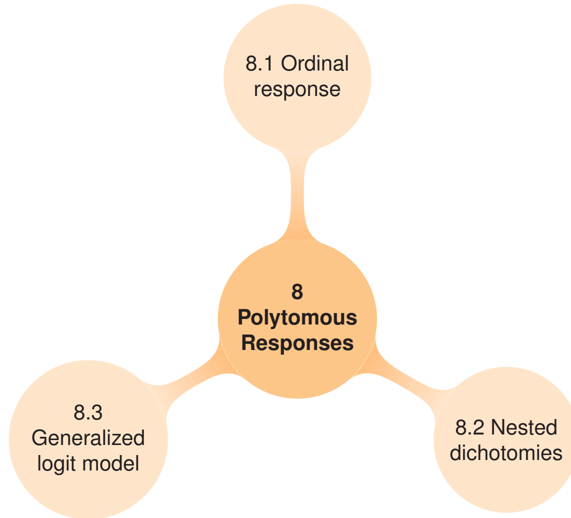


8

Models for Polytomous Responses

{ch:polytomous}

This chapter generalizes logistic regression models for a binary response to handle a multi-category (polytomous) response. Different models are available depending on whether the response categories are nominal or ordinal. Visualization methods for such models are mostly straight-forward extensions of those used for binary responses.

Polytomous response data arise when the outcome variable, Y, takes on $m > 2$ discrete values. For example, (a) patients may record that their improvement after treatment is "none," "some" or "marked;" (b) high school students may choose a general, vocational or academic program; (c) women's labor force participation may be recorded in a survey as not working outside the home, working part-time, or working full-time; (d) Canadian voters may express a preference for the Conservative, Liberal, NDP, Green party. These response categories may be considered *ordered*, as in case (a), or simply *nominal*, as in case (d), and sometimes the response can arguably be treated in either way, as in cases (b) and (c).

In this situation, there are several different ways to model the response probabilities. Let $\pi_{ij} \equiv \pi_j(\boldsymbol{x}_i)$ be the probability of response j for case or group i, given the predictors $\boldsymbol{x}_i$. Because $\sum_j \pi_{ij} = 1$, only $m-1$ of these probabilities are independent. The essential idea here is to construct a model for the polytomous (or multinomial) response composed of $m-1$ logit comparisons among the response categories in a similar way to how factors are treated in the predictor variables.

The simplest approach uses the ***proportional odds model***, described in Section 8.1. This model applies *only* when the response is ordinal (as in improvement after therapy) *and* an additional assumption (the proportional odds assumption) holds. This model can be fit using `polr()` in the

MASS (Ripley, 2015a) package, `lrm()` in the rms (Harrell, Jr., 2015) package, and `vglm()` in VGAM (Yee, 2015).

However, if the response is purely nominal (e.g., vote Conservative, Liberal, NDP, Green), or if the proportional odds assumption is untenable, another particularly simple strategy is to fit separate models to a set of $m-1$ ***nested dichotomies*** derived from the polytomous response (described in Section 8.2). This method allows you to resolve the differences among the m response categories into *independent* statistical questions (similar to orthogonal contrasts in ANOVA). For example, for women's labor force participation, it might be substantively interesting to contrast not working vs. (part-time and full-time) and then part-time vs. full-time for women who are working. You fit such nested dichotomies by running the $m-1$ binary logit models and combining the statistical results.

The most general approach is the ***generalized logit model***, also called the ***multinomial logit model***, described in Section 8.3. This model fits *simultaneously* the $m-1$ simple logit models against a baseline or reference category, for example, the last category, m. With a 3-category response, there are two generalized logits, $L_{i1} = \log(\pi_{i1}/\pi_{i3})$ and $L_{i2} = \log(\pi_{i2}/\pi_{i3})$, contrasting response categories 1 and 2 against category 3. In this approach, it doesn't matter which response category is chosen as the baseline, because all pairwise comparisons can be recovered from whatever is estimated. This model is conveniently fitted using `multinom()` in nnet (Ripley, 2015b).

8.1 Ordinal Response: Proportional Odds Model

{sec:ordinal}

For an ordered response Y, with categories $j = 1, 2, \ldots, m$, the ordinal nature of the response can be taken into account by forming logits based on the $m-1$ adjacent category cutpoints between successive categories. That is, if the cumulative probabilities are

$$\Pr(Y \le j \,|\, \boldsymbol{x}) = \pi_1(\boldsymbol{x}) + \pi_2(\boldsymbol{x}) + \cdots \pi_j(\boldsymbol{x}) \ ,$$

then the ***cumulative logit*** for category j is defined as

{eq:cumlogit}

$$L_j \equiv \text{logit}[\Pr(Y \le j \,|\, \boldsymbol{x})] = \log \frac{\Pr(Y \le j \,|\, \boldsymbol{x})}{\Pr(Y > j \,|\, \boldsymbol{x})} = \log \frac{\Pr(Y \le j \,|\, \boldsymbol{x})}{1 - \Pr(Y \le j \,|\, \boldsymbol{x})} \tag{8.1}$$

for $j = 1, 2, \ldots m-1$.

In our running example of responses to arthritis treatment, the actual response variable is `Improved`, with ordered levels `"None" < "Some" < "Marked"`. In this case, the cumulative logits would be defined as

$$L_1 = \log \frac{\pi_1(\boldsymbol{x})}{\pi_2(\boldsymbol{x}) + \pi_3(\boldsymbol{x})} = \text{logit (None vs. [Some or Marked])}$$

$$L_2 = \log \frac{\pi_1(\boldsymbol{x}) + \pi_2(\boldsymbol{x})}{\pi_3(\boldsymbol{x})} = \text{logit ([None or Some] vs. Marked)} \ ,$$

where $\boldsymbol{x}$ represents the predictors (sex, treatment and age).

The ***proportional odds model*** (PO) (McCullagh, 1980) proposes a simple and parsimonious account of these effects, where the predictors in (x) are constrained to have the same slopes for all cumulative logits,

{eq:propodds}

$$L_j = \alpha_j + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta} \qquad j = 1, \ldots, m-1 \ . \tag{8.2}$$

That is, the effect of the predictor x_i is the same, β_i, for all cumulative logits. The cumulative logits differ only in their intercepts. In this formulation, the $\{\alpha_j\}$ increase with j, because $\Pr(Y \le j \,|\, \boldsymbol{x})$

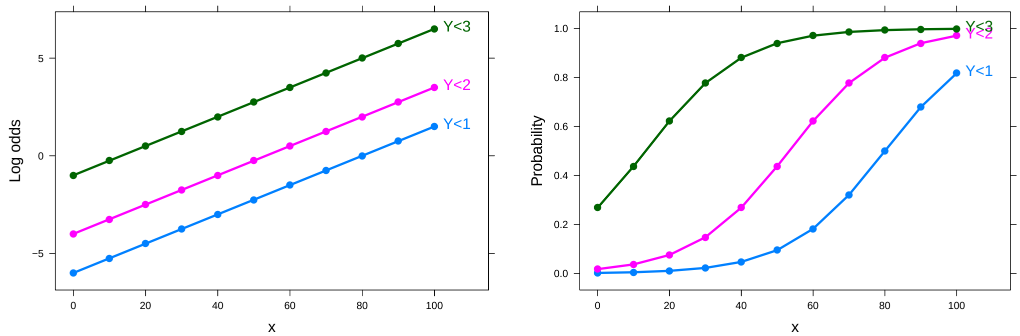


Figure 8.1: Proportional odds model for an ordinal response. The model assumes equal slopes for the cumulative response logits. Left: logit scale; right: probability scale. {fig:podds}

increases with j for fixed $\boldsymbol{x}$.[1] Figure 8.1 portrays the PO model for a single quantitative predictor x with $m = 4$ response categories.

The name "proportional odds" stems from the fact that under Eqn. (8.2), for fixed $\boldsymbol{x}$, the cumulative log odds (logits) for categories j and $j\prime$ are constant and their difference is $(\alpha_j - \alpha_{j\prime})$, so the odds have a constant ratio $\exp(\alpha_j - \alpha_{j\prime}) = \exp(\alpha_j)/\exp(\alpha_{j\prime})$, or are proportional. Similarly, the ratio of the cumulative odds of making a response $Y \leq j$ at values of the predictors $\boldsymbol{x} = \boldsymbol{x}_1$ are $\exp((\boldsymbol{x}_1 - \boldsymbol{x}_2)^\mathsf{T}\boldsymbol{\beta})$ times the odds of this response at $\boldsymbol{x} = \boldsymbol{x}_2$, so the log cumulative odds ratio is proportional to the difference between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

8.1.1 Latent variable interpretation

For a binary response, an alternative motivation for logistic regression regards the relation of the observed Y as arising from a continuous, unobserved, (latent) response variable, ξ representing the propensity for a "success" (1) rather than "failure" (0). The latent response is assumed to be linearly related to the predictors $\boldsymbol{x}$ according to

$$\xi_i = \alpha + \boldsymbol{x}_i^\mathsf{T}\beta + \epsilon_i = \alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \qquad (8.3)$$ {eq:latent}

However, we can only observe $Y_i = 1$ when ξ_i passes some threshold, that with some convenient scaling can be taken as $\xi_i > 0 \Longrightarrow Y_i = 1$.[2]

The latent variable motivation extends directly to an ordinal response under the PO model. We now assume that there is a set of $m - 1$ thresholds, $\alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}$ for the latent variable ξ_i in Eqn. (8.3) and we observe

$$Y_i = j \quad \text{if} \quad \alpha_{j-1} < \xi_i \leq \alpha_j \ ,$$

with appropriate modifications to the inequalities at the end points.

