Visualizing Categorical Data with SAS and R

Michael Friendly
York University
Short Course, 2012
Web notes: datavis.ca/courses/VCD/

Part 5: Polytomous response models

Topics:
- Proportional odds model
- Nested dichotomies
- Generalized logit models

Polytomous responses: Overview

- m categories \(\rightarrow (m - 1)\) comparisons (logits)
- Response categories ordered, e.g., None, Some, Marked improvement
  - Proportional odds model
    - Uses adjacent-category logits
    - Assumes slopes are the same for all \(m - 1\) logits; only intercepts vary
  - Nested dichotomies
    - Model each logit separately
    - \(G^2\) s are additive \(\rightarrow\) combined model
- Response categories unordered, e.g., vote NDP, Liberal, Tory, Green
  - Multinomial logistic regression
    - Uses generalized logits (LINK=GLOGIT) in PROC LOGISTIC
    - R: multinom() function in nnet package
  - Nested dichotomies
Fitting and graphing: Overview

SAS, using basic capabilities:
- output dataset contains predicted probabilities (and logits) and std errors
- Utility macros (LABELS, BARS, PSSCALE) allow plot customization

SAS, using ODS graphics (enhanced in Ver 9.2)
- plots= option for odds ratio, influence, etc
- effectplot statement can produce a variety of plots: boxplots, contour plots, interaction plots, etc.

R:
- Model objects contain all necessary information for plotting
- Basic diagnostic plots with `plot(model)`
- Fitted values with `predict();` customize with `points(), lines(),` etc.
- Effect plots most general

Proportional odds model

Consider a logistic regression model for each logit:

\[
\begin{align*}
\text{logit}(\theta_{ij1}) &= \alpha_1 + x_i^T \beta_1 & \text{None vs. Some/Marked} \\
\text{logit}(\theta_{ij2}) &= \alpha_2 + x_i^T \beta_2 & \text{None/Some vs. Marked}
\end{align*}
\]

- Proportional odds assumption: regression functions are parallel on the logit scale i.e., \( \beta_1 = \beta_2 \).

Ordinal response: Proportional odds model

Arthritis treatment data:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Treatment</th>
<th>None</th>
<th>Some</th>
<th>Marked</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Active</td>
<td>6</td>
<td>5</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>F</td>
<td>Placebo</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>32</td>
</tr>
<tr>
<td>M</td>
<td>Active</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>M</td>
<td>Placebo</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
</tbody>
</table>

Model logits for adjacent category cutpoints:

\[
\begin{align*}
\text{logit}(\theta_{ij1}) &= \log \frac{\pi_{ij1}}{\pi_{ij2} + \pi_{ij3}} = \text{logit ( None vs. [Some or Marked] )} \\
\text{logit}(\theta_{ij2}) &= \log \frac{\pi_{ij1} + \pi_{ij2}}{\pi_{ij3}} = \text{logit ( [None or Some] vs. Marked)}
\end{align*}
\]
**Proportional odds: Latent variable interpretation**

A simple motivation for the proportional odds model:

- Imagine a continuous, but *unobserved* response, $\xi$, a linear function of predictors
  \[ \xi_i = \beta^T x_i + \epsilon_i \]
- The *observed* response, $Y$, is discrete, according to some *unknown* thresholds, $\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$
- That is, the response, $Y = i$ if $\alpha_i \leq \xi_i < \alpha_{i+1}$
- Thus, intercepts in the proportional odds model $\sim$ thresholds on $\xi$

![Diagram of latent variable interpretation](image)

**Proportional odds: Fitting and plotting in SAS**

Similar to binary response models, except:

- Response variable has $m > 2$ levels; output dataset has `_LEVEL_` variable
- Must ensure that response levels are ordered as you want— use order=data or descending options.
- Validity of analysis depends on proportional odds assumption. Test of this assumption appears in PROC LOGISTIC output.

