beta\), 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

\[
df = \# \text{ parameters in full model} - \# \text{ parameters in reduced model} = (J - 1)K - K = (J - 2)K
\]
MLE’s

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

\[ p_{ij} = (p_{i1} + \ldots + p_{ij}) - (p_{i1} + \ldots + p_{i,j-1}) \]
\[ = F_{ij} - F_{i,j-1} \]

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

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

• The overall likelihood is

\[ L(\alpha, \beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} [p_{ij}(\alpha, \beta)]^{y_{ij}}, \]
• 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 (\(\text{AGE}_i\)), sex (\(\text{SEX}_i, 1=\text{male}, 0=\text{female}\)), treatment (\(\text{TRT}_i, 1 = \text{auranofin}, 0 = \text{placebo}\)), 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:

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

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

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

\[ F_{i4} = p_{i1} + p_{i2} + p_{i3} + p_{i4} = \text{prob very poor, poor, fair, or good} \]
The following ascii is in the current directory, and called art2.dat

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>41</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>39</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>35</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>35</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>65</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>42</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>52</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>63</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
/* 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

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Y</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>103</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCP1</td>
<td>1</td>
<td>0.9850</td>
<td>0.6395</td>
<td>2.3727</td>
<td>0.1235</td>
</tr>
<tr>
<td>INTERCP2</td>
<td>1</td>
<td>2.9290</td>
<td>0.6531</td>
<td>20.1114</td>
<td>0.0001</td>
</tr>
<tr>
<td>INTERCP3</td>
<td>1</td>
<td>4.7706</td>
<td>0.6904</td>
<td>47.7450</td>
<td>0.0001</td>
</tr>
<tr>
<td>INTERCP4</td>
<td>1</td>
<td>6.9144</td>
<td>0.7677</td>
<td>81.1098</td>
<td>0.0001</td>
</tr>
<tr>
<td>SEX</td>
<td>1</td>
<td>0.2648</td>
<td>0.2416</td>
<td>1.2018</td>
<td>0.2730</td>
</tr>
<tr>
<td>AGE</td>
<td>1</td>
<td>-0.0165</td>
<td>0.00978</td>
<td>2.8470</td>
<td>0.0915</td>
</tr>
<tr>
<td>TRT</td>
<td>1</td>
<td>0.6926</td>
<td>0.2181</td>
<td>10.0890</td>
<td>0.0015</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>-0.9190</td>
<td>0.1271</td>
<td>52.2807</td>
<td>0.0001</td>
</tr>
<tr>
<td>Variable</td>
<td>Odds Ratio</td>
<td>Lower</td>
<td>Upper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>--------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCP1</td>
<td>2.678</td>
<td>0.765</td>
<td>9.378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCP2</td>
<td>18.710</td>
<td>5.201</td>
<td>67.301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCP3</td>
<td>117.992</td>
<td>30.491</td>
<td>456.600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCP4</td>
<td>999.000</td>
<td>223.549</td>
<td>999.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX</td>
<td>1.303</td>
<td>0.812</td>
<td>2.092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>0.984</td>
<td>0.965</td>
<td>1.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRT</td>
<td>1.999</td>
<td>1.304</td>
<td>3.065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>0.399</td>
<td>0.311</td>
<td>0.512</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• We see that the assumption of parallel lines (proportional odds) is not violated since the test for proportional odds is not rejected:

\[ \text{Chi-Square} = 12.8763 \text{ with } 12 \text{ DF } (p=0.3781) \]

• 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.
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.

<table>
<thead>
<tr>
<th>Years</th>
<th>Normal</th>
<th>Mild</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 - 11</td>
<td>98</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12-18</td>
<td>51</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19-24</td>
<td>34</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>25-30</td>
<td>35</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>31-36</td>
<td>32</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>37-42</td>
<td>23</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>43-49</td>
<td>12</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>50-59</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Total resp</th>
<th>Total Frequency</th>
<th>Total Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 asever</td>
<td>7</td>
<td>44.00000</td>
<td></td>
</tr>
<tr>
<td>2 bmild</td>
<td>7</td>
<td>38.00000</td>
<td></td>
</tr>
<tr>
<td>3 normal</td>
<td>8</td>
<td>289.00000</td>
<td></td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.

**NOTE:** 2 observations having zero frequencies or weights were excluded since they 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

<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>5.0007</td>
<td>13</td>
<td>0.3847</td>
<td>0.9752</td>
</tr>
<tr>
<td>Pearson</td>
<td>4.6806</td>
<td>13</td>
<td>0.3600</td>
<td>0.9816</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Parameter</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 asever</td>
<td>1</td>
<td>-10.5728</td>
<td>1.3463</td>
<td>61.6776</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept bmild</td>
<td>1</td>
<td>-9.6672</td>
<td>1.3249</td>
<td>53.2392</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>lyear</td>
<td>1</td>
<td>2.5943</td>
<td>0.3813</td>
<td>46.2850</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

### Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
</table>
Thus, our estimated logs are

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

and

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

Or, for a person working for 20 years

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

and

\[
\text{odds}\left(\frac{\text{Severe or Mild}}{\text{Normal}}\right) = \exp(-9.6672 + 2.5943 \cdot \ln(20)) = 0.143
\]
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 $x_i$,
- Suppose we look at categories $j$ and $j + 1$, and we condition on the response being in one of these two categories

$$p_{ij}^* = P[Y_{ij} = 1|Y_{ij} + Y_{i,j+1} = 1, x_i]$$

$$= \frac{P[Y_{ij} = 1|x_i]}{P[Y_{ij} = 1|x_i] + P[Y_{i,j+1} = 1|x_i]}$$

$$= \frac{p_{ij}}{p_{ij} + p_{i,j+1}}$$
Then, consider the logit of being in category \( j \) (given that the response is category \( j \) or \( j + 1 \)).

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

Suppose we model this logit with

\[
\text{logit}(p_{ij}^*) = \log \left( \frac{p_{ij}}{p_{i,j+1}} \right) \\
= \alpha_j + \beta' x_i,
\]

for \( j = 1, \ldots, J - 1 \).

Note, \( \beta \) is the same for all \( j \).
• What is the interpretation of an element of the vector $\beta$, (assuming it is a scalar)
• As was the case with ordinary logistic regression, $\beta$ 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 \( j \)

\[
\log \left( \frac{p_{i,j-1}}{p_{ij}} \right) = \alpha_{j-1} + \beta' x_i,
\]

• For category \( j - 2 \) and \( j - 1 \),

\[
\log \left( \frac{p_{i,j-2}}{p_{i,j-1}} \right) = \alpha_{j-2} + \beta' x_i,
\]

• Then,

\[
\log \left( \frac{p_{i,j-2}}{p_{ij}} \right) = \log \left( \frac{p_{i,j-1}}{p_{ij}} \right) + \log \left( \frac{p_{i,j-2}}{p_{i,j-1}} \right)
= \text{"after a little algebra"}
= [\alpha_{j-1} + \beta' x_i] + [\alpha_{j-2} + \beta' x_i]
= [\alpha_{j-1} + \alpha_{j-2}] + [2\beta'] x_i
\]

• 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:

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

• In terms of interpretation and implementation, you do better to use the baseline category model or the proportional odds model.
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;