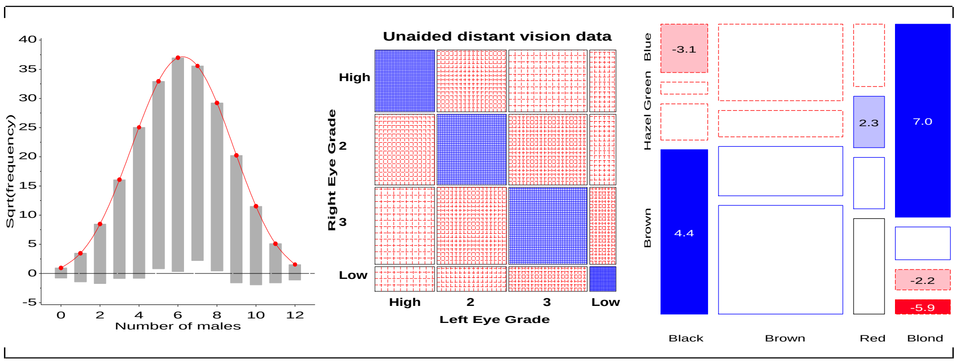
Visualizing Categorical Data with SAS and R

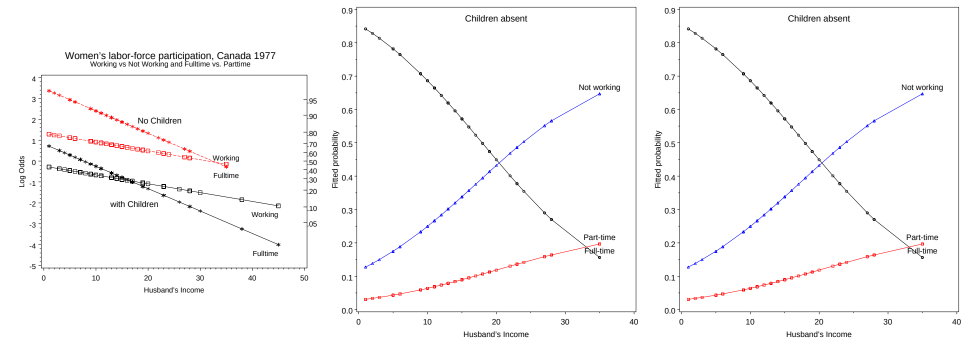
Michael Friendly

York University

Short Course, 2012
Web notes: `datavis.ca/courses/VCD/`



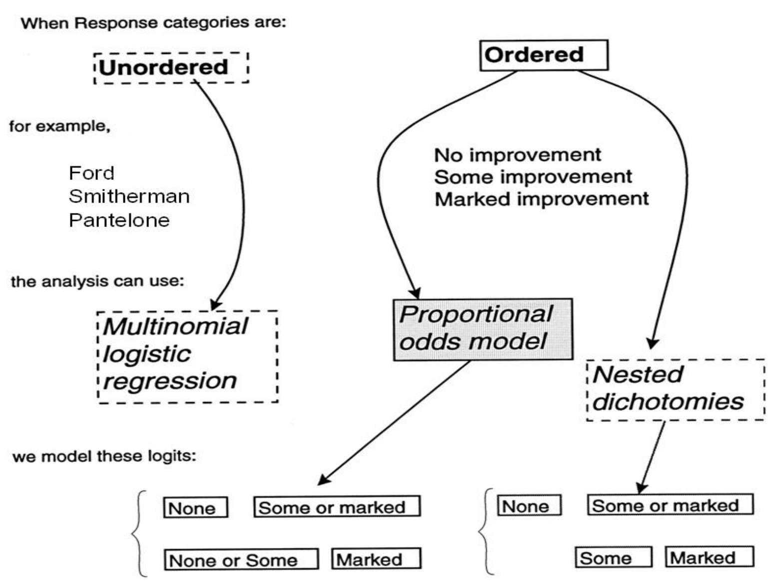
Part 5: Polytomous response models



Topics:

- Proportional odds model
- Nested dichotomies
- Generalized logit models

Polytomous responses: Overview



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

      | None | Some or Marked |
      | None or Some | Marked |

    - Assumes slopes are the same for all $m-1$ logits; only intercepts vary
  - Nested dichotomies

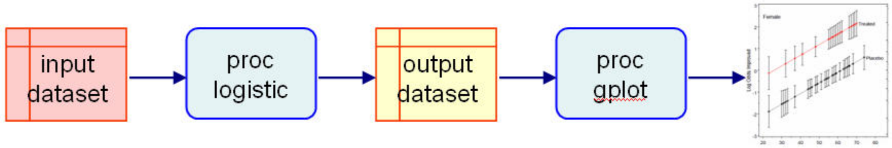
    | None | Some or Marked |
    | Some | Marked |

    - 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

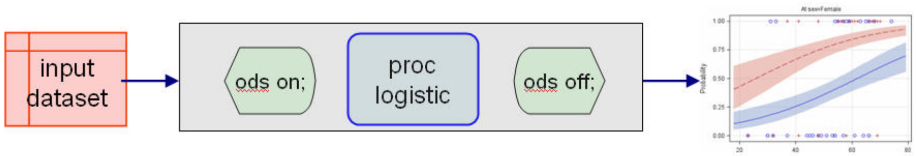
SAS, using basic capabilities:

- output dataset contains predicted probabilities (and logits) and std errors
- Utility macros (`LABELS`, `BARS`, `PSCALE`) allow plot customization



SAS, using ODS graphics (enhanced in Ver 9.2)

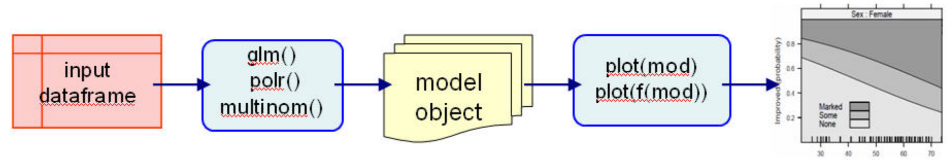
- `plots=` option for odds ratio, influence, etc
- `effectplot` statement can produce a variety of plots: boxplots, contour plots, interaction plots, etc.