Example, using dependent variable `improve`, with values 0, 1, and 2:

```sas
proc logistic data=arthrit descending;
  class sex (ref=last) treat (ref=first) / param=ref;
  model improve = sex treat age ;
  output out=results p=prob l=lower u=upper
    xbeta=logit stdxbeta=selogit / alpha=.33;
  proc print data=results(obs=6);
    id id treat sex;
    var improve _level_ prob lower upper logit;
    format prob lower upper logit selogit 6.3;
  run;
```
The response profile displays the ordering of the outcome variable (decreasing here)

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>improve</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>42</td>
</tr>
</tbody>
</table>

Test of Proportional Odds Assumption: OK

<table>
<thead>
<tr>
<th>Score Test for the Proportional Odds Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>2.4916</td>
</tr>
</tbody>
</table>

Odds ratios (exp(β_j))

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>sex Female vs Male</td>
<td>3.496</td>
<td>1.232 9.918</td>
</tr>
<tr>
<td>treat Treated vs Placebo</td>
<td>5.728</td>
<td>2.248 14.594</td>
</tr>
<tr>
<td>age</td>
<td>1.059</td>
<td>1.002 1.077</td>
</tr>
</tbody>
</table>

i.e., Treated 5.73 times as likely to show more improvement.

Parameter estimates (β_i):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>-4.6826</td>
<td>1.1949</td>
<td>15.3566</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-3.7836</td>
<td>1.1530</td>
<td>10.7680</td>
<td>0.0010</td>
</tr>
<tr>
<td>sex Female</td>
<td>1</td>
<td>1.2515</td>
<td>0.5321</td>
<td>5.5330</td>
<td>0.0187</td>
</tr>
<tr>
<td>treat Treated</td>
<td>1</td>
<td>1.7453</td>
<td>0.4772</td>
<td>13.3774</td>
<td>0.0003</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>0.0382</td>
<td>0.0185</td>
<td>4.2361</td>
<td>0.0396</td>
</tr>
</tbody>
</table>

To plot predicted probabilities in a single graph, combine values of TREAT and _LEVEL_

--- glogist2a.sas ---

Add error bars and legends:

--- glogist2a.sas ---
Proportional odds model fitting and plotting in SAS

Plot step:
```sas
options hby=0;
proc gplot data=results;
plot prob * age = treat1 / 
  vaxis=axis1 haxis=axis2 hminor=1 vminor=1 
  nolegend anno=bars name=glogist2a;
by sex;
axis1 label=(a=90 'Prob. Improvement (67% CI)') 
  order=(0 to 1 by .2); 
axis2 order=(20 to 80 by 10) 
  offset=(2,5);
symbol1 v=circle i=join line=3 c=black;
symbol2 v=circle i=join line=3 c=black;
symbol3 v=dot i=join line=1 c=red;
symbol4 v=dot i=join line=1 c=red;
run;
```

Effect plots using SAS ODS
```
ods graphics on;
proc logistic data=arthrit descending ;
class sex (ref=last) treat (ref=first) / param=ref;
model improve = sex treat age / clodds=wald expb;
effectplot slicefit(sliceby=improve plotby=Treat) / at(sex=all) clm alpha=0.33;
effectplot interaction(sliceby=improve x=Treat) / at(sex=all) clm alpha=0.33;
run;
ods graphics off;
```

Proportional odds models in R

Fitting: `polr()` in MASS package
```
The response, Improved has been defined as an ordered factor

> factor(Arthritis$Improved)
```

```r
[1] Some None None Marked Marked Marked None Marked None ...
[81] None Some Some Marked
Levels: None < Some < Marked
```

Fitting:
```
library(vcd)
library(car)  # for Anova()
arth.polr <- polr(Improved ~ Sex + Treatment + Age, 
data=Arthritis)
summary(arth.polr)
Anova(arth.polr)  # Type II tests
```
The `summary()` function gives standard statistical results:

```r
> summary(arth.polr)
```

Call:
```
polr(formula = Improved ~ Sex + Treatment + Age, data = Arthritis)
```

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SexMale</td>
<td>-1.251679</td>
<td>0.546365</td>
<td>-2.2909</td>
</tr>
<tr>
<td>TreatmentTreated</td>
<td>1.745289</td>
<td>0.475895</td>
<td>3.6674</td>
</tr>
<tr>
<td>Age</td>
<td>0.038162</td>
<td>0.018416</td>
<td>2.0722</td>
</tr>
</tbody>
</table>