[1] Some authors and some software describe the PO model in terms of $\text{logit}[\Pr(Y > j \mid \boldsymbol{x})]$, so the signs and order of the intercepts, α_j, are reversed.

[2] The latent variable derivation of logistic regression (and the related probit model) was fundamental in the history of statistical methods for discrete response outcomes. An early example was Thurstone's (1927) *Law of comparative judgment* designed to account for psychological preference by assuming an underlying latent continuum of "hedonic values." Similarly, the probit model arose from dose-response studies in toxicology (Bliss, 1934, Finney, 1947) where the number killed by some chemical agent was related to its type, dose or concentration. The idea of a latent variable was also at the heart of the development of factor analysis (Bentler, 1980) and latent class analysis (Lazarsfeld, 1950, 1954) was developed to treat the problem of classifying individuals into discrete latent classes from fallible measurements. See Bollen (2002) for a useful overview of latent variable models in the social sciences.

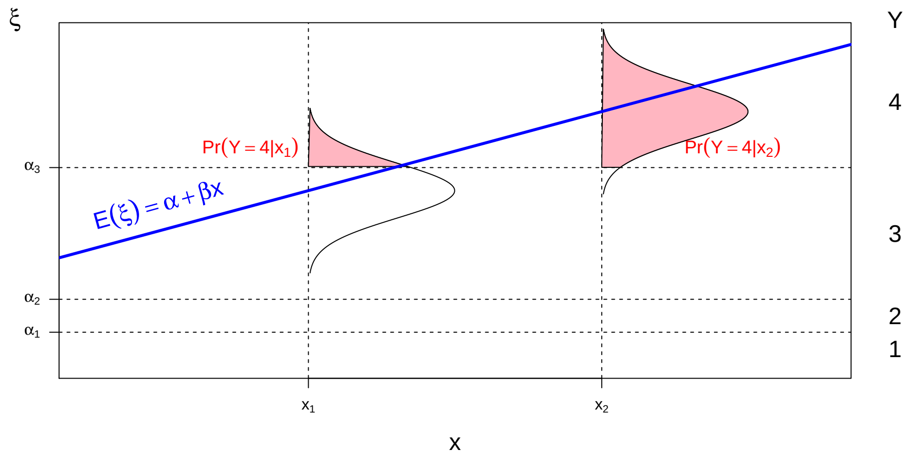


Figure 8.2: Latent variable representation of the proportional odds model for $m = 4$ response categories and a single quantitative predictor, x. *Source*: Adapted from Fox (2008, Fig 14.10), using code provided by John Fox. {fig:latent}

This is illustrated in Figure 8.2 for a response with $m = 4$ ordered categories and a single quantitative predictor, x. The observable response Y categories are shown on the right vertical axis, and the corresponding latent continuous variable ξ on the left axis together with the thresholds $\alpha_1, \alpha_2, \alpha_3$. The (conditional) logistic distribution of ξ is shown at two values of x, and the shaded areas under the curve give the conditional probabilities $\Pr(Y = 4 \,|\, x_i)$ for the two values x_1 and x_2.

8.1.2 Fitting the proportional odds model

As mentioned earlier, there are a number of different R packages that provide facilities for fitting the PO model. These have somewhat different capabilities for reporting results, testing hypotheses and plotting, so we generally use `polr()` in the MASS package, except where other packages offer greater convenience.

Unless the response variable has numeric values, it is important to ensure that it has been defined as an *ordered* factor (using `ordered()`). In the *Arthritis* data, the response, `Improved` was setup this way, as we can check by printing some of the values.[3]

```
> data("Arthritis", package="vcd")
> head(Arthritis$Improved, 8)

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

We fit the main effects model for the ordinal response using `polr()` as shown below. We also specify `Hess=TRUE` to have the function return the observed information matrix (called the Hessian), that is used in other operations to calculate standard errors.

```
> library(MASS)
> arth.polr <- polr(Improved ~ Sex + Treatment + Age,
+                   data=Arthritis, Hess=TRUE)
> summary(arth.polr)
```

[3] As an unordered factor, the levels would be treated as ordered alphabetically, i.e., `Marked`, `None`, `Some`.

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

Coefficients:
                   Value Std. Error t value
SexMale          -1.2517     0.5464   -2.29
TreatmentTreated  1.7453     0.4759    3.67
Age               0.0382     0.0184    2.07

Intercepts:
            Value  Std. Error t value
None|Some    2.532  1.057      2.395
Some|Marked  3.431  1.091      3.144

Residual Deviance: 145.46
AIC: 155.46
```

The output from the `summary()` method, shown above, gives the estimated coefficients ($\boldsymbol{\beta}$) and intercepts (α_j) labeled by the cutpoint on the ordinal response. It provides standard errors and t-values ($\beta_i / SE(\beta_i)$), but no significance tests or p-values. The car (Fox and Weisberg, 2015)::`Anova()` method gives the appropriate tests.

```
> library(car)
> Anova(arth.polr)

Analysis of Deviance Table (Type II tests)

Response: Improved
          LR Chisq Df Pr(>Chisq)
Sex           5.69  1    0.01708 *
Treatment    14.71  1    0.00013 ***
Age           4.57  1    0.03251 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

8.1.3 Testing the proportional odds assumption

The simplicity of the PO model is achieved only when the proportional odds model holds for a given data set. In essence, a test of this assumption involves a contrast between the PO model and a generalized logit NPO model that allows different effects (slopes) of the predictors across the response categories:

$$\text{PO}: \quad L_j = \alpha_j + \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta} \qquad j = 1, \dots, m-1 \tag{8.4}$$ {eq:po}

$$\text{NPO}: \quad L_j = \alpha_j + \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta}_j \qquad j = 1, \dots, m-1 \tag{8.5}$$ {eq:npo}

The most general test involves fitting both models and testing the difference in the residual deviance by a likelihood ratio test or using some other measure (such as AIC) for model comparison. The PO model (Eqn. (8.4)) has $(m-1)+p$ parameters, while the NPO model (Eqn. (8.5)) has $(m-1)(1+p) = m(1+p)$ parameters, which may be difficult to fit if this is large relative to the number of observations. An intermediate model, the ***partial proportional odds model*** (Peterson and Harrell, 1990) allows one subset of predictors, $\boldsymbol{x}_{po}$, to satisfy the proportional odds assumption (equal slopes), while the remaining predictors $\boldsymbol{x}_{npo}$ have slopes varying with the response level:

$$\text{PPO}: \quad L_j = \alpha_j + \boldsymbol{x}_{po}^{\mathsf{T}}\boldsymbol{\beta} + \boldsymbol{x}_{npo}^{\mathsf{T}}\boldsymbol{\beta}_j \qquad j = 1, \dots, m-1 \ . \tag{8.6}$$ {eq:ppo}

In R, the PO and NPO models can be readily contrasted by fitting them both using `vglm()` in the VGAM package. This defines the `cumulative` family of models and allows a `parallel` option. With `parallel=TRUE`, this is equivalent to the `polr()` model, except that the signs of the coefficients are reversed.

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

Call:
vglm(formula = Improved ~ Sex + Treatment + Age, family = cumulative(parallel = TRUE),
    data = Arthritis)

Coefficients:
   (Intercept):1    (Intercept):2          SexMale
        2.531990         3.430988         1.251671
TreatmentTreated              Age
       -1.745304        -0.038163

Degrees of Freedom: 168 Total; 163 Residual
Residual deviance: 145.46
Log-likelihood: -72.729
```

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

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

Call:
vglm(formula = Improved ~ Sex + Treatment + Age, family = cumulative(parallel = FALSE),
    data = Arthritis)

Coefficients:
     (Intercept):1      (Intercept):2          SexMale:1
          2.618539           3.431175           1.509827
         SexMale:2 TreatmentTreated:1 TreatmentTreated:2
          0.866434          -1.836929          -1.704011
             Age:1              Age:2
         -0.040866          -0.037294

Degrees of Freedom: 168 Total; 160 Residual
Residual deviance: 143.57
Log-likelihood: -71.787
```

The VGAM package defines a `coef()` method that can print the coefficients in a more readable matrix form giving the category cutpoints:

```
> coef(arth.po, matrix=TRUE)

                 logit(P[Y<=1]) logit(P[Y<=2])
(Intercept)            2.531990       3.430988
SexMale                1.251671       1.251671
TreatmentTreated      -1.745304      -1.745304
Age                   -0.038163      -0.038163

> coef(arth.npo, matrix=TRUE)

                 logit(P[Y<=1]) logit(P[Y<=2])
(Intercept)            2.618539       3.431175
SexMale                1.509827       0.866434
TreatmentTreated      -1.836929      -1.704011
Age                   -0.040866      -0.037294
```

In most cases, nested models can be tested using an `anova()` method, but the VGAM package has not implemented this for "vglm" objects. Instead, it provides an analogous function, `lrtest()`:

```
> 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
```

The LR test can be also calculated as “manually” shown below using the difference in residual deviance for the two models.

