Logistic Regression for Ordinal Response Variables
Edpsy/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology

I L L I N O I S
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

© Board of Trustees, University of Illinois

Spring 2017
Outline

Common models for ordinal responses:

- Cumulative logit model, or the “proportional odds” model.
- Adjacent categories logit model.
- Continuation ratio logits.

They differ in terms of

- How logits are formed.
- Some can summarize association with 1 parameter per predictor.
- Some can allow different models for different logits.
Logit Models for Ordinal Responses

The logit models for this situation

- Use the ordering of the categories in forming logits.
- Yield simpler models with simpler interpretations than nominal model.
- Are more powerful than nominal models.
Proportional Odds Model

or **Cumulative Logit Model**

Form logits (dichotomize categories of $Y$) incorporating the ordinal information.

**Cumulative Probabilities:**

- $Y = 1, 2, \ldots, J$ and order is relevant.
- $\{\pi_1, \pi_2, \ldots, \pi_J\}$. 
- $P(Y \leq j) = \pi_1 + \ldots + \pi_j = \sum_{k=1}^{j} \pi_k$ for $j = 1, \ldots, J - 1$.
- “Cumulative logits”

$$
\log \left( \frac{P(Y \leq j)}{P(Y > j)} \right) = \log \left( \frac{P(Y \leq j)}{1 - P(Y \leq j)} \right)
$$

$$
= \log \left( \frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J} \right) \quad \text{for} \quad j = 1, \ldots, J - 1
$$
“Proportional Odds Model”

\[
\text{logit}(P(Y \leq j)) = \log \left( \frac{P(Y \leq j)}{P(Y > j)} \right) = \alpha_j + \beta x \quad \text{for} \quad j = 1, \ldots, J-1
\]

- \(\alpha_j\) (intercepts) can differ.
- \(\beta\) (slope) is constant.
  - The effect of \(x\) is the same for all \(J-1\) ways to collapse \(Y\) into dichotomous outcomes.
  - A single parameter describes the effect of \(x\) on \(Y\) (versus \(J-1\) in the baseline model).
- Interpretation in terms of odds ratios.
Interpretation

For a given level of $Y$ (say $Y = j$)

$$\frac{P(Y \leq j|X = x_2)/P(Y > j|X = x_2)}{P(Y \leq j|X = x_1)/P(Y > j|X = x_1)} = \frac{P(Y \leq j|x_2)P(Y > j|x_1)}{P(Y \leq j|x_1)P(Y > j|x_2)}$$

$$= \exp(\alpha_j + \beta x_2)/\exp(\alpha_j + \beta x_1)$$

$$= \exp[\beta(x_2 - x_1)]$$

or log odds ratio $= \beta(x_2 - x_1)$.

The log cumulative odds ratio is proportional to the difference (distance) between $x_1$ and $x_2$.

Since the proportionality coefficient $\beta$ is constant, this model is called the “Proportional Odds Model”.
Properties of Model

- Note that the cumulative probabilities are given by
  \[
P(Y \leq j) = \frac{\exp(\alpha_j + \beta x)}{1 + \exp(\alpha_j + \beta x)}
\]

Since \( \beta \) is constant, curves of cumulative probabilities plotted against \( x \) are parallel.

- We can compute the probability of being in category \( j \) by taking differences between the cumulative probabilities.
  \[
P(Y = j) = P(Y \leq j) - P(Y \leq j - 1) \quad \text{for} \quad j = 2, \ldots, J
\]
  and
  \[
P(Y = 1) = P(Y \leq 1)
\]

Since \( \beta \) is constant, these probabilities are guaranteed to be non-negative.

