Logistic Regression II

Michael Friendly

Psych 6136

November 9, 2017
Donner Party: A graphic tale of survival & influence

History:
- Apr–May, 1846: Donner/Reed families set out from Springfield, IL to CA
- Jul: Bridger’s Fort, WY, 87 people, 23 wagons
Donner Party: A graphic tale of survival & influence

History:

- “Hasting’s Cutoff”, untried route through Salt Lake Desert, Wasatch Mtns. (90 people)
- Worst recorded winter: Oct 31 blizzard—Missed by 1 day, stranded at “Truckee Lake” (now Donner’s Lake, Reno)
  - Rescue parties sent out (“Dire necessity”, “Forelorn hope”, ...)
  - Relief parties from CA: 42 survivors (Mar–Apr, ’47)
Donner Party: Data

data("Donner", package="vcdExtra")
Donner$survived <- factor(Donner$survived, labels=c("no", "yes"))

library(car)
some(Donner, 12)

## family age sex survived death
## Breen, Peter  Breen  3 Male yes     <NA>
## Donner, George Donner 62 Male no 1847-03-18
## Donner, Jacob Donner 65 Male no 1846-12-21
## Foster, Jeremiah MurFosPik 1 Male no 1847-03-13
## Graves, Jonathan Graves 7 Male yes <NA>
## Graves, Mary Ann Graves 20 Female yes <NA>
## Graves, Nancy Graves 9 Female yes <NA>
## McCutchen, Harriet McCutchen 1 Female no 1847-02-02
## Reed, James Reed 46 Male yes <NA>
## Reed, Thomas Keyes Reed 4 Male yes <NA>
## Reinhardt, Joseph Other 30 Male no 1846-12-21
## Wolfinger, Doris FosdWolf 20 Female yes <NA>
Overview: a `gpairs()` plot

- Binary response: `survived`
- Categorical predictors: `sex`, `family`
- Quantitative predictor: `age`
- Q: Is the effect of age linear?
- Q: Are there interactions among predictors?
- This is a generalized pairs plot, with different plots for each pair
Survival decreases with age for both men and women.
- Women more likely to survive, particularly the young.
- Data is thin at older ages.
Using ggplot2

Basic plot: survived vs. age, colored by sex, with jittered points

```r
gg <- ggplot(Donner, 
  aes(age, as.numeric(survived=="yes"), color = sex)) + 
ylab("Survived") + 
geom_point(position = position_jitter(height = 0.02, width = 0))
```

Add conditional linear logistic regressions with
`stat_smooth(method="glm")`

```r
gg + stat_smooth(method = "glm", family = binomial, formula = y ~ x, 
  alpha = 0.2, size=2, aes(fill = sex))
```
Questions

- Is the relation of survival to age well expressed as a linear logistic regression model?
  - Allow a quadratic or higher power, using `poly(age, 2), poly(age, 3),`
    
    \[
    \text{logit}(\pi_i) = \alpha + \beta_1 x_i + \beta_2 x_i^2 \\
    \text{logit}(\pi_i) = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3
    \]
    
    \[
    \ldots
    \]
  - Use *natural spline* functions, `ns(age, df)`
  - Use non-parametric smooths, `loess(age, span, degree)`

- Is the relation the same for men and women? i.e., do we need an interaction of age and sex?
  - Allow an interaction of `sex * age` or `sex * f(age)`
  - Test goodness-of-fit relative to the main effects model
Fit separate quadratics for males and females
Fit separate loess smooths for males and females
Fitting models

Models with linear effect of age:

donner.mod1 <- glm(survived ~ age + sex,
data=Donner, family=binomial)
donner.mod2 <- glm(survived ~ age * sex,
data=Donner, family=binomial)

Anova(donner.mod2)

## Analysis of Deviance Table (Type II tests)
##
## Response: survived
## LR Chisq Df Pr(>Chisq)
## age 5.52 1 0.0188 *
## sex 6.73 1 0.0095 **
## age:sex 0.40 1 0.5269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Fitting models