```
> tab <- cbind(
+   Deviance = c(deviance(arth.npo), deviance(arth.po)),
+         df = c(df.residual(arth.npo), df.residual(arth.po))
+         )
> tab <- rbind(tab, diff(tab))
> rownames(tab) <- c("GenLogit", "PropOdds", "LR test")
> tab <- cbind(tab, pvalue=1-pchisq(tab[,1], tab[,2]))
> tab

         Deviance  df  pvalue
GenLogit 143.5741 160 0.81966
PropOdds 145.4579 163 0.83435
LR test    1.8838   3 0.59686
```

The `vglm()` can also fit partial proportional odds models, by specifying a formula giving the terms for which the PO assumption should be taken as `TRUE` or `FALSE`. Here we illustrate this using `parallel=FALSE ~ Sex`, to fit separate slopes for males and females, but parallel lines for the other predictors. The same model would be fit using `parallel=TRUE ~ Treatment + Age`.

```
> arth.ppo <- vglm(Improved ~ Sex + Treatment + Age, data=Arthritis,
+   family = cumulative(parallel=FALSE ~ Sex))
> coef(arth.ppo, matrix=TRUE)

                 logit(P[Y<=1]) logit(P[Y<=2])
(Intercept)            2.542452       3.615561
SexMale                1.483336       0.867362
TreatmentTreated      -1.775742      -1.775742
Age                   -0.039622      -0.039622
```

8.1.4 Graphical assessment of proportional odds

There are several graphical methods for visual assessment of the proportional odds assumption. These are all *marginal* methods, in that they treat the predictors one at a time. However, that provides one means to determine if a partial proportional odds model might be more appropriate. Harrell's work *Regression Modeling Strategies* (2001, Ch. 13–14) and the corresponding rms package provide an authoritative treatment and methods in R.

One simple idea is to plot the conditional mean or expected value $E(X \mid Y)$ 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. For comparison, one can also plot the estimated conditional means $\widehat{E}(X \mid Y = j)$ under the fitted PO model X as the only predictor. If the PO assumption holds for this X, the model-mean curve should be close to the data mean curve.

```
> library(rms)
> arth.po2 <- lrm(Improved ~ Sex + Treatment + Age, data=Arthritis)
> arth.po2


Logistic Regression Model

lrm(formula = Improved ~ Sex + Treatment + Age, data = Arthritis)

                      Model Likelihood     Discrimination    Rank Discrim.
                         Ratio Test           Indexes           Indexes
Obs             84    LR chi2      24.46    R2        0.291    C       0.750
 None           42    d.f.             3    g         1.335    Dxy     0.500
 Some           14    Pr(> chi2) <0.0001    gr        3.801    gamma   0.503
 Marked         28                          gp        0.280    tau-a   0.309
max |deriv| 1e-07                           Brier     0.187

                  Coef    S.E.   Wald Z Pr(>|Z|)
y>=Some           -2.5320 1.0570 -2.40  0.0166
y>=Marked         -3.4310 1.0911 -3.14  0.0017
Sex=Male          -1.2517 0.5464 -2.29  0.0220
Treatment=Treated  1.7453 0.4759  3.67  0.0002
Age                0.0382 0.0184  2.07  0.0382
```

The plot of conditional X means is produced using the `plot.xmean.ordinaly()` as shown below. It produces one marginal panel for each predictor in the model. For categorical predictors, it plots only the overall most frequent category. The resulting plot is shown in Figure 8.3.

```
> op <- par(mfrow=c(1,3))
> plot.xmean.ordinaly(Improved ~ Sex + Treatment + Age, data=Arthritis,
+                     lwd=2, pch=16, subn=FALSE)
> par(op)
```

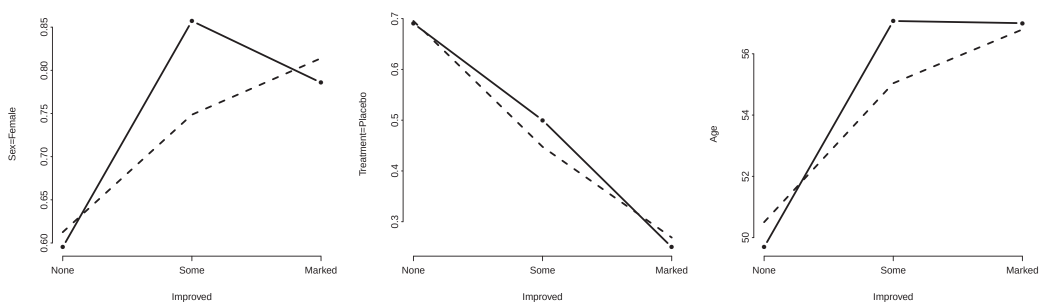


Figure 8.3: Visual assessment of ordinality and the proportional odds assumption for predictors in the Arthritis data. Solid lines connect the stratified means of X given Y. Dashed lines show the estimated expected value of X given Y=j if the proportional odds model holds for X. {fig:arth-rmsplot}

In Figure 8.3, there is some evidence that the effect of `Sex` is non-monotonic and the means differ from their model-implied values under the PO assumption. The effect of `Treatment` looks good by this method, and the effect of `Age` hints that the upper two categories may not be well-distinguished as an ordinal response.

Of course, this example has only a modest total sample size, and this method only examines the marginal effects of the predictors. Nevertheless, it is a useful supplement to the statistical methods described earlier.

8.1.5 Visualizing results for the proportional odds model

{sec:vis-propodds}

Results from the PO model (and other models for polytomous responses) can be graphed using the same ideas and methods shown earlier for a binary or binomial response. In particular, full-model plots (described earlier in Section 7.3.2) and effect plots (Section 7.3.3) are still very helpful.

But now there is the additional complication that the response variable has $m > 2$ levels and so needs to be represented by $m - 1$ curves or panels in addition to those related to the predictor variables.

8.1.5.1 Full-model plots

{sec:po-fullplots}

For full-model plots, we continue the idea of appending the fitted response probabilities (or logits) to the data frame and plotting these in relation to the predictors. The `predict()` method returns the highest probability category label by default (with `type="class"`), so to get the fitted probabilities you have to ask for `type="probs"`, as shown below.

```
> arth.fitp <- cbind(Arthritis,
+                    predict(arth.polr, type="probs"))
> head(arth.fitp)

  ID Treatment  Sex Age Improved    None    Some  Marked
1 57   Treated Male  27     Some 0.73262 0.13806 0.12932
2 46   Treated Male  29     None 0.71740 0.14443 0.13816
3 77   Treated Male  30     None 0.70960 0.14763 0.14277
4 17   Treated Male  32   Marked 0.69363 0.15400 0.15237
5 36   Treated Male  46   Marked 0.57025 0.19504 0.23471
6 23   Treated Male  58   Marked 0.45634 0.21713 0.32653
```

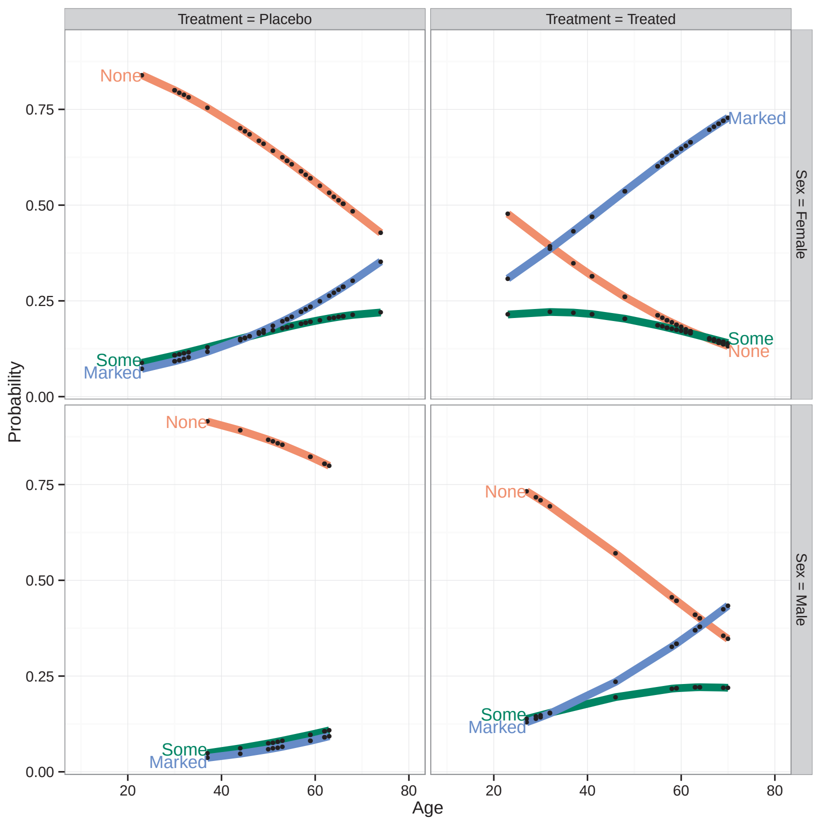
For plotting, it is most convenient to reshape these from wide to long format using `melt()` in the reshape2 (Wickham, 2014) package. The response category is named `Level`.