- In fitting this model to data, it must be simultaneous.
Example: HSB

Example: High School and Beyond

\[ X = \text{mean of 5 achievement test scores.} \]

\[ Y = \begin{cases} 
1 & \text{Academic} \\
2 & \text{General} \\
3 & \text{VoTech} 
\end{cases} \]

So the logit model is

\[
\begin{align*}
\text{Academic vs (Gen & VoTech):} & \quad \logit(Y \leq 1) = \alpha_1 + \beta x \\
\text{(Academic & Gen) vs VoTech:} & \quad \logit(Y \leq 2) = \alpha_2 + \beta x
\end{align*}
\]
Test of Proportional Odds Assumption

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8194</td>
<td>1</td>
<td>0.3653</td>
</tr>
</tbody>
</table>

If this test is significant, then proportional odds model is not a good one for the data. (Later we’ll talk about what to do if it’s significant.)
Example: Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>$e^\beta$</th>
<th>ASE</th>
<th>Wald</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>-6.8408</td>
<td>.6118</td>
<td>125.04</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-5.5138</td>
<td>.5866</td>
<td>88.37</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>.1330</td>
<td>1.142</td>
<td>.0118</td>
<td>127.64</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

For a 10 point increase in mean achievement, the odds ratio (for either case) equals

$$\exp(10(.1330)) = 3.78$$
Fitted Cumulative Probabilities

- P(Academic)
- P(Academic or General)

Fitted Cumulative Probabilities vs. Achievement Scores

C.J. Anderson (Illinois) Logistic Regression for Ordinal Responses
Fitted Category Probabilities

Fitted Probabilities

- \( P(\text{Academic}) \)
- \( P(\text{General}) \)
- \( P(\text{Vo/Tech}) \)

Achievement Scores

C.J. Anderson (Illinois)
Observed Proportions and Fitted $\pi_j$s

(Grouped only for plot)
Estimation in SAS

- LOGISTIC (maximum likelihood).
- CATMOD (weighted least squares).
- GENMOD
- NLP or NLMIXED (maximum likelihood).
- Others

For larger samples with categorical explanatory variables, results from MLE and WLS should be about same.
SAS Logistic & GENMOD Code

proc logistic ;
    model hsp = achieve;

In proc logistic, the cumulative logit model is the default if the response variable has more than 2 categories.

proc genmod;
    model = achieve / dist=multinomial link=clogit type3;

“clogit” for Cumulative Logit.
### SAS PROC LOGISTIC (edited) Output

#### Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Program</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>academic</td>
<td>308</td>
</tr>
<tr>
<td>2</td>
<td>general</td>
<td>145</td>
</tr>
<tr>
<td>3</td>
<td>vocation</td>
<td>147</td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.
Note on Example

We would get the exact same results regarding interpretation if we had used (i.e., put in descending option in proc LOGISTIC).

\[
Y = \text{high school program type} = \begin{cases} 1 & \text{VoTech} \\ 2 & \text{General} \\ 3 & \text{Academic} \end{cases}
\]

This reversal of the ordering of \( Y \) would

- Change the signs of the estimated parameters.
- Yield curves of cumulative probabilities that decrease (rather than increase).
- Basically the same results.
Example 2: PIRLS

US 2006 Progress in International Reading Literacy Study (PIRLS) responses to item “How often do you use the Internet as a source of information for school-related work?” with responses:

- Every day or almost every day ($y_1 = 746$, $p_1 = .1494$)
- Once or twice a week ($y_2 = 1,240$, $p_2 = .2883$)
- Once or twice a month ($y_3 = 1,377$, $p_3 = .2757$)
- Never or almost never ($y_4 = 1,631$, $p_4 = .3266$)

Predictors/Explanatory:

- Shortages at school.
- Time student spends in front of screen (electronic entertainment)
- Gender of student.
Graph of PIRLS Distribution

2006 US PIRLS on Internet Use for School

Response

Percent

1_Daily 2_Weekly 3_Monthly 4_Never
**Problem with Model?**

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>49.1500</td>
<td>6</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

\(H_o\): \(\beta_s\) are same over logits for all predictors.

\(H_a\): They are not all the same.
If Reject Proportional Odds Assumption

- If test is rejected this result could be due to large sample but not practical or substantively important. To investigate this fit separate logistic regressions to each logit.**
- Add additional terms.
- Try non-symmetric link function.
- Use a different ordinal model.
- Add dispersion parameters.
- Permit separate effects for some variables ("partial proportional odds")**
- Use the baseline model but use order to interpret the results.**
<table>
<thead>
<tr>
<th>Effect</th>
<th>Proportional Odds</th>
<th>Separate Binary Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P(Y = 1)/P(Y &gt; 1)$</td>
<td>$P(Y \leq 2)/P(Y &gt; 2)$</td>
</tr>
<tr>
<td>shortages</td>
<td>-0.2055</td>
<td>-0.0685</td>
</tr>
<tr>
<td>girl</td>
<td>0.2225</td>
<td>0.1223</td>
</tr>
<tr>
<td>screenT</td>
<td>0.0599</td>
<td>0.1904</td>
</tr>
</tbody>
</table>

- Shortages: Differ in terms of magnitude.
- Gender: Similar values.
- Screen Time: Different direction of effects.