Models with quadratic effect of age:

doner.mod3 <- glm(survived ~ poly(age, 2) + sex, 
data=Donner, family=binomial)
doner.mod4 <- glm(survived ~ poly(age, 2) * sex, 
data=Donner, family=binomial)
Anova(doner.mod4)

## Analysis of Deviance Table (Type II tests)
##
## Response: survived
##
## LR Chisq Df Pr(>Chisq)
## poly(age, 2) 9.91 2 0.0070 **
## sex 8.09 1 0.0044 **
## poly(age, 2):sex 8.93 2 0.0115 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
library(vcdExtra)
LRstats(donner.mod1, donner.mod2, donner.mod3, donner.mod4)

## Likelihood summary table:
## AIC  BIC  LR  Chisq  Df  Pr(>Chisq)
## donner.mod1 117 125 111.1  87 0.042  *
## donner.mod2 119 129 110.7  86 0.038  *
## donner.mod3 115 125 106.7  86 0.064  .
## donner.mod4 110 125  97.8  84 0.144  
## ---
## Signif. codes:  0 '***'  0.001 '**'  0.01 '*'  0.05 '.'  0.1 ' ' 1

<table>
<thead>
<tr>
<th></th>
<th>linear</th>
<th>non-linear</th>
<th>Δχ²</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>additive</td>
<td>111.128</td>
<td>106.731</td>
<td>4.396</td>
<td>0.036</td>
</tr>
<tr>
<td>non-additive</td>
<td>110.727</td>
<td>97.799</td>
<td>12.928</td>
<td>0.000</td>
</tr>
<tr>
<td>Δχ²</td>
<td>0.400</td>
<td>8.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.527</td>
<td>0.003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Who was influential?

```r
library(car)
res <- influencePlot(donner.mod3, id.col="blue", scale=8, id.n=2)
```
Why are they influential?

```r
idx <- which(rownames(Donner) %in% rownames(res))
# show data together with diagnostics
cbind(Donner[idx, 2:4], res)
```

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Sex</th>
<th>Survived</th>
<th>StudRes</th>
<th>Hat</th>
<th>CookD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breen, Patrick</td>
<td>51</td>
<td>Male</td>
<td>yes</td>
<td>2.501</td>
<td>0.09148</td>
<td>0.32354</td>
</tr>
<tr>
<td>Donner, Elizabeth</td>
<td>45</td>
<td>Female</td>
<td>no</td>
<td>-1.114</td>
<td>0.13541</td>
<td>0.03409</td>
</tr>
<tr>
<td>Graves, Elizabeth C.</td>
<td>47</td>
<td>Female</td>
<td>no</td>
<td>-1.019</td>
<td>0.16322</td>
<td>0.03418</td>
</tr>
<tr>
<td>Reed, James</td>
<td>46</td>
<td>Male</td>
<td>yes</td>
<td>2.098</td>
<td>0.08162</td>
<td>0.14364</td>
</tr>
</tbody>
</table>

- Patrick Breen, James Reed: Older men who survived
- Elizabeth Donner, Elizabeth Graves: Older women who died

Moral lessons of this story:
- Don’t try to cross the Donner Pass in late October; if you do, bring lots of food
- Plots of fitted models show *only* what is included in the model
- Discrete data often need smoothing (or non-linear terms) to see the pattern
- Always examine model diagnostics — preferably graphic
Polytomous responses: Overview

When response categories are:

- Unordered

for example:
Ford
Smitherman
Pantelone

the analysis can use:

- Multinomial logistic regression

we model these logits:

\[
\begin{align*}
\text{None} & \quad \text{Some or marked} \\
\text{None or Some} & \quad \text{Marked}
\end{align*}
\]

Ordered

No improvement
Some improvement
Marked improvement

Proportional odds model

Nested dichotomies

\[
\begin{align*}
\text{None} & \quad \text{Some or marked} \\
\text{Some} & \quad \text{Marked}
\end{align*}
\]
Polytomous responses: Overview