```
> library(reshape2)
> plotdat <- melt(arth.fitp,
+                 id.vars = c("Sex", "Treatment", "Age", "Improved"),
+                 measure.vars=c("None", "Some", "Marked"),
+                 variable.name = "Level",
+                 value.name = "Probability")
> ## view first few rows
> head(plotdat)

   Sex Treatment Age Improved Level Probability
1 Male   Treated  27     Some  None     0.73262
2 Male   Treated  29     None  None     0.71740
3 Male   Treated  30     None  None     0.70960
4 Male   Treated  32   Marked  None     0.69363
5 Male   Treated  46   Marked  None     0.57025
6 Male   Treated  58   Marked  None     0.45634
```

We can now plot `Probability` against `Age`, using `Level` to assign different colors to the lines for the response categories. `facet_grid()` is used to split the plot into separate panels by `Sex` and `Treatment`. In this example, the directlabels (Hocking, 2013) package is also used replace the default legend created by `ggplot()` with category labels on the curves themselves, which is easier to read.

```
> library(ggplot2)
> library(directlabels)
> gg <- ggplot(plotdat, aes(x = Age, y = Probability, colour = Level)) +
+     geom_line(size=2.5) + theme_bw() + xlim(10,80) +
+     geom_point(color="black", size=1.5) +
```

```
+     facet_grid(Sex ~ Treatment,
+                labeller = function(x, y) sprintf("%s = %s", x, y)
+                )
> direct.label(gg)
```



{fig:arth-polr1} **Figure 8.4:** Predicted probabilities for the proportional odds model fit to the Arthritis data

Although we now have three response curves in each panel, this plot is relatively easy to understand: (a) In each panel, the probability of no improvement decreases with age, while that for marked improvement increases. (b) It is easy to compare the placebo and treated groups in each row, showing that no improvement decreases, while marked improvement increases with the active treatment. (On the other hand, this layout makes it harder to compare panels vertically for males and females in each condition.) (c) The points show where the observations are located in each panel; so, we can see that the data is quite thin for males given the placebo.[4]

8.1.5.2 Effect plots

{sec:po-effplots}

For PO models fit using `polr()`, the effects (Fox *et al.*, 2015) package provides two different styles for plotting a given effect. By default, curves are plotted in separate panels for the different response levels of a given effect, together with confidence bands for predicted probabilities. This

[4] One way to improve (pun intended) this graph would be to show the points on the lines only for the actual level of `Improve` for each observation.

form provides confidence bands and rug plots for the observations, but the default vertical arrangement of the panels makes it harder to compare the trends for the different response levels. The alternative *stacked* format shows the changes in response level more directly, but doesn't provide confidence bands.

Figure 8.5 shows these two styles for the main effect of Age in the proportional odds model, arth.polr fit earlier.

```
> library(effects)
> plot(Effect("Age", arth.polr))
> plot(Effect("Age", arth.polr), style='stacked',
+      key.args=list(x=.55, y=.9))
```

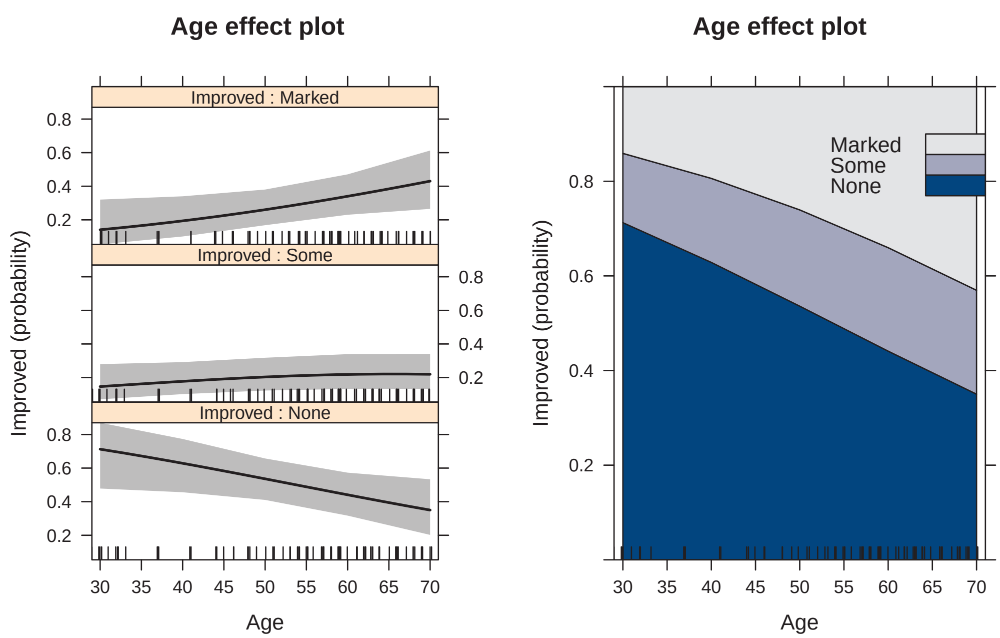


Figure 8.5: Effect plots for the effect of Age in the proportional odds model for the Arthritis data. Left: responses shown in separate panels. Right: responses shown in stacked format {fig:arth-po-eff1}

Even though this model includes only main effects, you can still plot the higher-order effects for more focal predictors in a coherent display. Figure 8.6 shows the predicted probabilities for all three predictors together. Again, visual comparison is easier horizontally for placebo versus treated groups, but you can also see that the prevalence of marked improvement is greater for females than for males.

```
> plot(Effect(c("Treatment", "Sex", "Age"), arth.polr),
+      style="stacked", key.arg=list(x=.8, y=.9))
```

Finally, the latent variable interpretation of the PO model provides for simpler plots on the logit scale. Figure 8.7 shows this plot for the effects of Treatment and Age (collapsed over Sex) produced with the argument latent=TRUE to Effect(). In this plot, there is a single line in each panel for the effect (slope) of Age on the log odds. The dashed horizontal lines give the thresholds between the adjacent response categories corresponding to the intercepts.

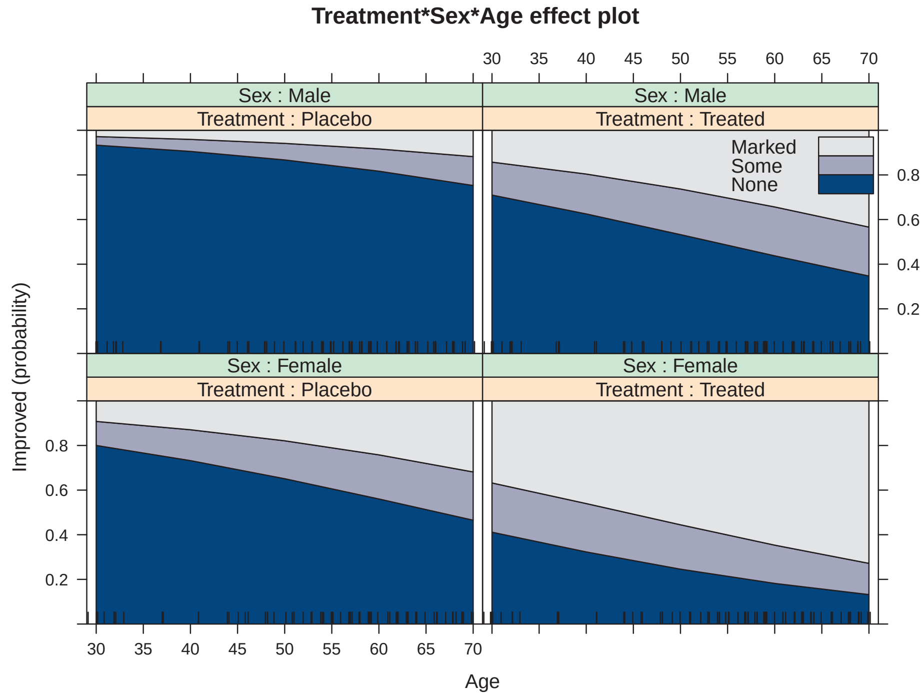
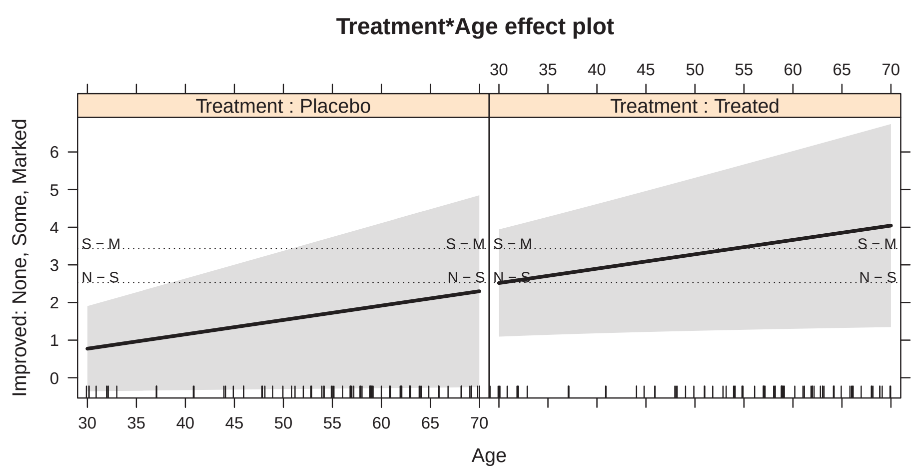


Figure 8.6: Effect plot for the effects of Treatment, Sex and Age in the Arthritis data. {fig:arth-po-eff2}