Not just statistical but also substantive differences
Partial Proportional Odds

Relax assumption for shortages and allow different parameters for it.

Edited Output from PROC NLMIXED:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std Error</th>
<th>DF</th>
<th>t</th>
<th>Pr</th>
<th>Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>-1.9499</td>
<td>0.08314</td>
<td>4377</td>
<td>-23.45</td>
<td>&lt; .0001</td>
<td>-0.00027</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>-0.8769</td>
<td>0.06464</td>
<td>4377</td>
<td>-13.57</td>
<td>&lt; .0001</td>
<td>0.000807</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>0.6976</td>
<td>0.07237</td>
<td>4377</td>
<td>9.64</td>
<td>&lt; .0001</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Girl</td>
<td>0.1138</td>
<td>0.04423</td>
<td>4377</td>
<td>2.57</td>
<td>.0101</td>
<td>0.000621</td>
</tr>
<tr>
<td>ScreenT</td>
<td>0.0471</td>
<td>0.02001</td>
<td>4377</td>
<td>2.35</td>
<td>.0187</td>
<td>-0.00057</td>
</tr>
<tr>
<td>Shortage 1</td>
<td>-0.0603</td>
<td>0.08061</td>
<td>4377</td>
<td>-0.75</td>
<td>.4543</td>
<td>0.000023</td>
</tr>
<tr>
<td>Shortage 2</td>
<td>-0.1394</td>
<td>0.04256</td>
<td>4377</td>
<td>-3.27</td>
<td>.0011</td>
<td>0.000087</td>
</tr>
<tr>
<td>Shortage 3</td>
<td>-0.2560</td>
<td>0.05864</td>
<td>4377</td>
<td>-4.37</td>
<td>&lt; .0001</td>
<td>-0.00045</td>
</tr>
</tbody>
</table>
Interpretation of Shortages

For fixed gender and screen time,

- The odds ratio daily versus more than daily usage for shortages $x + 1$ equals $\exp(-0.0603) = 0.94$ the odds for shortage $x \rightarrow$ equal odds.

- The odds ratio for daily or weekly use versus monthly or never for $x + 1$ shortages equals $\exp(-0.1394) = 0.87$ the odds for $x$ shortages.

- The odds ratio for monthly or more usage versus never for shortages $x + 1$ equals $\exp(-0.2560) = 0.77$

What does this mean:

- More shortages less frequently use computers?
- More shortage more frequently use computers?
Baseline Model but Use Order

All possible odds ratios: For 1 unit increase in shortage, the odds ratios for row versus column equal

<table>
<thead>
<tr>
<th></th>
<th>Daily</th>
<th>Weekly</th>
<th>Monthly</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>—</td>
<td>1.11</td>
<td>.98</td>
<td>.80</td>
</tr>
<tr>
<td>Weekly</td>
<td>0.90</td>
<td>—</td>
<td>.89</td>
<td>.72</td>
</tr>
<tr>
<td>Monthly</td>
<td>1.02</td>
<td>1.12</td>
<td>—</td>
<td>.81</td>
</tr>
<tr>
<td>Never</td>
<td>1.25</td>
<td>1.39</td>
<td>1.23</td>
<td>—</td>
</tr>
</tbody>
</table>

- The odds of Daily versus Weekly are 1.11 the odds for 1 unit more on shortages.
- For greater shortages, daily use of computers is more likely than weekly.
- For fewer shortages, monthly or never using computers is more likely than daily use.
Interpretation of Shortages

For better and more proper analysis of data see Anderson, Kim & Keller (2010) and see results for multinomial model.