- $m$ categories $\rightarrow (m - 1)$ comparisons (logits)
  - One part of the model for each logit
  - Similar to ANOVA where an $m$-level factor $\rightarrow (m - 1)$ contrasts (df)

- **Response categories** unordered, e.g., vote NDP, Liberal, Green, Tory
  - Multinomial logistic regression
    - Fits $m - 1$ logistic models for logits of category $i = 1, 2, \ldots m - 1$ vs. category $m$
      - Example:
        - NDP vs. Tory
        - Liberal vs. Tory
        - Green vs. Tory
  - This is the most general approach
  - R: `multinom()` function in `nnet`

- Can also use nested dichotomies
Polytomous responses: Overview

- **Response categories** *ordered*, e.g., None, Some, Marked improvement
  - Proportional odds model
    - Uses adjacent-category logits
    - Assumes slopes are *equal* for all *m* − 1 logits; only intercepts vary
    - R: `polr()` in MASS
  - Nested dichotomies
    - Model each logit separately
    - $G^2$ s are additive $\rightarrow$ combined model
Fitting and graphing: Overview

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
Ordinal response: Proportional odds model

Arthritis treatment data:

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

- Model logits for adjacent category cutpoints:

  \[
  \text{logit} (\theta_{ij1}) = \log \frac{\pi_{ij1}}{\pi_{ij2} + \pi_{ij3}} = \text{logit ( None vs. [Some or Marked] )}
  \]

  \[
  \text{logit} (\theta_{ij2}) = \log \frac{\pi_{ij1} + \pi_{ij2}}{\pi_{ij3}} = \text{logit ( [None or Some] vs. Marked )}
  \]
Consider a logistic regression model for each logit:

\[ \text{logit}(\theta_{ij1}) = \alpha_1 + \mathbf{x}_{ij}' \beta_1 \quad \text{None vs. Some/Marked} \]

\[ \text{logit}(\theta_{ij2}) = \alpha_2 + \mathbf{x}_{ij}' \beta_2 \quad \text{None/Some vs. Marked} \]

Proportional odds assumption: regression functions are parallel on the logit scale i.e., \( \beta_1 = \beta_2 \).
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 showing thresholds and response categories]
We can visualize the relation of the latent variable $\xi$ to the observed response $Y$, for two values, $x_1$ and $x_2$, of a single predictor, $X$ as shown below:
Proportional odds: Latent variable interpretation

For the Arthritis data, the relation of improvement to age is shown below (using the effects package)

![Graph showing the relation of improvement to age for the Arthritis data](image-url)
Proportional odds models in R

- **Fitting:** `polr()` in MASS package

The response, `Improved` has been defined as an *ordered* factor

```r
data(Arthritis, package="vcd")
head(Arthritis$Improved)
```

```text
## [1] Some None None Marked Marked Marked
## Levels: None < Some < Marked
```

Fitting:

```r
library(MASS)  # for polr()
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>-1.25168</td>
<td>0.54636</td>
<td>-2.2909</td>
</tr>
<tr>
<td>TreatmentTreated</td>
<td>1.74529</td>
<td>0.47589</td>
<td>3.6674</td>
</tr>
<tr>
<td>Age</td>
<td>0.03816</td>
<td>0.01842</td>
<td>2.0722</td>
</tr>
</tbody>
</table>

Intercepts:

<table>
<thead>
<tr>
<th></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
The `car::Anova()` function gives hypothesis tests for model terms:

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

Anova Table (Type II tests)

Response: Improved

<table>
<thead>
<tr>
<th></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>0.0170812 *</td>
</tr>
<tr>
<td>Treatment</td>
<td>14.7095</td>
<td>1</td>
<td>0.0001254 ***</td>
</tr>
<tr>
<td>Age</td>
<td>4.5715</td>
<td>1</td>
<td>0.0325081 *</td>
</tr>
</tbody>
</table>

---

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

- `anova()` gives Type I (sequential) tests — not usually useful
- Type II (partial) tests control for the effects of all other terms
Testing the proportional odds assumption

- The PO model is valid only when the slopes are equal for all predictors
- This can be tested by comparing this model to the generalized logit NPO model

\[
\text{PO : } L_j = \alpha_j + \mathbf{x}^T \beta_j \quad j = 1, \ldots, m - 1
\]

\[
\text{NPO : } L_j = \alpha_j + \mathbf{x}^T \beta \quad j = 1, \ldots, m - 1
\]

- A likelihood ratio test requires fitting both models calculating \( \Delta G^2 = G^2_{\text{NPO}} - G^2_{\text{PO}} \) with \( p \) df.
- This can be done using \texttt{vglm()} in the \texttt{VGAM} package
- The \texttt{rms} package provides a visual assessment, plotting the conditional mean \( E(X \mid Y) \) of a given predictor, \( X \), at each level of the ordered response \( Y \).
- If the response behaves ordinally in relation to \( X \), these means should be strictly increasing or decreasing with \( Y \).
Testing the proportional odds assumption

In **VGAM**, the PO model is fit using `family = cumulative(parallel=TRUE)`