```
> plot(Effect(c("Treatment", "Age"), arth.polr, latent=TRUE), lwd=3)
```



{fig:arth-po-eff3} **Figure 8.7:** Latent variable effect plot for the effects of Treatment and Age in the Arthritis data.

8.2 Nested dichotomies

{sec:nested}

The method of ***nested dichotomies*** provides another simple way to analyse a polytomous response in the framework of logistic regression (or other generalized linear models). This method does not require an ordinal response or special software. Instead, it uses the familiar binary logistic model and fits $m-1$ separate models for each of a hierarchically nested set of comparisons among the response categories.

Taken together, this set of models for the dichotomies comprises a complete model for the polytomous response. As well, these models are statistically independent, so test statistics such as G^2 or Wald tests can be added to give overall tests for the full polytomy.

For example, the response categories Y = {1,2,3,4} could be divided first as {1,2} vs. {3,4}, as shown in the left side of Figure 8.8. Then these two dichotomies could be divided as {1} vs. {2}, and {3} vs. {4}. Alternatively, these response categories could be divided as shown in the right side of Figure 8.8: first, {1} vs. {2,3,4}, then {2} vs {3,4}, and finally {3} vs. {4}.

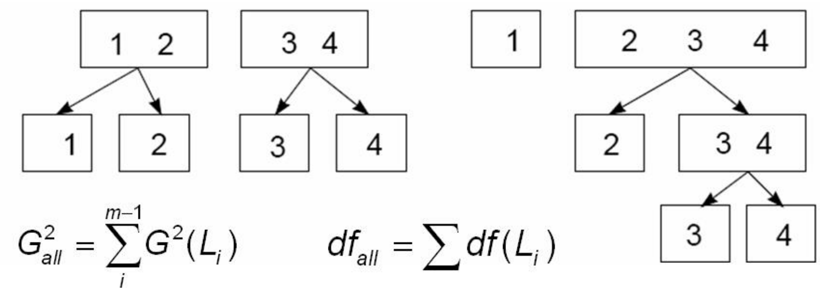


Figure 8.8: Nested dichotomies. The boxes show two different ways a four-category response can be represented as three nested dichotomies. Adapted from Fox (2008). {fig:nested2}

Such models make the most sense when there are substantive reasons for considering the response categories in terms of such dichotomies. Two examples are shown in Figure 8.9.

- For the `Arthritis` data, it is sensible to consider one dichotomy ("better"), with logit L_1, between the categories of `"None"` compared to `"Some"` or `"Marked"`. A second dichotomy, with logit L_2, would then distinguish between the some and marked response categories.
- For a second case where patients are classified into $m = 4$ psychiatric diagnostic categories, the first dichotomy, with logit L_1 distinguishes those considered normal from all others given a clinical diagnosis. Two other dichotomies are defined to further divide the non-normal categories.

Then, consider the separate logit models for these $m-1$ dichotomies, with different intercepts α_j and slopes $\boldsymbol{\beta}_j$ for each dichotomy,

$$\begin{aligned}
L_1 &= \alpha_1 + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta}_1 \\
L_2 &= \alpha_2 + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta}_2 \\
\vdots &= \vdots \\
L_{m-1} &= \alpha_{m-1} + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta}_{m-1}
\end{aligned}$$

{ex:wlfpart1}

EXAMPLE 8.1: Women's labor force participation

The data set *Womenlf* in the car package gives the result of a 1977 Canadian survey. It

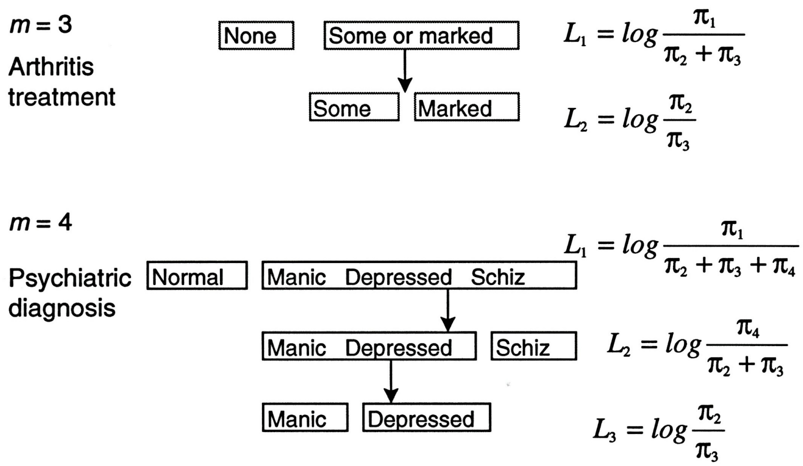


Figure 8.9: Examples of nested dichotomies and the corresponding logits {fig:nested1}

contains data for 263 married women of age 21–30 who indicated their working status (outside the home) as not working, working part time or working full time, together with their husband's income and a binary indicator of whether they had one or more young children in their household. (Another variable, region of Canada, had no effects in these analyses, and is not examined here.) This example follows Fox and Weisberg (2011, §5.8).

```
> library(car)    # for data and Anova()
> data("Womenlf", package = "car")
> some(Womenlf)

      partic hincome children   region
5   not.work      19  present  Ontario
89  not.work      35   absent  Ontario
98  fulltime      15   absent  Ontario
159 fulltime      22   absent  Ontario
163 fulltime       5  present Atlantic
199 fulltime      10   absent   Quebec
229 parttime      23  present   Quebec
236 fulltime       6   absent   Quebec
254 parttime      23  present   Quebec
256 not.work       9   absent   Quebec
```

In this example, it makes sense to consider a first dichotomy (`working`) between women who are not working, vs. those who are (full time or part time). A second dichotomy (`fulltime`) contrasts full time work vs. part time work, among those women who are working at least part time. These two binary variables are created in the data frame using the `recode()` function from the car package.

```
> # create dichotomies
> 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
39  not.work       9  present Atlantic     <NA>      no
66  not.work      15  present  Ontario     <NA>      no
```

```
77  parttime      38  present  Ontario       no     yes
113 not.work       1  present Atlantic     <NA>      no
115 parttime      13  present  Prairie       no     yes
164 fulltime       1   absent  Prairie      yes     yes
221 not.work       5   absent   Quebec     <NA>      no
228 fulltime       5   absent   Quebec      yes     yes
241 not.work      13  present   Quebec     <NA>      no
263 not.work      15  present   Quebec     <NA>      no
```

The tables below show how the response `partic` relates to the recoded binary variables, `working` and `fulltime`. Note that the `fulltime` variable is recoded to `NA` for women who are not working.

```
> with(Womenlf, table(partic, working))

          working
partic      no yes
  fulltime   0  66
  not.work 155   0
  parttime   0  42

> with(Womenlf, table(partic, fulltime, useNA = "ifany"))

          fulltime
partic      no yes <NA>
  fulltime   0  66    0
  not.work   0   0  155
  parttime  42   0    0
```

We proceed to fit two separate binary logistic regression models for the derived dichotomous variables. For the `working` dichotomy, we get the following results:

```
> mod.working <- glm(working ~ hincome + children, family = binomial,
+                    data = Womenlf)
> summary(mod.working)


Call:
glm(formula = working ~ hincome + children, family = binomial,
    data = Womenlf)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.677  -0.865  -0.777   0.929   1.997

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       1.3358     0.3838    3.48   0.0005 ***
hincome          -0.0423     0.0198   -2.14   0.0324 *
childrenpresent  -1.5756     0.2923   -5.39    7e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 356.15  on 262  degrees of freedom
Residual deviance: 319.73  on 260  degrees of freedom
AIC: 325.7

Number of Fisher Scoring iterations: 4
```

And, similarly for the `fulltime` dichotomy:

```
> mod.fulltime <- glm(fulltime ~ hincome + children, family = binomial,
+                     data = Womenlf)
> summary(mod.fulltime)


Call:
glm(formula = fulltime ~ hincome + children, family = binomial,
    data = Womenlf)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.405  -0.868   0.395   0.621   1.764

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.4778     0.7671    4.53  5.8e-06 ***
hincome          -0.1073     0.0392   -2.74   0.0061 **
childrenpresent  -2.6515     0.5411   -4.90  9.6e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144.34  on 107  degrees of freedom
Residual deviance: 104.49  on 105  degrees of freedom
  (155 observations deleted due to missingness)
AIC: 110.5

Number of Fisher Scoring iterations: 5
```

Although these were fit separately, we can view this as a combined model for the three-level response, with the following coefficients:

```
> cbind(working = coef(mod.working), fulltime = coef(mod.fulltime))

                  working fulltime
(Intercept)      1.335830  3.47777
hincome         -0.042308 -0.10727
childrenpresent -1.575648 -2.65146
```

Writing these out as equations for the logits, we have:

{eq:wlf-logits}

$$L_1 = \log \frac{\Pr(\text{working})}{\Pr(\text{notworking})} = 1.336 - 0.042\,\text{hincome} - 1.576\,\text{children} \tag{8.7}$$

$$L_2 = \log \frac{\Pr(\text{fulltime})}{\Pr(\text{parttime})} = 3.478 - 0.1072\,\text{hincome} - 2.652\,\text{children} \tag{8.8}$$

For both dichotomies, increasing income of the husband and the presence of young children decrease the log odds of a greater level of work. However, for those women who are working the effects of husband's income and and children are greater on the choice between full time and part time work than they are for all women on the choice between working and not working.

As we mentioned above, the use of nested dichotomies implies that the models fit to the separate dichotomies are statistically independent. Thus, we can additively combine χ^2 statistics and degrees of freedom to give overall tests for the polytomous response.

For example, here we define a function, `LRtest()` to calculate the likelihood ratio test of the hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}$ for all predictors simultaneously. We then use this to display these tests for each sub-model, as well as the combined model based on the sums of the test statistic and degrees of freedom.