Intercepts:

<table>
<thead>
<tr>
<th>Level</th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Some</td>
<td>2.5319</td>
<td>1.0571</td>
</tr>
<tr>
<td>Some</td>
<td>Marked</td>
<td>3.4309</td>
<td>1.0912</td>
</tr>
</tbody>
</table>

Residual Deviance: 145.4579
AIC: 155.4579

```r
> Anova(arth.polr)  # Type II tests
```

Anova Table (Type II tests)

<table>
<thead>
<tr>
<th>Response</th>
<th>LR Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>5.6880</td>
<td>1</td>
<td>0.0170</td>
</tr>
<tr>
<td>Treatment</td>
<td>14.7095</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>4.5715</td>
<td>1</td>
<td>0.0325</td>
</tr>
</tbody>
</table>

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Proportional odds models in R: Plotting

Making visual comparisons easier:

```r
> library(effects)
> plot(effect("Treatment:Age", arth.polr))
```

- The default plot shows all details
- But, is harder to compare across treatment and response levels

```r
> plot(effect("Sex:Age", arth.polr), style='stacked')
```

The `plot()` function in the `effects` package can be used to visualize the effects of predictors on the outcome variable. This is particularly useful when the effects are not linear or when there are interactions between predictors.

Proportional odds models in R: Plotting

Making visual comparisons easier:

```r
> plot(effect("Treatment:Age", arth.polr), style='stacked')
```

```r
> plot(effect("Sex:Age", arth.polr), style='stacked')
```
Proportional odds model
Proportional odds models in R

These plots are even simpler on the logit scale, using latent=TRUE to show the cutpoints between response categories

```r
plot(effect("Treatment:Age", arth.polr, latent=TRUE))
```

Treatment*Age effect plot

Age

Improved: None, Some, Marked

0
1
2
3
4
5
6
7

N − S  S − M  N − S  S − M

: Treatment Placebo

30 40 50 60 70

N − S  S − M  N − S  S − M

: Treatment Treated

25 / 64

Nested dichotomies
Basic ideas

Polytomous response: Nested dichotomies

- $m$ categories $\rightarrow (m - 1)$ comparisons (logits)
- If these are formulated as $(m - 1)$ nested dichotomies:
  - Each dichotomy can be fit using the familiar binary-response logistic model,
  - the $m - 1$ models will be statistically independent ($G^2$ statistics will be additive)
  - (Need some extra work to summarize these as a single, combined model)
- This allows the slopes to differ for each logit

G^2_{all} = \sum_{i=1}^{m-1} G^2(L_i) \quad df_{all} = \sum df(L_i)

Example: Women’s Labour-Force Participation

Data: Social Change in Canada Project, York ISR (Fox, 1997)

- Response: not working outside the home ($n=155$), working part-time ($n=42$) or working full-time ($n=66$)
- Model as two nested dichotomies:
  - Working ($n=106$) vs. NotWorking ($n=155$)
  - Working full-time ($n=66$) vs. working part-time ($n=42$).
- Predictors:
  - Children? — 1 or more minor-aged children
  - Husband’s Income — in $1000s
  - Region of Canada (not considered here)
Example: Women's Labour-Force Participation

```sas
proc format;
  value labour /* labour-force participation */
    1 = 'working full-time'
    2 = 'working part-time'
    3 = 'not working';
  value kids /* children in the household */
    0 = 'Children absent'
    1 = 'Children present';
data wlfpart;
  input case labour husinc children region;
  working = labour < 3;
  if working then
    fulltime = (labour = 1);
  datalines;
  1 3 15 1 3
  2 3 13 1 3
  3 3 45 1 3
  4 3 23 1 3
  5 3 19 1 3
  6 3 7 1 3
  7 3 15 1 3
  8 1 7 1 3
  9 3 15 1 3
  ... more data lines ...
```

First, try proportional odds model for labour

```sas
proc logistic data=wlfpart;
  model labour = husinc children;
  title2 'Proportional Odds Model: Fulltime/Parttime/NotWorking';
```

The score test rejects the 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>18.5638</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
```

This indicates that the slopes differ for at least one of husinc and children. Note: The score test is known to be overly sensitive. Use a more stringent \( \alpha \) to reject.

Fit separate models for each of working and fulltime:

```sas
proc logistic data=wlfpart nosimple descending;
  model working = husinc children;
  output out=resultw p=predict xbeta=logit;
  title2 'Nested Dichotomies';
```

```sas
proc logistic data=wlfpart nosimple descending;
  model fulltime = husinc children;
  output out=resultf p=predict xbeta=logit;
```

descending option used to model the \( \Pr(Y = 1) \) – working, or fulltime

output statements → datasets for plotting

Join for plotting:

```sas
data both;
  set resultw resultsf;
  ... more data lines ...
```

Output for WORKING dichotomy:

```
Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Wald</th>
<th>Pr &gt; Chi-Square</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCPT</td>
<td>1</td>
<td>1.3358</td>
<td>0.3838</td>
<td>12.1165</td>
<td>0.0005</td>
<td>1.021</td>
</tr>
<tr>
<td>HUSINC</td>
<td>1</td>
<td>-0.0423</td>
<td>0.0198</td>
<td>4.5751</td>
<td>0.0324</td>
<td>0.959</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>1</td>
<td>-1.5756</td>
<td>0.2923</td>
<td>29.0651</td>
<td>0.0001</td>
<td>0.197</td>
</tr>
</tbody>
</table>
```

Output for FULLTIME dichotomy:

```
Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Wald</th>
<th>Pr &gt; Chi-Square</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCPT</td>
<td>1</td>
<td>3.4778</td>
<td>0.7671</td>
<td>20.5537</td>
<td>0.0001</td>
<td>1.429</td>
</tr>
<tr>
<td>HUSINC</td>
<td>1</td>
<td>-0.1073</td>
<td>0.0392</td>
<td>7.5063</td>
<td>0.0061</td>
<td>0.898</td>
</tr>
<tr>
<td>CHILDREN</td>
<td>1</td>
<td>-2.6515</td>
<td>0.5411</td>
<td>24.0135</td>
<td>0.0001</td>
<td>0.071</td>
</tr>
</tbody>
</table>
```

\[
\log \left( \frac{\Pr(working)}{\Pr(not working)} \right) = 1.336 - 0.042 \text{ H$} - 1.576 \text{ kids}
\]

\[
\log \left( \frac{\Pr(fulltime)}{\Pr(parttime)} \right) = 3.478 - 0.107 \text{ H$} - 2.652 \text{ kids}
\]
Combined tests for Nested Dichotomies

- Nested dichotomies → $\chi^2$ tests and df for the separate logits are independent
- → add, to give tests for the full $m$-level response (manually)

Global tests of BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Response</th>
<th>ChiSq</th>
<th>DF</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>working</td>
<td>36.4184</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>fulltime</td>
<td>39.8468</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>76.2652</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Wald tests:

Wald tests of maximum likelihood estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Response</th>
<th>WaldChiSq</th>
<th>DF</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>working</td>
<td>12.1164</td>
<td>1</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>fulltime</td>
<td>20.5536</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>52.6700</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>children</td>
<td>working</td>
<td>29.0650</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>fulltime</td>
<td>24.0134</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>53.0784</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>husinc</td>
<td>working</td>
<td>4.5750</td>
<td>1</td>
<td>0.0324</td>
</tr>
<tr>
<td></td>
<td>fulltime</td>
<td>7.5062</td>
<td>1</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>ALL</td>
<td>12.0813</td>
<td>2</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

Model visualization

- Join output datasets (resultsw and resultsf)
- Combine Response & Children → event
- plot logit * husinc = event; → separate lines

Model visualization in SAS