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:

Sex	Treatment	Improvement None	Some	Marked	Total
F	Active	6	5	16	27
F	Placebo	19	7	6	32
M	Active	7	2	5	14
M	Placebo	10	0	1	11

- Model logits for adjacent category cutpoints:

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

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

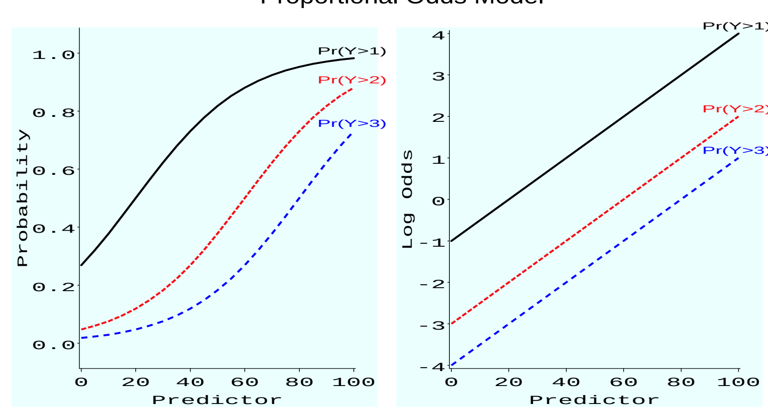
- Consider a logistic regression model for each logit:

$$\text{logit}(\theta_{ij1}) = \alpha_1 + \mathbf{x}'_{ij}\boldsymbol{\beta}_1 \qquad \text{None vs. Some/Marked}$$

$$\text{logit}(\theta_{ij2}) = \alpha_2 + \mathbf{x}'_{ij}\boldsymbol{\beta}_2 \qquad \text{None/Some vs. Marked}$$

- Proportional odds assumption: regression functions are parallel on the logit scale i.e., $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$.

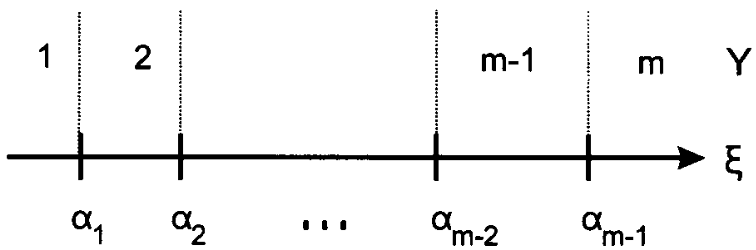
Proportional Odds Model



Proportional odds: Latent variable interpretation

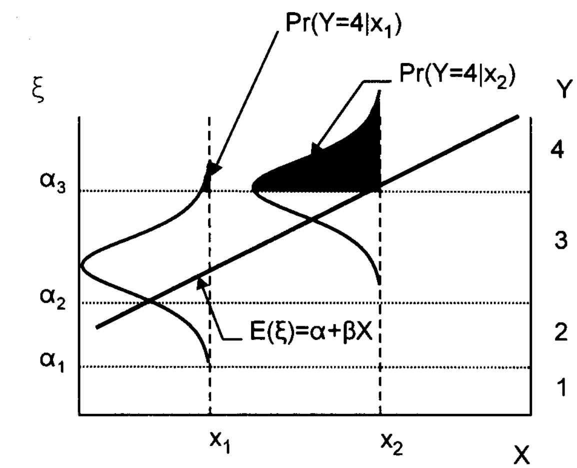
A simple motivation for the proportional odds model:

- Imagine a continuous, but *unobserved* response, ξ, a linear function of predictors
$$\xi_i = \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_i + \epsilon_i$$
- 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 ξ



Proportional odds: Latent variable interpretation

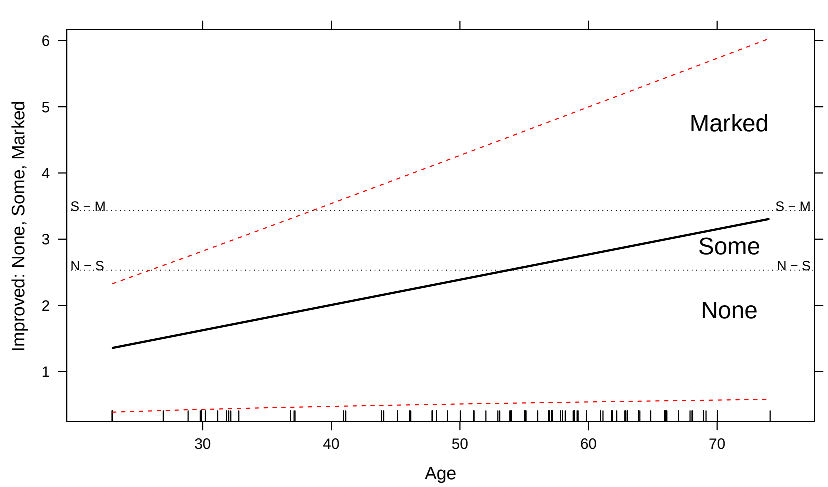
We can visualize the relation of the latent variable ξ to the observed response Y, for two values, x_1 and x_2, of a single predictor, X as shown below:



Proportional odds: Latent variable interpretation

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

Arthritis data: Age effect, latent variable scale



Proportional odds model: Fitting and plotting

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:

glogist2a.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)

```
                Response Profile

      Ordered                     Total
        Value      improve    Frequency

            1            2           28
            2            1           14
            3            0           42
```

Test of Proportional Odds Assumption: OK

```
    Score Test for the Proportional Odds Assumption

          Chi-Square       DF     Pr > ChiSq
              2.4916        3         0.4768
```

Parameter estimates (β_i):

```
                Analysis of Maximum Likelihood Estimates

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

Intercept 2           1    -4.6826     1.1949      15.3566       <.0001
Intercept 1           1    -3.7836     1.1530      10.7680       0.0010
sex       Female      1     1.2515     0.5321       5.5330       0.0187
treat     Treated     1     1.7453     0.4772      13.3774       0.0003
age                   1     0.0382     0.0185       4.2361       0.0396
```


Odds ratios ($\exp(\beta_i)$)

```
                       Odds Ratio Estimates

                                Point          95% Wald
   Effect                    Estimate      Confidence Limits

   sex   Female vs Male         3.496       1.232       9.918
   treat Treated vs Placebo     5.728       2.248      14.594
   age                          1.039       1.002       1.077
```

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

Output data set (RESULTS) for plotting:

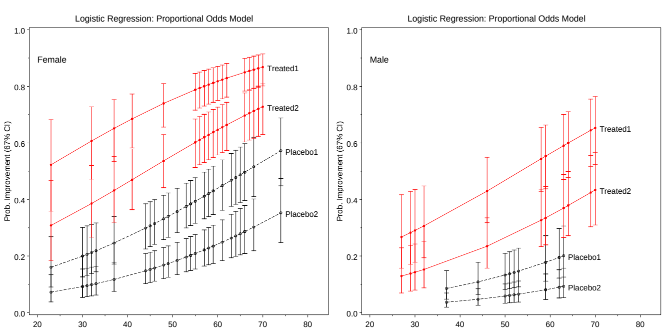
```
id   treat    sex  improve  _LEVEL_   prob   lower   upper   logit

57  Treated   Male     1       2      0.129  0.069   0.229  -1.907
57  Treated   Male     1       1      0.267  0.157   0.417  -1.008
 9  Placebo   Male     0       2      0.037  0.019   0.069  -3.271
 9  Placebo   Male     0       1      0.085  0.048   0.149  -2.372
46  Treated   Male     0       2      0.138  0.076   0.238  -1.830
46  Treated   Male     0       1      0.283  0.171   0.429  -0.931
...
```


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

··· glogist2a.sas ···

```
   *-- combine treatment and _level_, set error bar color;
data results;
   set results;
   treatl = trim(treat)||put(_level_,1.0);
   if treat='Placebo' then col='BLACK';
                      else col='RED';
proc sort data=results;
   by sex treatl age;
```

··· plot prob * age = treatl; by sex;



Add error bars and legends:

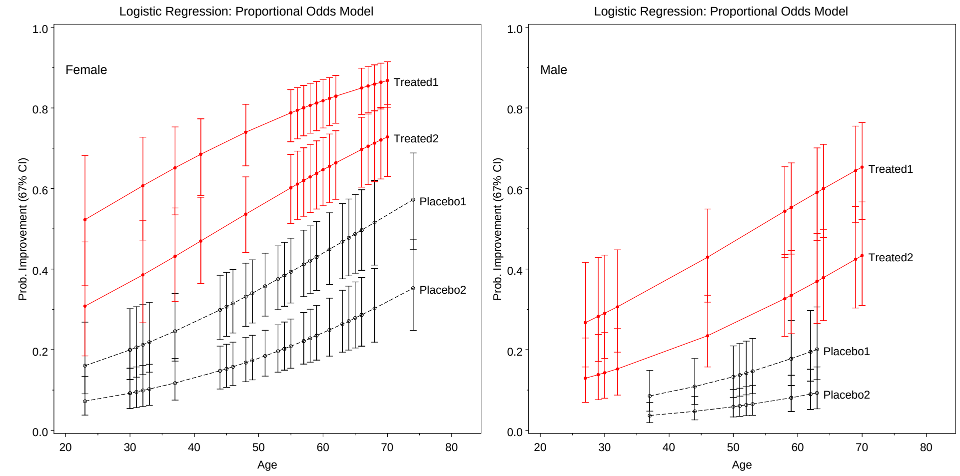
··· glogist2a.sas ···

```
   *-- Error bars, on prob scale;
%bars(data=results, var=prob,
   class=age, cvar=treatl, by=age,
   lower=lower, upper=upper,
   color=col, out=bars);
proc sort data=bars;
   by sex treatl age;

   *-- Custom legends, for treat-level and sex;
%label(data=results, y=prob, x=age, xoff=1, cvar=treatl,
   by=sex, subset=last.treatl, out=label1, pos=6, text=treatl);
%label(data=results, y=0.9, x=20, size=2,
   by=sex, subset=first.sex, out=label2, pos=6, text=sex);

   *-- Combine the annotate data sets;
data bars;
   set label1 label2 bars;
   by sex;
```

Plot step:

··· glogist2a.sas

```
goptions hby=0;
proc gplot data=results;
   plot prob * age = treatl /
        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;
```

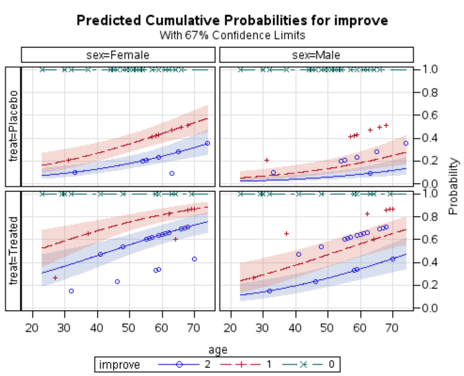



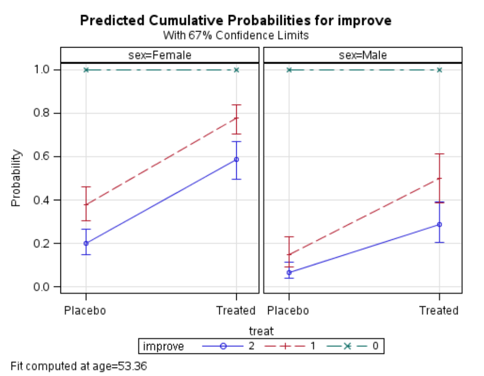
- Intercept1: Marked , Some | None
- Intercept2: Marked | Some, None
- On logit scale, these would be parallel lines
- Effects of age, treatment, sex similar to what we saw before

Effect plots using SAS ODS

arthritis-propodds-ods.sas

```
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)
```

```
 [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:

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

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

Coefficients:
                      Value Std. Error   t value
SexMale          -1.25167969 0.54636501 -2.290922
TreatmentTreated  1.74528949 0.47589542  3.667380
Age               0.03816199 0.01841628  2.072187

Intercepts:
            Value   Std. Error t value
None|Some    2.5319  1.0571     2.3952
Some|Marked  3.4309  1.0912     3.1442

Residual Deviance: 145.4579
AIC: 155.4579
```