```
> LRtest <- function(model)
+   c(LRchisq = model$null.deviance - model$deviance,
+     df = model$df.null - model$df.residual)
>
> tab <- rbind(working = LRtest(mod.working),
+              fulltime = LRtest(mod.fulltime))
> tab <- rbind(tab, All = colSums(tab))
> tab <- cbind(tab, pvalue = 1- pchisq(tab[,1], tab[,2]))
> tab

         LRchisq df     pvalue
working   36.418  2 1.2355e-08
fulltime  39.847  2 2.2252e-09
All       76.265  4 1.1102e-15
```

Similarly, you can carry out tests of individual predictors, $H_0 : \boldsymbol{\beta}_i = \mathbf{0}$ for the polytomy by adding the separate χ^2s from `Anova()`.

```
> Anova(mod.working)

Analysis of Deviance Table (Type II tests)

Response: working
         LR Chisq Df Pr(>Chisq)
hincome   4.82637  1   0.028028 *
children 31.32288  1 2.1849e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> Anova(mod.fulltime)

Analysis of Deviance Table (Type II tests)

Response: fulltime
         LR Chisq Df Pr(>Chisq)
hincome    8.9813  1  0.0027275 **
children  32.1363  1 1.4373e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For example, the test for husband's income gives $\chi^2 = 4.826 + 8.981 = 13.807$ with 2 df.

As before, you can plot the fitted values from such models, either on the logit scale (for the separate logit equations) or in terms of probabilities for the various responses. The general idea is the same: obtain the fitted values from `predict()` using data frame containing the values of the predictors. However, now we have to combine these for each of the sub-models.

We calculate these values below, on both the logit scale and the response scale of probabilities. The `newdata` argument to `predict()` is constructed as the combinations of values for `hincome` and `children`.[5]

```
> predictors <- expand.grid(hincome = 1 : 50,
+                           children =c('absent', 'present'))
> fit <- data.frame(predictors,
+     p.working = predict(mod.working, predictors, type = "response"),
+     p.fulltime = predict(mod.fulltime, predictors, type = "response"),
+     l.working = predict(mod.working, predictors, type = "link"),
+     l.fulltime = predict(mod.fulltime, predictors, type = "link")
+ )
> print(some(fit, 5), digits = 3)
```

[5] Alternatively, using the predictor values in the *Womenlf* data would give the fitted values for the cases in the data, and allow a more data-centric plot as shown in Figure 8.4.

```
   hincome children p.working p.fulltime l.working l.fulltime
10      10   absent     0.714      0.917     0.913      2.405
38      38   absent     0.432      0.355    -0.272     -0.598
71      21  present     0.244      0.194    -1.128     -1.426
72      22  present     0.237      0.177    -1.171     -1.534
75      25  present     0.215      0.135    -1.298     -1.855
```

One wrinkle here is that the probabilities for working full time and part time are conditional on working. We calculate the unconditional probabilities as shown below and choose to display the probability of *not* working as the complement of working.

```
> fit <- within(fit, {
+   full <- p.working * p.fulltime
+   part <- p.working * (1 - p.fulltime)
+   not  <- 1 - p.working
+   })
```

Plotting these fitted values using ggplot2 (Wickham and Chang, 2015) would require reshaping the `fit` data frame from wide to long format. Instead, we use R base graphics to produce plots of the probabilities and log odds. This method doesn't automatically give plots in separate panels, so a `for`-loop is used to generate panels for the levels of `children`. We set up an empty plot frame (`type="n"`) for each panel and then use `lines()` to plot the fitted probabilities. Using `par(mfrow=c(1,2))` places these plots in two side-by-side panels in a single display. The lines below give the plot shown in Figure 8.10.

TODO: DM: this code is quite complicated. Why not using melt + ggplot2? – MF: Pls fix!

```
> op <- par(mfrow=c(1,2), mar=c(5,4,4,1)+.1)
> Hinc <- 1:max(fit$hincome)
> for ( kids in c("absent", "present") ) {
+   dat <- subset(fit, children==kids)
+   plot( range(Hinc), c(0,1), type="n", cex.lab=1.25,
+         xlab="Husband's Income", ylab='Fitted Probability',
+         main = paste("Children", kids))
+   lines(Hinc, dat$not,  lwd=3, col="black", lty=1)
+   lines(Hinc, dat$part, lwd=3, col="blue",  lty=2)
+   lines(Hinc, dat$full, lwd=3, col="red",   lty=5)
+   if (kids=="absent") {
+     legend("topright", lty=c(1,2,5), lwd=3, col=c("black", "blue", "red"),
+            legend=c('not working', 'part-time', 'full-time'))
+     }
+ }
> par(op)
```

We can see how that the decision not to work outside the home increases strongly with husband's income, and is higher when there are children present. As well, among working women, the decision to work full time as opposed to part time decreases strongly with husband's income, and is less likely with young children.

Similarly, we plot the fitted logits for the two dichotomies in `l.working` and `l.fulltime` as shown below, giving Figure 8.11.

```
> op <- par(mfrow=c(1,2), mar=c(5,4,1,1)+.1)
> for ( kids in c("absent", "present") ) {
+   dat <- subset(fit, children==kids)
+   plot( range(Hinc), c(-4,5), type="n", cex.lab=1.25,
+         xlab="Husband's Income", ylab='Fitted log odds')
+   lines(Hinc, dat$l.working,  lwd=3, col="black", lty=1)
+   lines(Hinc, dat$l.fulltime, lwd=3, col="blue",  lty=2)
+   text(25, 4.5, paste("Children", kids), cex=1.4)
+   if (kids=="absent") {
```

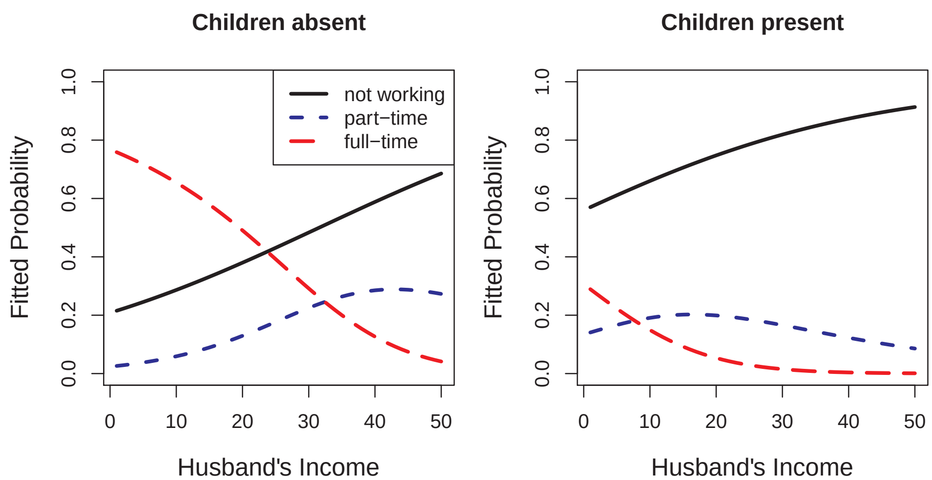


Figure 8.10: Fitted probabilities from the models for nested dichotomies fit to the data on women's labor force participation.
{fig:wlf-fitted-prob}

```
+     legend("bottomleft", lty=1:2, lwd=3, col=c("black", "blue"),
+            legend=c('working', 'full-time'))
+     }
+ }
> par(op)
```

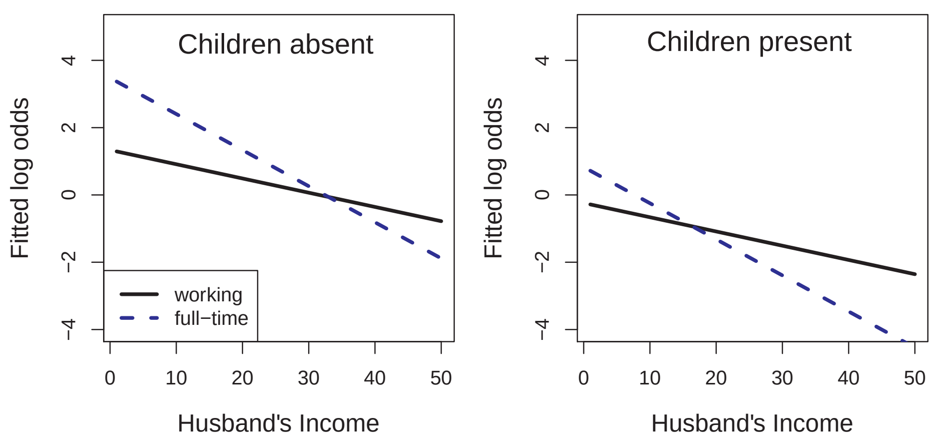


Figure 8.11: Fitted log odds from the models for nested dichotomies fit to the data on women's labor force participation.
{fig:wlf-fitted-logit}

This is essentially a graph of the fitted equations for L_1 and L_2 shown in Eqn. (8.7). It shows how the choice of full time work as opposed to part time depends more strongly on husband's income among women who are working than does the choice of working at all among all women. It also illustrates why the proportional odds assumption would not be reasonable for this data: that would require equal slopes for the two lines within each panel.