When take into account hierarchical structure, missing data and unequal probability sampling (particularly of the school), the impact of shortages of computer use quite different.
Final Comments on Cumulative Logit Models

- It takes into account the ordering of the categories of the response variable.
- One probability is monotonically increasing as a function of $x$. (see figure of estimated probabilities from HSB example).
- One probability is monotonically decreasing as a function of $x$. (see figure of estimated probabilities).
- Curves of probabilities for intermediate categories are uni-modal with the modes (maximum) corresponding to the order of the categories.
- The conclusions regarding the relationship between $Y$ and $x$ are not affected by the response category.
Final Comments on Cumulative Logit Models

- The specific combination of categories examined does not lead to substantially different conclusions regarding the relationship between responses and $x$.
- IRT connection: Samejima’s (1969) graded response model for polytomous items is the same as the proportional odds model except that $x$ is a latent continuous variable.
Adjacent–Categories Logit Models

Rather than using all categories in forming logits, we can just use $J - 1$ pairs of them.

To incorporate the ordering of the response, we use adjacent categories:

$$\log \left( \frac{\pi_j}{\pi_{j+1}} \right) \quad j = 1, \ldots, J - 1$$

The logit model for one (continuous) explanatory variable $x$ is

$$\log \left( \frac{\pi_j}{\pi_{j+1}} \right) = \alpha_j + \beta x \quad j = 1, \ldots, J - 1$$
Adjacent–Categories Logit Models

- This model is a special case of the baseline model. (shown below)

- It would not work for the PIRLs example

<table>
<thead>
<tr>
<th></th>
<th>Daily</th>
<th>Weekly</th>
<th>Monthly</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>—</td>
<td>1.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weekly</td>
<td>0.90</td>
<td>—</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Monthly</td>
<td>1.12</td>
<td>—</td>
<td></td>
<td>0.81</td>
</tr>
<tr>
<td>Never</td>
<td>1.23</td>
<td>—</td>
<td></td>
<td>—</td>
</tr>
</tbody>
</table>

- If we had a single $\beta$, these odds ratios would all be equal.
An Example for Adjacent Categories

GSS Happiness data from Agresti (2013):

- Response variable is happiness with categories 1 = very happy, 2 = pretty happy, and 3 = not too happy.
- Predictors are
  - Race with categories 1 = black and 0 = white.
  - Number of traumatic events that happened to respondent or relatives in the last year. Values range from 0 to 5.
- Estimated model:

\[
\log(P(Y_i = j)/P(Y_i = j+1)) = \hat{\alpha}_j - 0.357(\text{traumatic})_i - 1.84(\text{race})_i
\]

Note: \(\hat{\alpha}_1 = 2.532\) and \(\hat{\alpha}_2 = 3.028\)
Interpretation

Estimated model:

$$\log(P(Y = j)/P(Y = j+1)) = \hat{\alpha}_j - 0.357(\text{traumatic}) - 1.842(\text{race});$$

- Given number of traumatic events, the estimated odds of being very happy versus pretty happy for whites are $$\exp(1.842) = 6.31$$ times the odds for blacks.
- Given number of traumatic events, the estimated odds of being pretty happy versus not too happy for whites are $$\exp(1.842) = 6.31$$ times the odds for blacks — the same.
- Given race, the estimated odds of very happy versus pretty happy for $$x$$ traumatic events are $$1/\exp(-.357) = 1.429$$ times the odds for $$x + 1$$ events.
- Odd ratio for pretty happy versus not too happy are the same as above.
Estimation

- CATMOD: Weighted least squares, but if there are 0s, need to add a small number to each cell.
- CATMOD: Maximum likelihood estimation involves a design matrix that puts restrictions on parameters of the baseline model.
- NLMIXED: MLE for baseline but modify to correspond to adjacent categories.
CATMOD and WLS

```sas
title 'Check for zeros';
proc freq data=gss;
tables race*trauma*happy / nopercent norow nocol sparse out=table;
data fillin;
  set table;
  count2=count+.01;
title 'Adjacent Categories (WLS)';
proc catmod data=fillin;
  weight count2;
  response alogits;
  population race trauma;
  direct trauma race;
  model happy = _response_ race trauma;

Will be run in lecture
```
To Use NLMIXED

We make use of the fact that the adjacent categories models is a special case of the baseline model.