```sas
proc gplot data=both;
plot logit * husinc = event / anno=lbl nolegend frame vaxis=axis1;
axis1 label=(a=90 'Log Odds') order=(-5 to 4);
title2 'Working vs Not Working and Fulltime vs. Parttime';
symbol1 v=dot h=1.5 i=join l=3 c=red;
symbol2 v=dot h=1.5 i=join l=1 c=black;
symbol3 v=circle h=1.5 i=join l=3 c=red;
symbol4 v=circle h=1.5 i=join l=1 c=black;
```

Women's labor-force participation, Canada 1977

Working vs Not Working and Fulltime vs. Parttime

No Children

with Children

Working

Fulltime

Working

Fulltime

with Children

No Children

Husband's Income

0 10 20 30 40 50

Log Odds

-5

-4

-3

-2

-1

0

1

2

3

4

-5

-4

-3

-2

-1

0

1

2

3

4

-5

-4

-3

-2

-1

0

1

2

3

4

-5

-4

-3

-2

-1

0

1

2

3

4

-5

-4

-3

-2

-1

0

1

2

3

4

0 10 20 30 40 50

Husband's Income

0 10 20 30 40 50

logit * husinc = event;
Nested dichotomies in R

In R, the steps are similar—first create new variables, working and fulltime, using the `recode()` function in the `car` package:

```r
> library(car)  # for data and Anova()
> data(Womenlf)
> Womenlf <- within(Womenlf,
+   working <- recode(partic, 
+                     "'not.work' = 'no'; else = 'yes' ")
+   fulltime <- recode (partic,
+                     "'fulltime' = 'yes'; 'parttime' = 'no'; 'not.work' = NA")
)>
```

```
partic hincome children region fulltime working
31 not.work 13 present Ontario <NA> no
34 not.work 9 absent Ontario <NA> no
55 parttime 9 present Atlantic no yes
86 fulltime 27 absent BC yes yes
96 not.work 17 present Ontario <NA> no
141 not.work 14 present Ontario <NA> no
189 fulltime 9 present Atlantic yes yes
234 fulltime 5 absent Quebec yes yes
240 not.work 13 present Quebec <NA> no
```

Nested dichotomies in R: fitting

Then, fit models for each dichotomy:

```r
> contrasts(children)
> contrasts(children) <- 'contr.treatment'
> mod.working <- glm(working ~ hincome + children, family=binomial, data=Womenlf)
> mod.fulltime <- glm(fulltime ~ hincome + children, family=binomial, data=Womenlf)
```

Some output from `summary(mod.working)`:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) 1.33583   0.38376  3.481 0.0005 ***
hincome    -0.04231   0.01978 -2.139 0.0324 *
childrenpresent -1.57565   0.29226 -5.391 7e-08 ***
```

Some output from `summary(mod.fulltime)`:

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)  
(Intercept) 3.47777   0.76711  4.534 5.80e-06 ***
hincome    -0.10727   0.03915 -2.740 0.00615 **
childrenpresent -2.65146   0.54108 -4.900 9.57e-07 ***
```

Nested dichotomies in R: plotting

For plotting, we need to calculate the predicted probabilities (or logits) over a grid of combinations of the predictors in each sub-model, using the `predict()` function.

```r
> pred <- expand.grid(hincome=1:45, children=c('absent', 'present'))
> # get fitted values for both sub-models
> p.work <- predict(mod.working, pred, type='response')
> p.fulltime <- predict(mod.fulltime, pred, type='response')
```

The fitted value for the fulltime dichotomy is **conditional** on working outside the home; multiplying by the probability of working gives the **unconditional** probability.

```r
> p.full <- p.work * p.fulltime
> p.part <- p.work * (1 - p.fulltime)
> p.not <- 1 - p.work
```
Polytomous response: Generalized Logits

- Models the probabilities of the \( m \) response categories as \( m - 1 \) logits comparing each of the first \( m - 1 \) categories to the last (reference) category.
- Logits for any pair of categories can be calculated from the \( m - 1 \) fitted ones.
- With \( k \) predictors, \( x_1, x_2, \ldots, x_k \), for \( j = 1, 2, \ldots, m - 1 \),
  \[ L_{jm} \equiv \log \left( \frac{\pi_{ij}}{\pi_{im}} \right) = \beta_{0j} + \beta_{1j} x_1 + \beta_{2j} x_2 + \cdots + \beta_{kj} x_k \]
  \[ = \beta_j^T x_i \]
- One set of fitted coefficients, \( \beta_j \) for each response category except the last.
- Each coefficient, \( \beta_{hj} \), gives the effect on the log odds of a unit change in the predictor \( x_h \) that an observation belongs to category \( j \) vs. category \( m \).
- Probabilities in response categories are calculated as:
  \[ \pi_{ij} = \frac{\exp(\beta_j^T x_i)}{\sum_{j=1}^{m-1} \exp(\beta_j^T x_i)}, \ j = 1, \ldots, m - 1; \quad \pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij} \]

Example: Women's Labour Force Participation

Graphs:

```
wlfpart5.sas ···
1 title 'Generalized logit model';
2 proc logistic data=wlfpart;
3   model labor = husinc children / link=glogit;
4   output out=results p=predict xbeta=logit;
```

```
Response profile:
<table>
<thead>
<tr>
<th>Ordered</th>
<th>labor</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td></td>
<td>Frequency</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>155</td>
</tr>
</tbody>
</table>

Logits modeled use labor=3 as the reference category.
```

Note: Not working is the baseline category
Overall and Type III tests:

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>77.6106</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>76.4850</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>58.4351</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Type III Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>husinc</td>
<td>2</td>
<td>12.8159</td>
<td>0.0016</td>
</tr>
<tr>
<td>children</td>
<td>2</td>
<td>53.9806</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

These are comparable to the combined tests for the nested dichotomies models.

Coefficients:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>labor</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>1</td>
<td>1</td>
<td>1.9828</td>
<td>0.4842</td>
<td>16.7709</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>1</td>
<td>-1.4323</td>
<td>0.5925</td>
<td>5.8445</td>
<td>0.0156</td>
</tr>
<tr>
<td>husinc</td>
<td>1</td>
<td>1</td>
<td>-0.0972</td>
<td>0.0281</td>
<td>11.9762</td>
<td>0.0005</td>
</tr>
<tr>
<td>children</td>
<td>2</td>
<td>1</td>
<td>-2.5586</td>
<td>0.3622</td>
<td>49.9008</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>children</td>
<td>1</td>
<td>1</td>
<td>0.0215</td>
<td>0.4690</td>
<td>0.0021</td>
<td>0.9635</td>
</tr>
</tbody>
</table>

i.e., the fitted models are:

\[
\log \left( \frac{\Pr(\text{fulltime})}{\Pr(\text{not working})} \right) = 1.983 - 0.097 H$ - 2.56 \text{ kids}
\]

\[
\log \left( \frac{\Pr(\text{parttime})}{\Pr(\text{not working})} \right) = -1.432 - 0.0069 H$ - 0.0215 \text{ kids}
\]

**Interpretation:** Signs for husinc and children are understandable, but need to make a plot!

Example: Women's Labour Force Participation

```sas
proc sort data=results; by children husinc _level_; *-- Curve labels; %label(data=results, x=husinc, y=predict, cvar=_level_, by=children, subset=last._level_, text=put(_level_, labor.), pos=2, size=2, out=labels1);
%label(data=results, x=20, y=0.85, by=children, subset=last.children, text=put(children, kids.), pos=2, out=labels2);
data labels;
set labels1 labels2; by children;
goptions hby=0;
proc gplot data=results; plot predict * husinc = _level_ / axis1 order=(0 to .9 by .1) anno=labels nolegend;
    by children;
    axis1 order=(0 to .9 by .1) label=(a=90);
    symbol i=join v=circle c=black;
    symbol2 i=join v=square c=red;
    symbol3 i=join v=triangle c=blue;
run;
```

- **logit** gives the two fitted log odds vs Not working
- **predict** gives the predicted probability for each category of labor
Generalized logit models

Example: Women’s Labour Force Participation

Graphs:

- Children absent
- Children present

Husband’s Income

<table>
<thead>
<tr>
<th>Fitted probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
</tr>
</tbody>
</table>

- Full-time
- Part-time
- Not working

Husband’s Income

| 0 10 20 30 40 |

Generalized logit models in R

Generalized logit models in R: Fitting

- In R, the generalized logit model can be fit using the `multinom()` function in the `nnet` package.
- For interpretation, it is useful to reorder the levels of partic so that `not.work` is the baseline level.