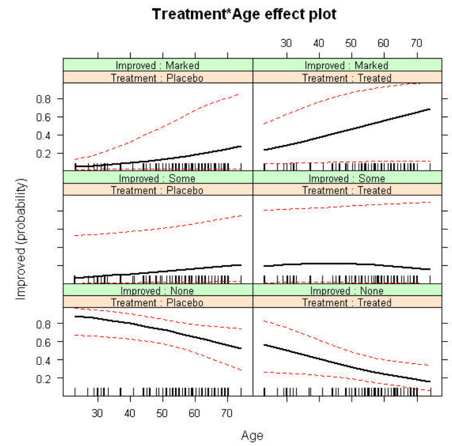
```
> Anova(arth.polr)      # Type II tests
```

```
Anova Table (Type II tests)

Response: Improved
          LR Chisq Df Pr(>Chisq)
Sex         5.6880  1  0.0170812 *
Treatment  14.7095  1  0.0001254 ***
Age         4.5715  1  0.0325081 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Proportional odds models in R: Plotting

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

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

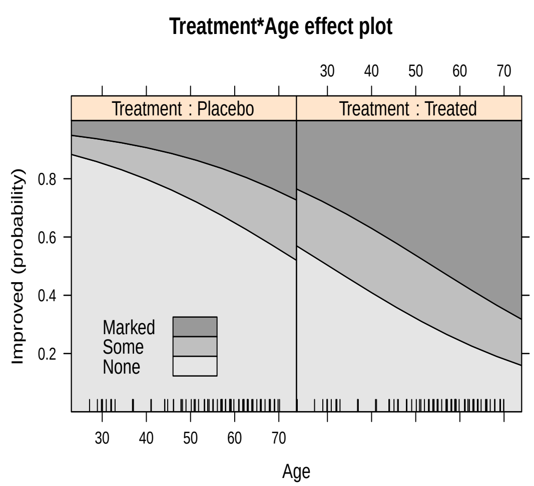


- The default plot shows all details
- But, is harder to compare across treatment and response levels

Proportional odds models in R: Plotting

Making visual comparisons easier:

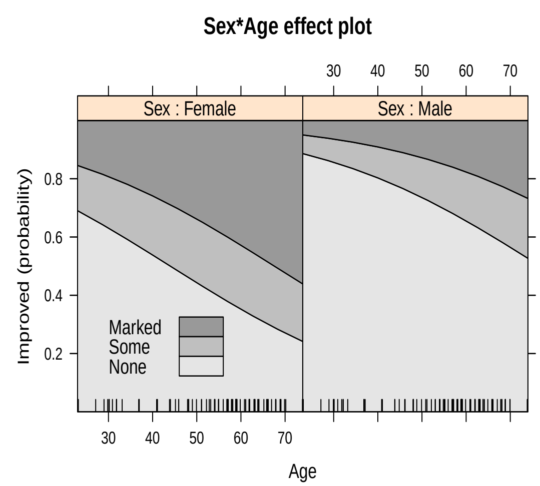
```
> plot(effect("Treatment:Age", arth.polr), style='stacked')
```



Proportional odds models in R: Plotting

Making visual comparisons easier:

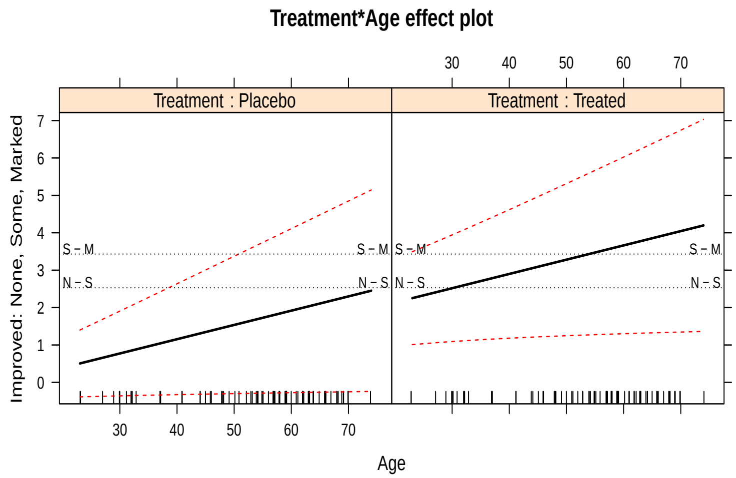
```
> plot(effect("Sex:Age", arth.polr), style='stacked')
```



Proportional odds models in R: Plotting

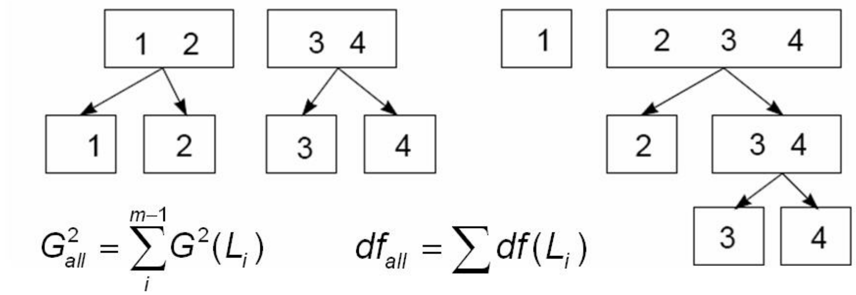
These plots are even simpler on the logit scale, using `latent=TRUE` to show the cutpoints between response categories

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



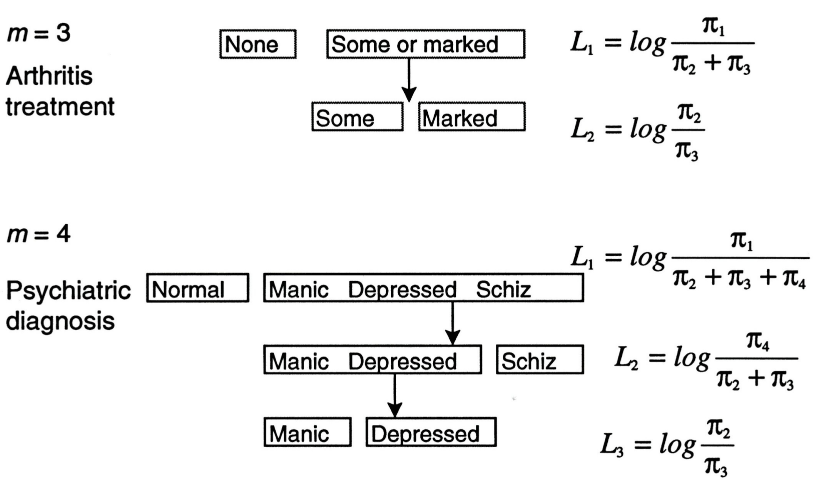
Polytomous response: Nested dichotomies

- 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



$$G^2_{all} = \sum_i^{m-1} G^2(L_i) \qquad df_{all} = \sum df(L_i)$$

Nested dichotomies: Examples



$$L_1 = log \frac{\pi_1}{\pi_2 + \pi_3} \qquad L_2 = log \frac{\pi_2}{\pi_3}$$

$$L_1 = log \frac{\pi_1}{\pi_2 + \pi_3 + \pi_4} \qquad L_2 = log \frac{\pi_4}{\pi_2 + \pi_3} \qquad L_3 = log \frac{\pi_2}{\pi_3}$$

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

wlfpart.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 ...
```


Example: Women's Labour-Force Participation

First, try proportional odds model for labour

```
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

                Chi-Square       DF      Pr > ChiSq
                   18.5638        2          <.0001
```

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 α to reject.