Baseline odds \(=\) Product of adjacent categories odds , and logarithm of odds equals sum

\[
\log \left( \frac{\pi_{ij}}{\pi_{iJ}} \right) = \log \left( \frac{\pi_{ij}}{\pi_{ij+1}} \right) + \log \left( \frac{\pi_{i(j+1)}}{\pi_{i(j+2)}} \right) + \ldots \log \left( \frac{\pi_{i(J-1)}}{\pi_{iJ}} \right)
\]

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

e.g., Taking a simple model for the adjacent categories,

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

\[
= \sum_{k=j}^{J-1} \alpha_k + \beta (J - j) x_i
\]

\[\alpha_j^*, \beta_j^* \]
**Proportional Odds**

**Adjacent-Categories**

**Continuation-ratio**

---

**NLMIXED & MLE**

```plaintext
title 'Adjacent Categories (MLE)';
proc nlmixed data=gss;  * ← un-collapsed data;
  parms a1=0.1 a2=0.1 br=0.1 bt=0.1;
  /* Linear predictors */
  eta1 = a1 + br*(3-1)*race + bt*(3-1)*trauma;
  eta2 = a2 + br*(3-2)*race + bt*(3-2)*trauma;
  /* Define likelihood */
  if happy=1 then prob= exp(eta1)/(1 + exp(eta1) + exp(eta2));
  if happy=2 then prob= exp(eta2)/(1 + exp(eta1) + exp(eta2));
  if happy=3 then prob= 1/(1 + exp(eta1) + exp(eta2));
  /* To make sure that probabilities are valid ones */
  p = (prob>0 and prob<=1)*prob + (prob<=0)*1e-8 + (prob>1);
  loglike = log(p);
  /* Specify distribution for response variable */
  model happy ~ general(loglike);
```
### Comparison of WLS & MLE

#### Weighted Least Squares Estimates from CATMOD

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.1148</td>
<td>0.3698</td>
<td>9.09</td>
<td>0.0026</td>
</tr>
<tr>
<td>RESPONSE_1</td>
<td>1.7341</td>
<td>0.3056</td>
<td>32.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>race</td>
<td>1.7317</td>
<td>0.8031</td>
<td>4.65</td>
<td>0.0311</td>
</tr>
<tr>
<td>trauma</td>
<td>0.2053</td>
<td>0.1876</td>
<td>1.20</td>
<td>0.2740</td>
</tr>
</tbody>
</table>

#### MLE from NLMIXED

| Parameter | Estimate | Error  | df | t    | Pr > |t| |
|-----------|----------|--------|----|------|-------|---|
| a1        | 2.5315   | 0.7464 | 23 | 3.39 | 0.0025 |
| a2        | 3.0276   | 0.5740 | 23 | 5.27 | <.0001 |
| br        | -1.8424  | 0.6419 | 23 | -2.87| 0.0087 |
| bt        | -0.3570  | 0.1640 | 23 | -2.18| 0.0400 |
Adjacent Categories or Proportional Odds Model? (from Agresti, 2013)

- Both tend to fit (or not) for a particular data set.
- If prefer effects to refer to individual categories, use adjacent categories.
- If want to use entire scale for each logit or hypothesize underlying continuous latent variable, use proportional odds model.
- Effects for proportional odds tend to be larger because whole scale is used.
- Proportional odds models not effected by choice and number of response categories.
- Adjacent is more general than proportional odds model—if replace $\beta$ by $\beta_j$ in the adjacent model, cumulative probabilities will be in correct order—this isn’t true for the partial proportional odds model.
Adjacent Categories for Ordered Grouped Data

- Recall... General Social Survey (1994) data from before.
  - Item 1: A working mother can establish just as warm and secure of a relationship with her children as a mother who does not work.
  - Item 2: Working women should have paid maternity leave.
- When using $u_i = i$ and $v_j = j$ as scores and fitting the independence log-linear model and the uniform association model:
  $$\log(\mu_{ij}) = \lambda + \lambda_i^l + \lambda_j^ll + \beta_{ij}$$
- Results from model fitting:

<table>
<thead>
<tr>
<th>Model/Test</th>
<th>$df$</th>
<th>$G^2$</th>
<th>$p$</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>12</td>
<td>44.96</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Uniform Assoc</td>
<td>11</td>
<td>8.67</td>
<td>.65</td>
<td>$\hat{\beta} = .24$, $ASE = .0412$</td>
</tr>
</tbody>
</table>
Adjacent Categories for Ordered Grouped Data