△

8.3 Generalized logit model

{sec:genlogit}

The generalized logit (or multinomial logit) approach models the probabilities of the m response categories directly as a set of $m-1$ logits. These compare each of the first $m-1$ categories to the last category, which serves as the baseline.[6] The logits for any other pair of categories can be retrieved from the $m-1$ fitted ones.

When there are p predictors, $x_1, x_2, \ldots, x_p$, which may be quantitative or categorical, the generalized logit model expresses the logits as

{eq:glogit1}

$$
\begin{aligned}
L_{jm} \equiv \log \frac{\pi_{ij}}{\pi_{im}} &= \beta_{0j} + \beta_{1j}\, x_{i1} + \beta_{2j}\, x_{i2} + \cdots + \beta_{kj}\, x_{ip} \quad j = 1, \ldots, m-1 \\
&= \boldsymbol{x}_i{}^{\mathsf{T}} \boldsymbol{\beta}_j \qquad (8.9)
\end{aligned}
$$

Thus, there is one set of fitted coefficients, $\boldsymbol{\beta}_j$ for each response category except the last. Each coefficient, β_{hj}, gives the effect, for a unit change in the predictor x_h, on the log odds that an observation had a response in category $Y = j$, as opposed to category $Y = m$.

The probabilities themselves can be expressed as

$$
\begin{aligned}
\pi_{ij} &= \frac{\exp(\boldsymbol{x}_i{}^{\mathsf{T}} \boldsymbol{\beta}_j)}{1 + \sum_{\ell=1}^{m-1} \exp(\boldsymbol{x}_i{}^{\mathsf{T}} \boldsymbol{\beta}_j)} \qquad j = 1, 2, \ldots m-1 \\
\pi_{im} &= 1 - \sum_{i=1}^{m-1} \pi_{ij} \qquad \text{for } Y = m
\end{aligned}
$$

Parameters in the $m-1$ equations Eqn. (8.9) can be used to determine the probabilities or the predicted log odds for any pair of response categories by subtraction. For instance, for an arbitrary pair of categories, a and b, and two predictors, x_1 and x_2,

$$
\begin{aligned}
L_{ab} &= \log \frac{\pi_{ia}/\pi_{im}}{\pi_{ib}/\pi_{im}} \\
&= \log \frac{\pi_{ia}}{\pi_{im}} - \log \frac{\pi_{ib}}{\pi_{im}} \\
&= (\beta_{0a} - \beta_{0b}) + (\beta_{1a} - \beta_{1b}) x_{i1} + (\beta_{2a} - \beta_{2b}) x_{i2}
\end{aligned}
$$

For example, the coefficient for x_{i1} in L_{ab} is just $(\beta_{1a} - \beta_{1b})$. Similarly, the predicted logit for any pair of categories can be calculated as

$$
\hat{L}_{ab} = \hat{L}_{am} - \hat{L}_{bm} \ .
$$

The generalized logit model can be fit most conveniently in R using the function `multinom()` in the nnet package and the effects package has a set of methods for "multinom" models. These models can also be fit using VGAM and the mlogit (Croissant, 2013) package.

{ex:wlfpart2}

EXAMPLE 8.2: Women's labor force participation

To illustrate this method, we fit the generalized logit model to the women's labor force participation data as explained below. The response, `partic` is a character factor, and, by default `multinom()` treats these in alphabetical order and uses the *first* level as the baseline category.

[6]When the response is a factor, any category can be selected as the baseline level using `relevel()`.

```
> levels(Womenlf$partic)

[1] "fulltime" "not.work" "parttime"
```

Although the multinomial model does not depend on the baseline category, it makes interpretation easier to choose `"not.work"` as the reference level, which we do with `relevel()`.[7]

```
> # choose not working as baseline category
> Womenlf$partic <- relevel(Womenlf$partic, ref = "not.work")
```

We fit the main effects model for husband's income and children as follows. As we did with `polr()` (Section 8.1), specifying `Hess=TRUE` saves the Hessian and facilitates calculation of standard errors and hypothesis tests.

```
> library(nnet)
> wlf.multinom <- multinom(partic ~ hincome + children,
+                          data = Womenlf, Hess = TRUE)

# weights:  12 (6 variable)
initial  value 288.935032
iter  10 value 211.454772
final  value 211.440963
converged
```

The `summary()` method for "multinom" objects doesn't calculate test statistics for the estimated coefficients by default. The option `Wald=TRUE` produces Wald z-test statistics, calculated as $z = \beta/SE(\beta)$.

```
> summary(wlf.multinom, Wald = TRUE)

Call:
multinom(formula = partic ~ hincome + children, data = Womenlf,
    Hess = TRUE)

Coefficients:
         (Intercept)    hincome childrenpresent
fulltime      1.9828 -0.0972321       -2.558605
parttime     -1.4323  0.0068938        0.021456

Std. Errors:
         (Intercept)  hincome childrenpresent
fulltime     0.48418 0.028096         0.36220
parttime     0.59246 0.023455         0.46904

Value/SE (Wald statistics):
         (Intercept)  hincome childrenpresent
fulltime      4.0953 -3.46071       -7.064070
parttime     -2.4176  0.29392        0.045744

Residual Deviance: 422.88
AIC: 434.88
```

Notice that the coefficients, their standard errors and the Wald test z values are printed in separate tables. The first line in each table pertains to the logit comparing full time work with the not working reference level; the second line compares part time work against not working.

For those who like p-values for significance tests, you can calculate these from the results returned by the `summary()` method in the `Wald.ratios` component, using the standard normal asymptotic approximation:

[7] Alternatively, we could declare `partic` an *ordered* factor, using `ordered()`.

```
> stats <- summary(wlf.multinom, Wald = TRUE)
> z <- stats$Wald.ratios
> p <- 2 * (1 - pnorm(abs(z)))
> zapsmall(p)

         (Intercept) hincome childrenpresent
fulltime     0.00004 0.00054         0.00000
parttime     0.01562 0.76882         0.96351
```

The interpretation of these tests is that both husband's income and presence of children have highly significant effects on the comparison of working full time as opposed to not working, while neither of these predictors are significant for the comparison of working part time vs. not working.

So far, we have assumed that the effects of husband's income and presence of young children are additive on the log odds scale. We can test this assumption by allowing an interaction of those effects and testing it for significance.

```
> wlf.multinom2 <- multinom(partic ~ hincome * children,
+                           data = Womenlf, Hess = TRUE)

# weights:  15 (8 variable)
initial  value 288.935032
iter  10 value 210.797079
final  value 210.714841
converged

> Anova(wlf.multinom2)

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 ***
hincome:children      1.5  2    0.48378
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test for the interaction term, `hincome:children` is not significant, so we can abandon this model.

Full model plots of the fitted values can be plotted as shown earlier in Example 8.1: obtain the fitted values over a grid of the predictors and plot these.

```
> predictors <- expand.grid(hincome = 1 : 50,
+                           children = c("absent", "present"))
> fit <- data.frame(predictors,
+                   predict(wlf.multinom, predictors, type = "probs")
+                   )
```

Plotting these fitted values gives the plot shown in Figure 8.12.

```
> op <- par(mfrow=c(1,2), mar=c(5,4,4,1)+.1)
> Hinc <- 1:max(fit$hincome)
> for ( kids in c("absent", "present") ) {
+   dat <- subset(fit, children==kids)
+         plot( range(Hinc), c(0,1), type="n", cex.lab=1.25,
+                 xlab="Husband's Income", ylab='Fitted Probability',
+                 main = paste("Children", kids))
+         lines(Hinc, dat$not.work, lwd=3, col="black", lty=1)
+         lines(Hinc, dat$parttime, lwd=3, col="blue",  lty=2)
+         lines(Hinc, dat$fulltime, lwd=3, col="red",   lty=5)
```