```r
library(VGAM)
arth.po <- vglm(Improved ~ Sex + Treatment + Age, data=Arthritis, 
family = cumulative(parallel=TRUE))
```

The more general NPO model can be fit using `parallel=FALSE`.

```r
arth.npo <- vglm(Improved ~ Sex + Treatment + Age, data=Arthritis, 
family = cumulative(parallel=FALSE))
```

The LR test says the PO model is OK:

```r
VGAM::lrtest(arth.npo, arth.po)
```

```
## Likelihood ratio test
##
## Model 1: Improved ~ Sex + Treatment + Age
## Model 2: Improved ~ Sex + Treatment + Age
## #Df LogLik Df Chisq Pr(>Chisq)
## 1 160 -71.8
## 2 163 -72.7 3 1.88 0.6
```
Full-model plot of predicted probabilities:

- **Intercept1**: [Marked, Some] vs. [None]
- **Intercept2**: [Marked] vs. [Some, None]
- On logit scale, these would be parallel lines
- Effects of age, treatment, sex similar to what we saw before
Proportional odds models in R: Plotting

- **Plotting**: `plot(effect())` in `effects` package

```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
Proportional odds models in R: Plotting

Making visual comparisons easier:

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

![Treatment*Age effect plot](Image)
Proportional odds models in R: Plotting

Making visual comparisons easier:

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

![Sex*Age effect plot](image)
Proportional odds models in R: Plotting

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

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

![Treatment*Age effect plot](image)
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

\[
G^2_{all} = \sum_{i} G^2(L_i) \quad df_{all} = \sum df(L_i)
\]
Nested dichotomies: Examples

$m = 3$

Arthritis treatment

None \rightarrow Some or marked

Some \rightarrow Marked

$L_1 = \log \frac{\pi_1}{\pi_2 + \pi_3}$

$L_2 = \log \frac{\pi_2}{\pi_3}$

$m = 4$

Psychiatric diagnosis

Normal, Manic, Depressed, Schiz

Manic, Depressed \rightarrow Schiz

Manic, Depressed

$L_1 = \log \frac{\pi_1}{\pi_2 + \pi_3 + \pi_4}$

$L_2 = \log \frac{\pi_4}{\pi_2 + \pi_3}$

$L_3 = \log \frac{\pi_2}{\pi_3}$
Example: Women’s Labour-Force Participation

Data: *Social Change in Canada Project*, York ISR, car::Womenlf data

- **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).

\[ L_1: \begin{array}{c}
\text{not working} \\
\text{part-time, full-time}
\end{array} \]

\[ L_2: \begin{array}{c}
\text{part-time} \\
\text{full-time}
\end{array} \]

- **Predictors:**
  - Children? — 1 or more minor-aged children
  - Husband’s Income — in $1000s
  - Region of Canada (not considered here)
Nested dichotomies: Combined tests

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

<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 for each coefficient:

<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>32.6700</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>
Nested dichotomies: recoding