Suppose that we consider item 2 as the response variable and model adjacent category logits with the restriction that $\beta_j = \beta = a$ constant.

$$
\log \left( \frac{\mu_i(j+1)}{\mu_{ij}} \right) = \lambda + \lambda_i + \lambda_{j+1} + \beta i(j + 1)
$$

$$
-(\lambda + \lambda_i + \lambda_{j+1} + \beta ij)
$$

$$
= (\lambda_{j+1} - \lambda_{j}) + \beta (ij + i - ij)
$$

$$
= \alpha^*_{j} + \beta i
$$

So the estimated local odds ratio equals (and the effect of response on item 1 on item 2 for adjacent categories)

$$
e^{\hat{\beta}} = e^{24} = 1.28$$
In this approach, the order of the categories of the response variable is used to form \((J - 1)\) logits as follows:

\[
\log \left( \frac{\pi_1}{\pi_2} \right), \log \left( \frac{\pi_1 + \pi_2}{\pi_3} \right), \ldots, \log \left( \frac{\pi_1 + \ldots + \pi_{J-1}}{\pi_J} \right)
\]

or

\[
\log \left( \frac{\pi_1}{\pi_2 + \ldots + \pi_J} \right), \log \left( \frac{\pi_2}{\pi_3 + \ldots + \pi_J} \right), \ldots, \log \left( \frac{\pi_{J-1}}{\pi_J} \right)
\]

These are called “continuation–ratio logits.”
Continuation–ratios Logit

- Just apply regular binary logistic regression to each one.
- The fitting is separate (no restrictions on parameters across the logits).
- The sum of the separate $df$ and $G^2$ provide an overall global goodness of fit test and measure.
Example


$n = 978$ of 20-22 year old men from NYLS.

<table>
<thead>
<tr>
<th>Race</th>
<th>Father’s education</th>
<th>Employment Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>In school</td>
</tr>
<tr>
<td>White/other</td>
<td>≤ 12 yr</td>
<td>204</td>
</tr>
<tr>
<td>Black</td>
<td>≤ 12 yr</td>
<td>100</td>
</tr>
<tr>
<td>White/other</td>
<td>&gt; 12 yr</td>
<td>78</td>
</tr>
<tr>
<td>Black</td>
<td>&gt; 12 yr</td>
<td>12</td>
</tr>
</tbody>
</table>

Best baseling/multinomial model was (R,F).
### Example

<table>
<thead>
<tr>
<th>Logit Model</th>
<th>df</th>
<th>$G^2$</th>
<th>p</th>
<th>df</th>
<th>$G^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>null</td>
<td>3</td>
<td>19.1576</td>
<td>&lt; .01</td>
<td>3</td>
<td>16.5575</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>(F)</td>
<td>2</td>
<td>17.8385</td>
<td>&lt; .01</td>
<td>2</td>
<td>9.7941</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>(R)</td>
<td>2</td>
<td>2.6484</td>
<td>.27</td>
<td>2</td>
<td>6.0043</td>
<td>.05</td>
</tr>
<tr>
<td>(F,R)</td>
<td>1</td>
<td>2.3879</td>
<td>.12</td>
<td>1</td>
<td>1.3512</td>
<td>.25</td>
</tr>
</tbody>
</table>

- Test between (R) and (F,R),
  
  \[ G^2((R)|(F, R)) = 6.0043 − 1.3512 = 4.6531, \ df = 1, \ p = .03. \]
  
- Total: \[ G^2 = 2.3879 + 1.3512 = 3.7391, \ df = 3, \ p = .29 \]
  
- Only Race is needed for $P(\text{School})/P(\text{Working})$.
  
- Father’s education and race needed for $P(\text{School or Working})/P(\text{Inactive})$. 

---

C.J. Anderson (Illinois)  
Logistic Regression for Ordinal Responses  
Spring 2017
Recommendation

The overriding determinate of which model you should reflect the goals of the analysis.


Research Questions:

- What predicts whether a woman has even been screened?
- Among those who have ever been screened, what predicts whether screening is up to date?

What model should (did) we use?