Fit separate models for each of working and fulltime:

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

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 $\rightarrow$ datasets for plotting
- Join for plotting:

```
data both;
   set resultsw resultsf;
   ...
```


Output for WORKING dichotomy:

```
                  Analysis of Maximum Likelihood Estimates

           Parameter Standard     Wald       Pr >        Odds
Variable DF  Estimate    Error  Chi-Square Chi-Square    Ratio

INTERCPT 1     1.3358   0.3838    12.1165     0.0005       .
HUSINC   1    -0.0423   0.0198     4.5751     0.0324     0.959
CHILDREN 1    -1.5756   0.2923    29.0651     0.0001     0.207
```

Output for FULLTIME dichotomy:

```
                  Analysis of Maximum Likelihood Estimates

           Parameter Standard     Wald       Pr >        Odds
Variable DF  Estimate    Error  Chi-Square Chi-Square    Ratio

INTERCPT 1     3.4778   0.7671    20.5537     0.0001       .
HUSINC   1    -0.1073   0.0392     7.5063     0.0061     0.898
CHILDREN 1    -2.6515   0.5411    24.0135     0.0001     0.071
```

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

$$\log\left(\frac{\Pr(\text{fulltime})}{\Pr(\text{parttime})}\right) = 3.478 - 0.107\,\text{H\$} - 2.652\,\text{kids}$$

Combined tests for Nested Dichotomoies

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

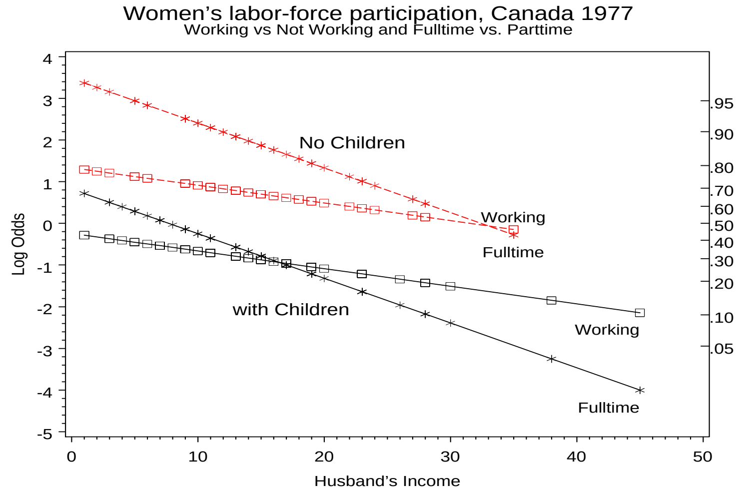
```
                     Global tests of BETA=0
                                                      Prob
Test                  Response          ChiSq     DF     ChiSq

Likelihood Ratio      working         36.4184      2    <.0001
                      fulltime        39.8468      2    <.0001
                      ALL             76.2652      4    <.0001
```

Wald tests:

```
          Wald tests of maximum likelihood estimates
                                                  Prob
    Variable      Response      WaldChiSq    DF     ChiSq

    Intercept     working         12.1164     1    0.0005
                  fulltime        20.5536     1    <.0001
                  ALL             32.6700     2    <.0001

    children      working         29.0650     1    <.0001
                  fulltime        24.0134     1    <.0001
                  ALL             53.0784     2    <.0001

    husinc        working          4.5750     1    0.0324
                  fulltime         7.5062     1    0.0061
                  ALL             12.0813     2    0.0024
```


Model visualization

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



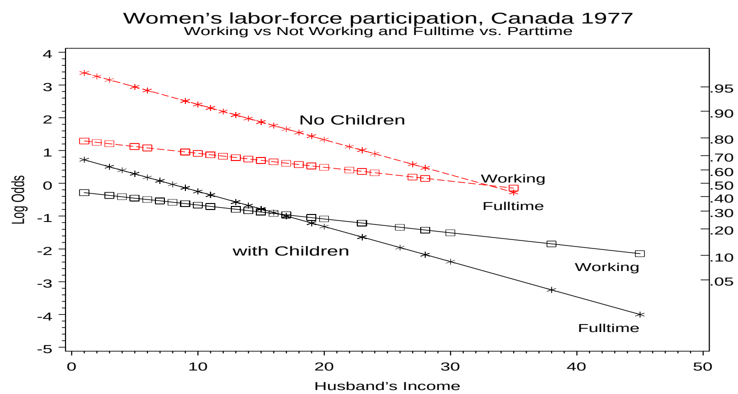
Model visualization

- Join output datasets (`resultsw` and `resultsf`)
- Combine Response & Children → `event`

```
*-- Join the results datasets to create one plot;
data both;
   set resultw(in=inw)      /* working  */
       resultf(in=inf);     /* fulltime */
   if inw then do;
      if children=1 then event='Working, with Children ';
      else event='Working, no Children ';
   end;
   else do;
      if children=1 then event='Fulltime, with Children ';
      else event='Fulltime, no Children ';
 end;
```


Model visualization

```
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;
```



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:

```
> 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
```


Nested dichotomies in R: fitting

Then, fit models for each dichotomy:

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

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

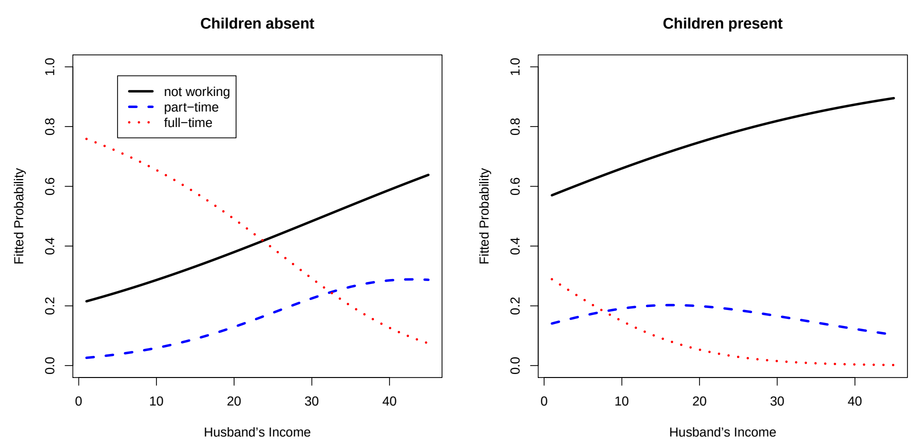
```
> pred <- expand.grid(hincome=1:45, children=c('absent', 'present'))
> # get fitted values for both sub-models
> p.work <- predict(mod.working, pred, type='response')
> p.fulltime <- predict(mod.fulltime, pred, type='response')
```