```
+   if (kids=="absent") {
+   legend("topright", lty=c(1,2,5), lwd=3, col=c("black", "blue", "red"),
+     legend=c('not working', 'part-time', 'full-time'))
+     }
+ }
> par(op)
```

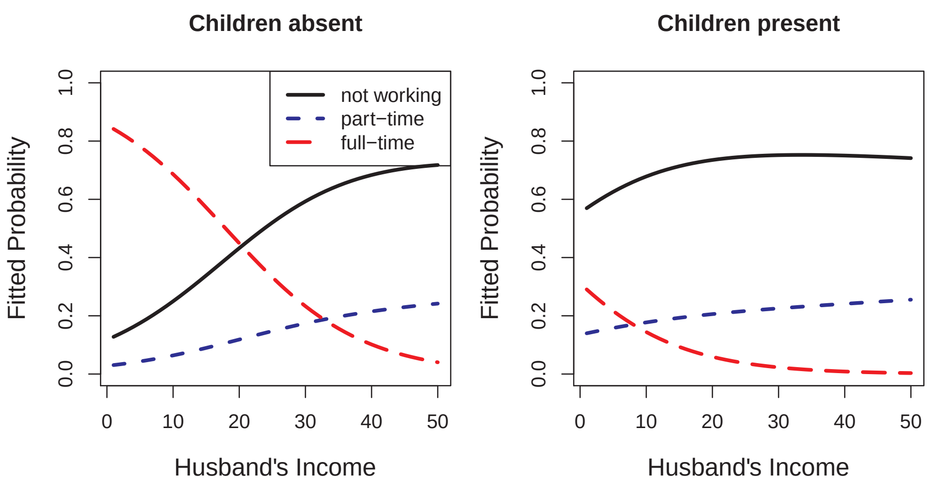


Figure 8.12: Fitted probabilities from the generalized logit model fit to the data on women's labor force participation. {fig:wlf-multi-prob}

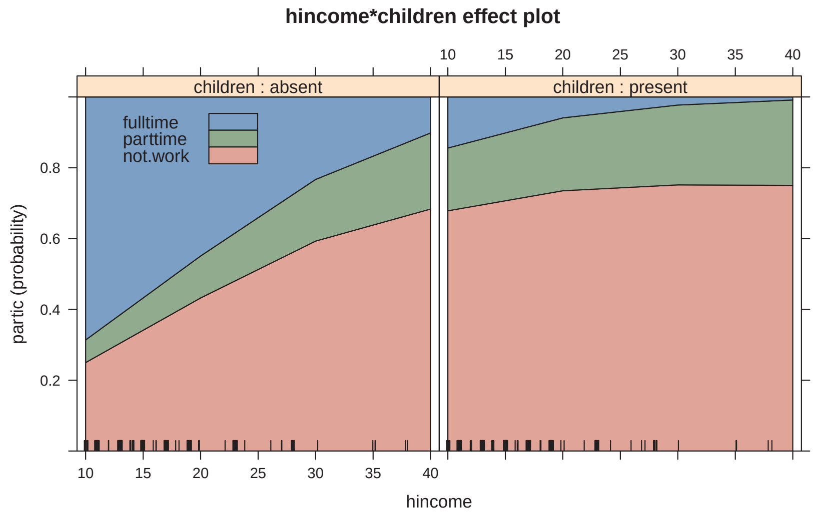
The results shown in this plot are roughly similar to those obtained from the nested dichotomy models, graphed in Figure 8.10. However, the predicted probabilities of not working under the generalized logit model rise more steeply with husband's income for women with no children and level off sooner for women with young children.

The effects package has special methods for "multinom" models. It treats the response levels in the order given by `levels()`, so before plotting we use `ordered()` to arrange levels in their natural order. The `update()` method provides a simple way to get a new fitted model; in the call, the model formula `. ~ .` means to fit the same model as before, i.e., `partic ~ hincome + children`.

```
> levels(Womenlf$partic)

[1] "not.work" "fulltime" "parttime"

> Womenlf$partic <- ordered(Womenlf$partic,
+                           levels=c("not.work", "parttime", "fulltime"))
> wlf.multinom <- update(wlf.multinom, . ~ .)

# weights:  12 (6 variable)
initial  value 288.935032
iter  10 value 211.454772
final  value 211.440963
converged
```

As illustrated earlier, you can use `plot(allEffects(model), ...)` to plot all the high-order terms in the model, either with separate curves for each response level (`style="lines"`) or as cumulative filled polygons (`style="stacked"`). Here, we simply plot the effects for the combinations of husband's income and children in stacked style, giving a plot (Figure 8.13) that is analogous to the full-model plot shown in Figure 8.12.

```
> plot(Effect(c("hincome", "children"), wlf.multinom),
+      style = "stacked", key.args = list(x = .05, y = .9))
```



{fig:wlf-multi-effect} **Figure 8.13:** Effect plot for the probabilities of not working and working part time and full time from the generalized logit model fit to the women's labor force data.

△

8.4 Chapter summary

{sec:ch08-summary}

- Polytomous responses may be handled in several ways as extensions of binary logistic regression. These methods require different fitting functions in R, however the graphical methods for plotting results are relatively straight-forward extensions of those used for binary responses.
- The *proportional odds model* (Section 8.1) is simple and convenient, but its validity depends on an assumption of equal slopes for adjacent-category logits.
- *Nested dichotomies* (Section 8.2) among the response categories give a set of statistically independent, binary logistic submodels. These may be regarded as a single, combined model for the polytomous response.
- *Generalized logit models* (Section 8.3) provide the most general approach. These may be used to construct submodels comparing any pair of categories.

8.5 Lab exercises

{sec:ch08-exercises} {lab:8.1}

Exercise 8.1 For the women's labor force participation data (*`Womenlf`*) the response variable, `partic`, can be treated as ordinal by using

```
> Womenlf$partic <- ordered(Womenlf$partic,
+                           levels=c('not.work', 'parttime', 'fulltime'))
```

Use the methods in Section 8.1 to test whether the proportional odds model holds for these data.

{lab:8.2}

Exercise 8.2 The data set *housing* in the MASS package gives a $3 \times 3 \times 4 \times 2$ table in frequency form relating (a) satisfaction (`Sat`) of residents with their housing (High, Medium, Low), (b) perceived degree of influence (`Infl`) they have on the management of the property (High, Medium, Low), (c) `Type` of rental (Tower, Atrium, Apartment, Terrace), and (d) contact (`Cont`) residents have with other residents (Low, High). Consider satisfaction as the ordinal response variable.

(a) Fit the proportional odds model with additive (main) effects of housing type, influence in management and contact with neighbors to this data. (Hint: Using `polr()`, with the data in frequency form, you need to use the `weights` argument to supply the `Freq` variable.)
(b) Investigate whether any of the two-factor interactions among `Infl`, `Type` and `Cont` add substantially to goodness of fit of this model. (Hint: use `stepAIC()`, with the scope formula `~ .^2` and `direction="forward"`.)
(c) For your chosen model from the previous step, use the methods of Section 8.1.5 to plot the probabilities of the categories of satisfaction.
(d) Write a brief summary these analyses, interpreting *how* satisfaction with housing depends on the predictor variables.

{lab:8.3}

Exercise 8.3 The data *TV* on television viewing was analyzed using correspondence analysis in Example 6.4, ignoring the variable `Time` and extended in Exercise 6.9. Treating `Network` as a three-level response variable, fit a generalized logit model (Section 8.3) to explain the variation in viewing in relation to `Day` and `Time`. The *TV* data is a three-way table, so you will need to convert it to a frequency data frame first.

```
> data("TV", package="vcdExtra")
> TV.df <- as.data.frame.table(TV)
```

(a) Fit the main-effects model, `Network ~ Day + Time` with `multinom()`. Note that you will have to supply the `weights` argument because each row of `TV.df` represents the number of viewers in the `Freq` variable.
(b) Prepare an effects plot for the fitted probabilities in this model.
(c) Interpret these results in comparison to the correspondence analysis analysis in Example 6.4.

{lab:8.4}

Exercise 8.4 * Refer to Exercise 5.10 for a description of the *Vietnam* data set in vcdExtra (Friendly, 2015). The goal here is to fit models for the polytomous `response` varialble in relation to `year` and `sex`. {lab:logist-vietnam}

(a) Fit the proportional odds model to these data, allowing an interaction of `year` and `sex`.
(b) Is there evidence that the proportional odds assumption does not hold for this data set? Use the methods described in Section 8.1 to assess this.
(c) Fit the multinomial logistic model, also allowing an interaction. Use car::`Anova()` to assess the model terms.
(d) Produce an effect plot for this model and describe the nature of the interaction.
(e) Fit the simpler multinomial model in which there is no effect of year for females and the effect of year is linear for males (on the logit scale). Test whether this model is significantly worse than the general multinomial model with interaction.

References

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31(1), 419–456.

Bliss, C. I. (1934). The method of probits. *Science*, 79(2037), 38–39.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53(1), 605–634.

Croissant, Y. (2013). *mlogit: multinomial logit model*. R package version 0.2-4.

Finney, D. J. (1947). *Probit analysis*. Cambridge England: University Press.

Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks, CA: Sage, 2nd edn.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. Thousand Oaks CA: Sage, 2nd edn.

Fox, J. and Weisberg, S. (2015). *car: Companion to Applied Regression*. R package version 2.0-25/r421.

Fox, J., Weisberg, S., Friendly, M., and Hong, J. (2015). *effects: Effect Displays for Linear, Generalized Linear, and Other Models*. R package version 3.0-4/r200.

Friendly, M. (2015). *vcdExtra: vcd Extensions and Additions*. R package version 0.6-7.

Harrell, Jr, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Graduate Texts in Mathematics. New York: Springer.

Harrell, Jr., F. E. (2015). *rms: Regression Modeling Strategies*. R package version 4.3-0.

Hocking, T. D. (2013). *directlabels: Direct labels for multicolor plots in lattice or ggplot2*. R package version 2013.6.15.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttmann, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen, eds., *Studies in Social Psychology in World War II, vol. IV, Measurement and Prediction*, (pp. 362–412). Princeton, NJ: Princeton University Press.

Lazarsfeld, P. F. (1954). A conceptual introduction to latent structure analysis. In P. F. Lazarsfeld, ed., *Mathematical Thinking in the Social Sciences*, (pp. 349–387). Glencoe, IL: Free Press.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society*, 42, 109–142.

Peterson, B. and Harrell, Jr, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39, 205–217.

Ripley, B. (2015a). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. R package version 7.3-40.

Ripley, B. (2015b). *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*. R package version 7.3-9.

Thurstone, L. L. (1927). A law of comparitive judgment. *Psychological Review*, 34, 278–286.

Wickham, H. (2014). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package.* R package version 1.4.1.

Wickham, H. and Chang, W. (2015). *ggplot2: An Implementation of the Grammar of Graphics*. R package version 1.0.1.

Yee, T. W. (2015). *VGAM: Vector Generalized Linear and Additive Models*. R package version 0.9-7.