In R, first create new variables, working and fulltime, using the recode() function in the car:

```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")
)>
> some(Womenlf)
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
180 fulltime 13 absent BC yes yes
189 fulltime 9 present Atlantic yes yes
234 fulltime 5 absent Quebec yes yes
240 not.work 13 present Quebec <NA> no
```
Then, fit models for each dichotomy:

```r
> 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: interpretation

Write out the predictions for the two logits, and compare coefficients:

\[
\log \left( \frac{\Pr(\text{working})}{\Pr(\text{not working})} \right) = 1.336 - 0.042 \ H$ - 1.576 \ \text{kids}
\]
\[
\log \left( \frac{\Pr(\text{fulltime})}{\Pr(\text{parttime})} \right) = 3.478 - 0.107 \ H$ - 2.652 \ \text{kids}
\]

Better yet, plot the predicted log odds for these equations:
Nested dichotomies: plotting

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

- **type='response'** gives these on the probability scale, whereas
- **type='link'** (the default) gives these on the logit scale.

```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
```
Nested dichotomies in R: plotting

The plot below was produced using the basic R functions `plot()`, `lines()` and `legend()`. See the file `wlf-nested.R` on the course web page for details.
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 $j = 1, 2, \ldots, m - 1$,

$$L_{jm} \equiv \log \left( \frac{\pi_{ij}}{\pi_{im}} \right) = \beta_{0j} + \beta_{1j} x_{i1} + \beta_{2j} x_{i2} + \cdots + \beta_{kj} x_{ik}$$

$$= \beta_j^T x_i$$

- 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:

$$\pi_{ij} = \frac{\exp(\beta_j^T x_i)}{\sum_{j=1}^{m-1} \exp(\beta_j^T x_i)} , \ j = 1, \ldots, m - 1 ; \quad \pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij}$$
Generalized logit models: Fitting

- In R, the generalized logit model can be fit using the `multinom()` function in the `nnet`
- 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:

```
Analysis of Deviance Table (Type II tests)

Response: partic
          LR Chisq Df Pr(>Chisq)
hincome  15.2  2 0.00051  ***
children 63.6  2 1.6e-14  ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
Generalized logit models: 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

```r
library(effects)
plot(effect("hincome*children", mod.multinom), style="stacked")
```
Generalized logit models: Plotting

- 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)
```
Political knowledge & party choice in Britain

Example from Fox & 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

Fit using `multinom()` in the `nnet` package

```r
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)

Response: vote

<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?

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

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

Coefficients:

```
                      (Intercept)        age gender male economic.cond.national
Labour                -0.8734   -0.01980    0.1126       0.5220
Liberal Democrat      -0.7185   -0.01460    0.0914       0.1451
                      economic.cond.household Blair Hague Kennedy Europe
Labour                0.178632   0.8236  -0.8684      0.2396  -0.001706
Liberal Democrat      0.007725   0.2779  -0.7808      0.6557     0.068412
                      political.knowledge Europe:political.knowledge
Labour                0.6583      -0.1589
Liberal Democrat      1.1602      -0.1829
```

Std. Errors:

```
                      (Intercept)        age gender male economic.cond.national
Labour                0.6908     0.005364    0.1694       0.1065
Liberal Democrat      0.7344     0.005643    0.1780       0.1100
...                       ...                       ...
```

Residual Deviance: 2233
AIC: 2277
BEPS data: Initial look, relative multiple barcharts

How does party choice—Liberal democrat, Labour, Conservative vary with political knowledge and Europe attitude (high="Eurosceptic")?
BEPS data: Effect plots to the rescue!
Age effect: Older more likely to vote Conservative
BEPS data: Effect plots to the rescue!

Attitude toward European integration $\times$ political knowledge effect:

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

- **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
  - R: MASS::polr(); VGAM::vglm()

- **Nested dichotomies**
  - Applies to ordered or unordered categories
  - Fit $m - 1$ *separate* independent models $\rightarrow$ Additive $\chi^2$ values
  - R: only need glm()

- **Generalized (multinomial) logistic regression**
  - Fit $m - 1$ logits as a *single* model
  - Results usually comparable to nested dichotomies
  - R: nnet::multinom()