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

```
> 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 $j = 1, 2, \ldots, m-1$,

$$
\begin{aligned}
L_{jm} \equiv \log\left(\frac{\pi_{ij}}{\pi_{im}}\right) &= \beta_{0j} + \beta_{1j}\, x_{i1} + \beta_{2j}\, x_{i2} + \cdots + \beta_{kj}\, x_{ik} \\
&= \boldsymbol{\beta}_j^{\mathsf{T}} \mathbf{x}_i
\end{aligned}
$$

  - One set of fitted coefficients, $\boldsymbol{\beta}_j$ for each response category except the last.
  - Each coefficient, β_{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 caegories are calculated as:

$$
\pi_{ij} = \frac{\exp(\boldsymbol{\beta}_j^{\mathsf{T}} \mathbf{x}_i)}{\sum_{j=1}^{m-1} \exp(\boldsymbol{\beta}_j^{\mathsf{T}} \mathbf{x}_i)}\,,\; j = 1, \ldots, m-1\,; \qquad \pi_{im} = 1 - \sum_{j=1}^{m-1} \pi_{ij}
$$

Polytomous response: Generalized Logits

Fitting generalized logit models with SAS:

- Use PROC LOGISTIC with LINK=GLOGIT option.
  - output dataset → fitted probabilities, $\widehat{\pi}_{ij}$ for all m categories
  - Overall tests and specific tests for each predictor, for all m categories

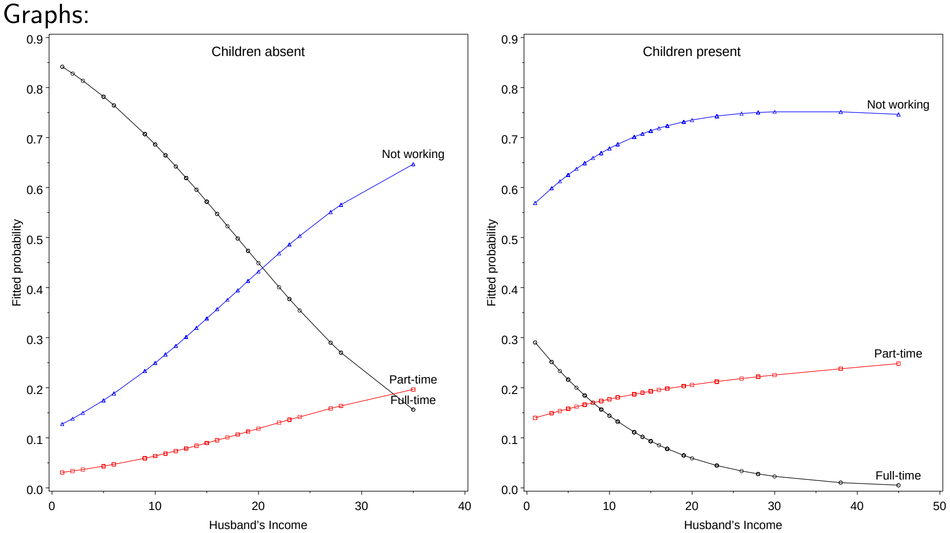
```
proc logistic data=wlfpart;
   model labor = husinc children / link=glogit;
   output out=results p=predict xbeta=logit;
```

- Can also use PROC CATMOD with RESPONSE=LOGITS statement.
  - Same model, same predicted probabilities
  - Different syntax, output dataset format, plotting steps
  - Quantitative variables: direct statement

```
proc catmod data=wlfpart;
   direct husinc;
   model labor = husinc children;
   response logits / out=results;
```


Example: Women's Labour Force Participation

Graphs:



Example: Women's Labour Force Participation

wlfpart5.sas ···

```
title 'Generalized logit model';
proc logistic data=wlfpart;
   model labor = husinc children / link=glogit;
   output out=results p=predict xbeta=logit;
```

Response profile:

```
                Ordered                          Total
                  Value         labor        Frequency

                      1             1               66
                      2             2               42
                      3             3              155

      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

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	77.6106	4	<.0001
Score	76.4850	4	<.0001
Wald	58.4351	4	<.0001

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
husinc	2	12.8159	0.0016
children	2	53.9806	<.0001

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

Coefficients:

Analysis of Maximum Likelihood Estimates

Parameter	labor	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	1.9828	0.4842	16.7709	<.0001
Intercept	2	1	-1.4323	0.5925	5.8445	0.0156
husinc	1	1	-0.0972	0.0281	11.9762	0.0005
husinc	2	1	0.00689	0.0235	0.0863	0.7689
children	1	1	-2.5586	0.3622	49.9008	<.0001
children	2	1	0.0215	0.4690	0.0021	0.9635

i.e., the fitted models are:

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

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

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

output dataset `results` (for plots):

case	labor	husinc	children	_LEVEL_	logit	predict
1	3	15	1	1	-2.03423	0.09333
1	3	15	1	2	-1.30743	0.19305
1	3	15	1	3	.	0.71363
2	3	13	1	1	-1.83977	0.11142
2	3	13	1	2	-1.32122	0.18715
2	3	13	1	3	.	0.70143
3	3	45	1	1	-4.95114	0.00528
3	3	45	1	2	-1.10067	0.24830
3	3	45	1	3	.	0.74642
4	3	23	1	1	-2.81207	0.04464
4	3	23	1	2	-1.25230	0.21238
4	3	23	1	3	.	0.74298
5	3	19	1	1	-2.42315	0.06486
5	3	19	1	2	-1.27987	0.20346
5	3	19	1	3	.	0.73168
6	3	7	1	1	-1.25639	0.18478
6	3	7	1	2	-1.36257	0.16616
...						