```r
Womenlf$partic <- ordered(Womenlf$partic, levels=c('not.work', 'parttime', 'fulltime'))
library(nnet)
mod.multinom <- multinom(partic ~ hincome + children, data=Womenlf)
summary(mod.multinom, Wald=TRUE)
Anova(mod.multinom)
```

The `Anova()` tests are similar to what we got from summing these tests from the two nested dichotomies:

<table>
<thead>
<tr>
<th>LR Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hincome</td>
<td>15.2</td>
<td>2</td>
</tr>
<tr>
<td>children</td>
<td>63.6</td>
<td>2</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ' ***' 0.001 ' **' 0.01 ' *' 0.05 ' . ' 0.1 ' ' 1

Generalized logit models in R: Plotting

- As before, it is much easier to interpret a model from a plot than from coefficients, but this is particularly true for polytomous response models.
- `style="stacked"` shows cumulative probabilities.

```
library(effects)
plot(effect("hincome*children", mod.multinom), style="stacked")
```

You can also view the effects of husband’s income and children separately in this main effects model with `plot(allEffects)`.

```
plot(allEffects(mod.multinom), ask=FALSE)
```

Generalized logit models in R: Plotting

- You can also view the effects of husband’s income and children separately in this main effects model with `plot(allEffects(mod.multinom), ask=FALSE)`.

```
plot(allEffects(mod.multinom), ask=FALSE)
```

 hincome effect plot

<table>
<thead>
<tr>
<th>partic (probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
</tr>
</tbody>
</table>

 children effect plot

<table>
<thead>
<tr>
<th>partic (probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.6</td>
</tr>
</tbody>
</table>

hincome

| 0 10 20 30 40 |

children

| absent | present | 0 10 20 30 40 |

Political knowledge & party choice in Britain

Example from Fox and Andersen (2006): Data from 1997 British Election Panel Survey (BEPS)

- **Response**: Party choice—Liberal democrat, Labour, Conservative
- **Predictors**
  - Europe: 11-point scale of attitude toward European integration (high="Eurosceptic")
  - Political knowledge: knowledge of party platforms on European integration ("low"=0–3="high")
  - Others: Age, Gender, perception of economic conditions, evaluation of party leaders (Blair, Hague, Kennedy)—1.5 scale
- **Model**
  - Main effects of Age, Gender, economic conditions (national, household)
  - Main effects of evaluation of party leaders
  - Interaction of attitude toward European integration with political knowledge

**BEPS data: Fitting**

In R, generalized (multinomial) response models are fit using `multinom()` in the `nnet` package

```
library(effects)  # data, plots
library(car)      # for Anova()
library(nnet)     # for multinom()

multinom.mod <- multinom(vote ~ age + gender + economic.cond.national +
                         economic.cond.household + Blair + Hague + Kennedy +
                         Europe*political.knowledge, data=BEPS)

Anova(multinom.mod)
```

Anova Table (Type II tests)

<table>
<thead>
<tr>
<th></th>
<th>LR Chisq</th>
<th>Df</th>
<th>Pr(&gt;Chisq)</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>13.9</td>
<td>2</td>
<td>0.00097 ***</td>
</tr>
<tr>
<td>gender</td>
<td>0.5</td>
<td>2</td>
<td>0.79726</td>
</tr>
<tr>
<td>economic.cond.national</td>
<td>30.6</td>
<td>2</td>
<td>2.3e-07 ***</td>
</tr>
<tr>
<td>economic.cond.household</td>
<td>5.7</td>
<td>2</td>
<td>0.05926 .</td>
</tr>
<tr>
<td>Blair</td>
<td>135.4</td>
<td>2</td>
<td>&lt; 2e-16 ***</td>
</tr>
<tr>
<td>Hague</td>
<td>166.8</td>
<td>2</td>
<td>&lt; 2e-16 ***</td>
</tr>
<tr>
<td>Kennedy</td>
<td>68.9</td>
<td>2</td>
<td>1.1e-15 ***</td>
</tr>
<tr>
<td>Europe</td>
<td>78.0</td>
<td>2</td>
<td>&lt; 2e-16 ***</td>
</tr>
<tr>
<td>political.knowledge</td>
<td>55.6</td>
<td>2</td>
<td>8.6e-13 ***</td>
</tr>
<tr>
<td>Europe:political.knowledge</td>
<td>50.8</td>
<td>2</td>
<td>9.3e-12 ***</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**BEPS data: Interpretation?**

How to understand the **nature** of these effects on party choice?