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

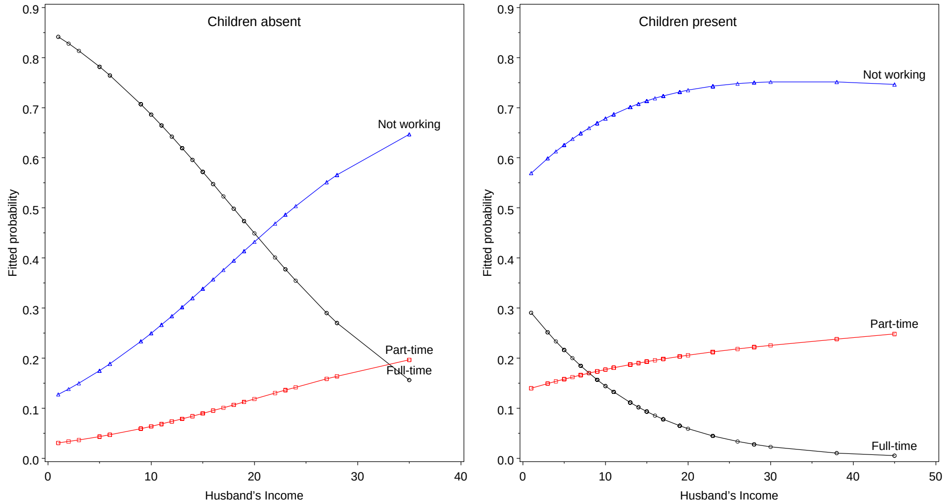
Example: Women's Labour Force Participation

··· wlfpart5.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, out=labels1);

   *-- Panel labels;
%label(data=results, x=20, y=0.85,
   by=children, subset=last.children, text=put(children, kids.),
   pos=2, size=2, out=labels2);
data labels;
   set labels1 labels2;
   by children;

goptions hby=0;
proc gplot data=results;
   plot predict * husinc = _level_ /
      vaxis=axis1 hm=1 vm=1 anno=labels nolegend;
   by children;
   axis1 order=(0 to .9 by .1) label=(a=90);
   symbol1 i=join v=circle   c=black;
   symbol2 i=join v=square   c=red;
   symbol3 i=join v=triangle c=blue;
   run;
```

Example: Women's Labour Force Participation

Graphs:



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.

```
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)
```

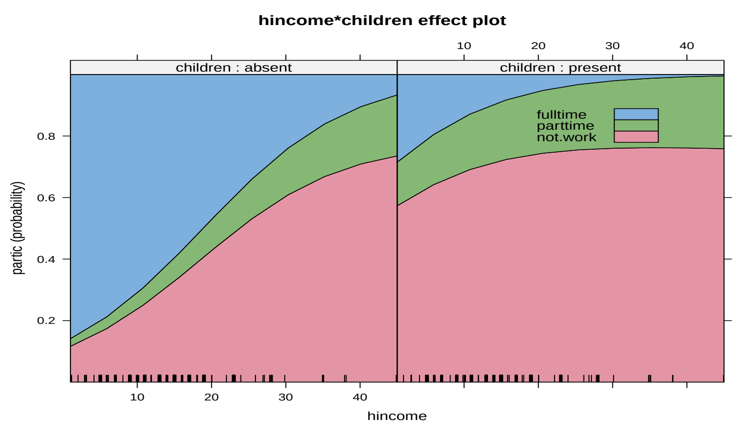
The `Anova()` tests are similar to what we got from summing these tests from the two nested dichotomies:

```
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 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

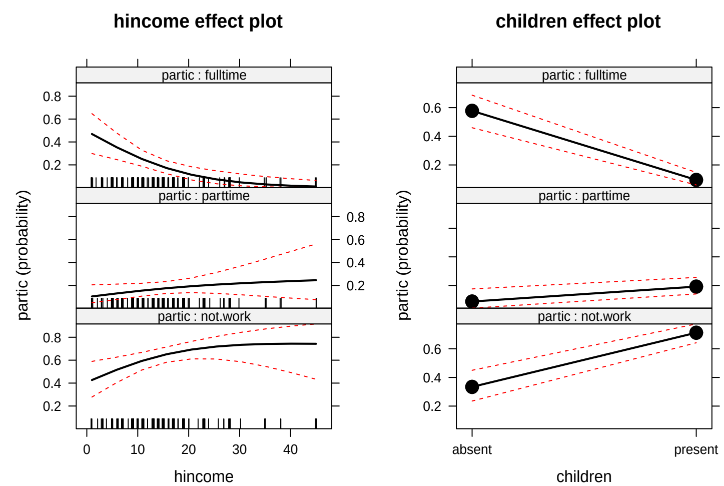
```
library(effects)
plot(effect("hincome*children", mod.multinom), style="stacked")
```



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))`.

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



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)

Response: vote
                           LR Chisq Df Pr(>Chisq)
age                            13.9  2    0.00097 ***
gender                          0.5  2    0.79726
economic.cond.national         30.6  2    2.3e-07 ***
economic.cond.household         5.7  2    0.05926 .
Blair                         135.4  2    < 2e-16 ***
Hague                         166.8  2    < 2e-16 ***
Kennedy                        68.9  2    1.1e-15 ***
Europe                         78.0  2    < 2e-16 ***
political.knowledge            55.6  2    8.6e-13 ***
Europe:political.knowledge     50.8  2    9.3e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

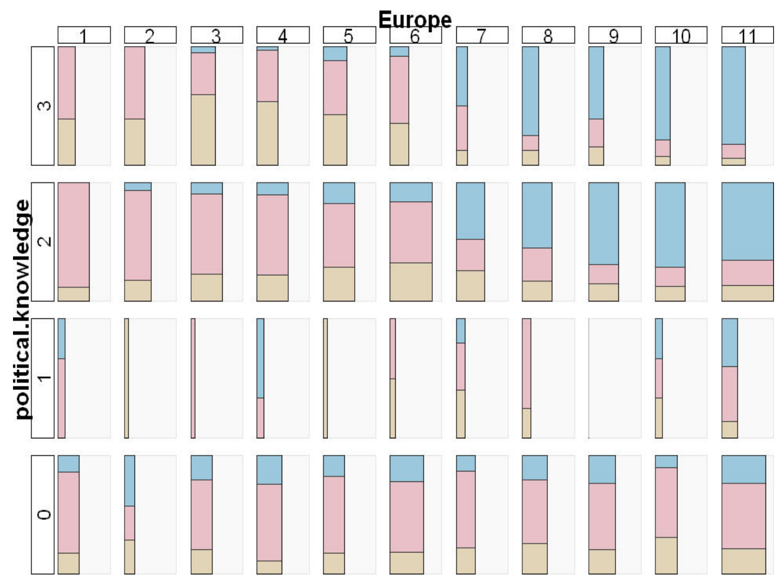

BEPS data: Interpretation?

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

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

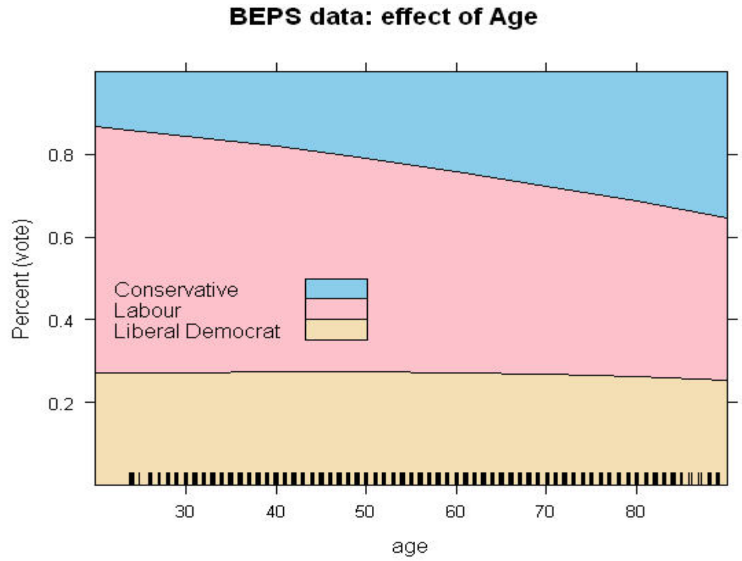
```
Call:
multinom(formula = vote ~ age + gender + economic.cond.national +
    economic.cond.household + Blair + Hague + Kennedy + Europe *
    political.knowledge, data = BEPS)

Coefficients:
                 (Intercept)      age gendermale 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 gendermale 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



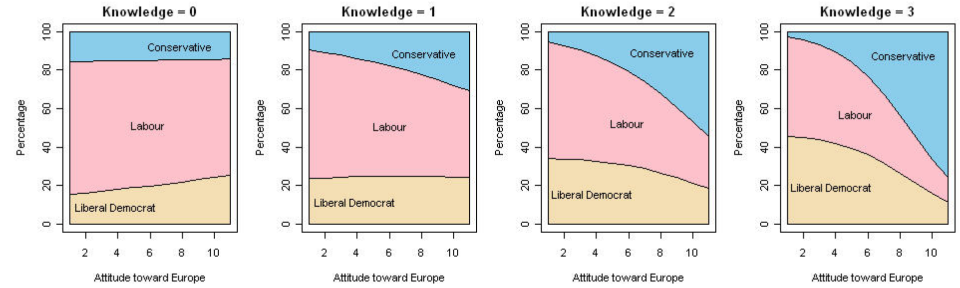
BEPS data: Effect plots to the rescue!

Age effect: Older more likely to vote Conservative



BEPS data: Effect plots to the rescue!

Attitude toward European integration × political knowledge effect:



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

Summary: Part 5

- **Polytomous responses**
  - m response categories → $m-1$ 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 asumption to be met
  - SAS: `PROC LOGISTIC`; R: `polr()`
- **Nested dichotomies**
  - Applies to ordered or unordered categories
  - Fit $m-1$ independent models → Additive χ^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** ∼ 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 ∼ 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

Atkinson, A. C. Two graphical displays for outlying and influential observations in regression. *Biometrika*, 68:13–20, 1981.

Bangdiwala, K. Using SAS software graphical procedures for the observer agreement chart. *Proceedings of the SAS User's Group International Conference*, 12:1083–1088, 1987.

Bickel, P. J., Hammel, J. W., and O'Connell, J. W. Sex bias in graduate admissions: Data from Berkeley. *Science*, 187:398–403, 1975.

Bowker, A. H. Bowker's test for symmetry. *Journal of the American Statistical Association*, 43:572–574, 1948.

Cowles, M. and Davis, C. The subject matter of psychology: Volunteers. *British Journal of Social Psychology*, 26:97–102, 1987.

Dawson, R. J. M. The "unusual episode" data revisited. *Journal of Statistics Education*, 3(3), 1995.

Fox, J. Effect displays for generalized linear models. In Clogg, C. C., editor, *Sociological Methodology, 1987*, pp. 347–361. Jossey-Bass, San Francisco, 1987.

References II

Fox, J. *Applied Regression Analysis, Linear Models, and Related Methods*. Sage Publications, Thousand Oaks, CA, 1997.

Fox, J. and Andersen, R. Effect displays for multinomial and proportional-odds logit models. *Sociological Methodology*, 36:225–255, 2006.

Friendly, M. Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89:190–200, 1994.

Friendly, M. Conceptual and visual models for categorical data. *The American Statistician*, 49:153–160, 1995.

Friendly, M. Extending mosaic displays: Marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, 8(3): 373–395, 1999.

Friendly, M. Multidimensional arrays in SAS/IML. In *Proceedings of the SAS User's Group International Conference*, volume 25, pp. 1420–1427. SAS Institute, 2000.

Friendly, M. Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56(4):316–324, 2002.

References III

Friendly, M. and Kwan, E. Effect ordering for data displays. *Computational Statistics and Data Analysis*, 43(4):509–539, 2003.

Hartigan, J. A. and Kleiner, B. Mosaics for contingency tables. In Eddy, W. F., editor, *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, pp. 268–273. Springer-Verlag, New York, NY, 1981.

Hoaglin, D. C. and Tukey, J. W. Checking the shape of discrete distributions. In Hoaglin, D. C., Mosteller, F., and Tukey, J. W., editors, *Exploring Data Tables, Trends and Shapes*, chapter 9. John Wiley and Sons, New York, 1985.

Koch, G. and Edwards, S. Clinical efficiency trials with categorical data. In Peace, K. E., editor, *Biopharmaceutical Statistics for Drug Development*, pp. 403–451. Marcel Dekker, New York, 1988.

Landis, J. R. and Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174., 1977.

Mersey, L. Report on the loss of the "Titanic" (S. S.). Parliamentary command paper 6352, 1912.

Ord, J. K. Graphical methods for a class of discrete distributions. *Journal of the Royal Statistical Society, Series A*, 130:232–238, 1967.

References IV

Ramsay, F. L. and Schafer, D. W. *The Statistical Sleuth*. Duxbury, Belmont, CA, 1997.

Srole, L., Langner, T. S., Michael, S. T., Kirkpatrick, P., Opler, M. K., and Rennie, T. A. C. *Mental Health in the Metropolis: The Midtown Manhattan Study*. NYU Press, New York, 1978.

Tufte, E. R. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

Tukey, J. W. Some graphic and semigraphic displays. In Bancroft, T. A., editor, *Statistical Papers in Honor of George W. Snedecor*, pp. 292–316. Iowa State University Press, Ames, IA, 1972.

Tukey, J. W. *Exploratory Data Analysis*. Addison Wesley, Reading, MA, 1977.

van der Heijden, P. G. M. and de Leeuw, J. Correspondence analysis used complementary to loglinear analysis. *Psychometrika*, 50:429–447, 1985.

Williams, D. A. Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics*, 36:181–191, 1987.