```
> summary(multinom.mod)
```

Call:
`multinom(formula = vote ~ age + gender + economic.cond.national +
         economic.cond.household + Blair + Hague + Kennedy + Europe *
         political.knowledge, data = BEPS)`

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>age</th>
<th>gender</th>
<th>male</th>
<th>economic.cond.national</th>
<th>Liberal Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour</td>
<td>-0.8734</td>
<td>-0.01980</td>
<td>0.1126</td>
<td>0.5220</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberal Democrat</td>
<td>-0.7185</td>
<td>-0.01460</td>
<td>0.0914</td>
<td>0.1451</td>
<td></td>
<td></td>
</tr>
<tr>
<td>economic.cond.household</td>
<td>0.178632</td>
<td>0.8236</td>
<td>-0.8684</td>
<td>0.2396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blair</td>
<td>0.178632</td>
<td>0.8236</td>
<td>-0.8684</td>
<td>0.2396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hague</td>
<td>0.007725</td>
<td>0.2779</td>
<td>-0.7808</td>
<td>0.6557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kennedy</td>
<td>0.007725</td>
<td>0.2779</td>
<td>-0.7808</td>
<td>0.6557</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European knowledge</td>
<td>0.6683</td>
<td>-0.1829</td>
<td>1.1602</td>
<td>-0.1589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberal Democrat</td>
<td>0.6683</td>
<td>-0.1829</td>
<td>1.1602</td>
<td>-0.1589</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Std. Errors:

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>age</th>
<th>gender</th>
<th>male</th>
<th>economic.cond.national</th>
<th>Liberal Democrat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour</td>
<td>0.6908</td>
<td>0.005364</td>
<td>0.1694</td>
<td>0.1065</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberal Democrat</td>
<td>0.7394</td>
<td>0.006564</td>
<td>0.1780</td>
<td>0.1100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Residual Deviance: 2233
AIC: 2277
Generalized logit models

**BEPS data: Effect plots to the rescue!**

Age effect: Older more likely to vote Conservative

![BEPS data: effect of Age](image)

Attitude toward European integration \( \times \) political knowledge effect:

- Low political knowledge: little relation between attitude and political choice
- As knowledge increases: more Eurosceptic views more likely to support Conservatives
- \( \Rightarrow \) detailed understanding of complex models depends strongly on visualization!

Summary: Part 5

**Polytomous responses**
- \( m \) response categories \( \rightarrow \) \( m - 1 \) comparisons (logits)
- Different models for ordered vs. unordered categories

**Proportional odds model**
- Simplest approach for ordered categories: Same slopes for all logits
- Requires proportional odds assumption to be met
- SAS: PROC LOGISTIC; R: `polr()`

**Nested dichotomies**
- Applies to ordered or unordered categories
- Fit \( m - 1 \) independent models \( \rightarrow \) Additive \( \chi^2 \) values
- SAS: PROC LOGISTIC; R: `glm()`

**Generalized (multinomial) logistic regression**
- Fit \( m - 1 \) logits as a single model
- Results usually comparable to nested dichotomies
- SAS: PROC LOGISTIC, LINK=GLOGIT; R: `(multinom)`

Visualizing Categorical Data: What we've learned

- **Categorical data**
  - Table form vs. case form
  - Non-parametric methods vs. model-based methods
  - Response models vs. association models

- **Graphical methods for categorical data**
  - Frequency data more naturally displayed as count \( \sim \) area
  - Sieve diagram, fourfold & mosaic display: compare observed vs. expected
  - Discrete response data benefits from: smoothing, effect plots
  - Graphical principles: Visual comparisons, effect ordering, small multiples

- **Theory into practice**
  - To be useful, statistical methods must be:
    - available— implemented in standard software
    - accessible— easy to use (or at least easier)
    - VCD provides \( \sim 40 \) general macros and SAS/IML programs
    - The `vcd` package for R does the same for R users.
    - Effective statistical graphics is still hard work— 80/20 rule
References I


References II


References III


References IV


