Minicurso 3
Modeling Ordinal Categorical Data

Alan Agresti
University of Florida, USA

This is copyrighted material and cannot be used other ways without author permission.
Short Course
Modeling Ordinal Categorical Data

Alan Agresti
Distinguished Professor Emeritus
University of Florida, USA

Presented for SINAPE
Porto Alegre, Brazil
July 26, 2016

© Alan Agresti, 2016
Ordinal categorical responses

- Patient quality of life (excellent, good, fair, poor)
- Pain (none, little, considerable, severe)
- Diagnostic evaluation (definitely normal, probably normal, equivocal, probably abnormal, definitely abnormal)
- Political philosophy (very liberal, slightly liberal, moderate, slightly conservative, very conservative)
- Categorization of an inherently continuous variable, such as body mass index, \( \text{BMI} = \frac{\text{weight (kg)}}{\text{height (m)}}^2 \),
  measured as (\(< 18.5, 18.5-25, 25-30, > 30\))
  for (underweight, normal weight, overweight, obese)

For ordinal response variable \( y \) with \( c \) categories, our focus is on modeling how \( P(y = j), \quad j = 1, 2, \ldots, c \), depends on explanatory variables \( x \) (categorical and/or quantitative).

The models treat \( y \) at fixed \( x \) as having a \textit{multinomial} distribution.
Focus of short course


Emphasis on concepts and examples rather than theory, derivations, or technical details.

We assume familiarity with basic categorical data methods, such as chi-squared tests and binary logistic regression.

Examples of model fitting use R, SAS, but output provided to illustrate interpretation, not to teach software.

For R for ordinal models, Thomas Yee’s VGAM library is especially useful; see [www.stat.auckland.ac.nz/~yee/VGAM](http://www.stat.auckland.ac.nz/~yee/VGAM). Also useful is detailed R tutorial by Laura Thompson to accompany *Categorical Data Analysis*, linked at R section of [www.stat.ufl.edu/~aa/cda/cda.html](http://www.stat.ufl.edu/~aa/cda/cda.html).
Do we really need special models for ordinal responses?

Why not just assign scores to the ordered categories and use ordinary regression methods (as many methodologists do)?

- With categorical data, there is nonconstant variance, so ordinary least squares (OLS) is not optimal.
- At some settings of explanatory variables, estimated mean from the model fit may fall below lowest score or above highest score.

For binary response with scores (1, 0), \( E(y) = P(y = 1) \), and this approach simplifies to linear probability model, \( P(y = 1) = \alpha + \beta'x \). This rarely works with multiple explanatory variables.

- With categorical data, we may want estimates of category probabilities rather than means.
- Regardless of fitting method or distributional assumption, ceiling effects and floor effects can cause bias in results.
Example: Floor effect (Sec. 1.3 of *OrdCDA*)

For ordinal responses, it’s often realistic to imagine an underlying continuous latent variable $y^*$. Suppose

$$y^* = 20.0 + 0.6x - 40z + \varphi$$

for $x \sim \text{uniform}(0, 100)$, $P(z = 0) = P(z = 1) = 0.5$, $\varphi \sim N(0, 10^2)$. For a random sample of size $n = 100$, suppose

$$y = 1 \text{ if } y^* \leq 20, \quad y = 2 \text{ if } 20 < y^* \leq 40, \quad y = 3 \text{ if } 40 < y^* \leq 60,$$

$$y = 4 \text{ if } 60 < y^* \leq 80, \quad y = 5 \text{ if } y^* > 80.$$ 

Suppose we fit model $y = \alpha + \beta_1 x + \beta_2 z + \beta_3 (x \cdot z) + \varphi$ to investigate effects and possible interactions. For $x < 50$ with $z = 1$, a high probability that observations fall in lowest category of $y$. Because of *floor effect*, least squares line when $z = 1$ has only half the slope of least squares line when $z = 0$. Interaction is statistically and practically significant. Such spurious effects would not occur with an ordinal model. (We’ll see why later.)
Cumulative Logit Model with Proportional Odds

(Sec. 3.2–3.5 of *OrdCDA*)

$y$ an ordinal response ($c$ categories), $x$ an explanatory variable

Model $P(y \leq j), \ j = 1, 2, \cdots, c - 1$, using logits

$$
\text{logit}[P(y \leq j)] = \log[P(y \leq j)/P(y > j)] \\
= \alpha_j + \beta x, \ j = 1, \ldots, c - 1.
$$

This is called a *cumulative logit* model.

As in ordinary logistic regression, effects described by odds ratios (comparing odds of being below vs. above any point on the response scale, so *cumulative odds ratios* are natural).

For fixed $j$, looks like ordinary logistic regression for binary response (below $j$, above $j$). See figure on next page for $c = 4$ categories.

Binary logistic regression model is special case for $c = 2$. 
Model satisfies

$$\log \frac{P(y \leq j | x_1)}{P(y > j | x_2)} = \beta(x_1 - x_2)$$

for all $j$ (Proportional odds property).

\begin{itemize}
  \item $\beta = \text{cumulative log odds ratio}$ for 1-unit increase in predictor
  \item Model assumes effect $\beta$ is identical for every “cutpoint” for cumulative probability, $j = 1, \ldots, c - 1$.
  \item For $r \times c$ table with scores $(1, 2, \ldots, r)$ for rows, $e^\beta$ is assumed uniform value for cumulative odds ratio.
  \item R functions \textit{vglm} in VGAM library and \textit{polr} (proportional odds logistic regression) in MASS library.
\end{itemize}
Example: Detecting trend in dose response

Effect of intravenous medication doses on patients with subarachnoid hemorrhage trauma (p. 207, *OrdCDA*):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>59 (28%)</td>
<td>25</td>
<td>46</td>
<td>48</td>
<td>32 (15%)</td>
</tr>
<tr>
<td>Low dose</td>
<td>48 (25%)</td>
<td>21</td>
<td>44</td>
<td>47</td>
<td>30 (16%)</td>
</tr>
<tr>
<td>Med dose</td>
<td>44 (21%)</td>
<td>14</td>
<td>54</td>
<td>64</td>
<td>31 (15%)</td>
</tr>
<tr>
<td>High dose</td>
<td>43 (22%)</td>
<td>4</td>
<td>49</td>
<td>58</td>
<td>41 (21%)</td>
</tr>
</tbody>
</table>

Model with linear effect of dose on cumulative logits for outcome (assigning scores $x = 1, 2, 3, 4$ to ordinal $x$),

$$\text{logit}[P(y \leq j)] = \alpha_j + \beta x$$

has ML estimate $\hat{\beta} = -0.176$ ($SE = 0.056$). Likelihood-ratio test of $H_0$: $\beta = 0$ has test statistic $= 9.6$ ($df = 1$, $P = 0.002$), based on twice difference in maximized log likelihoods compared to model with $\beta = 0$. 
R for modeling dose-response data, using \texttt{vglm} function in \texttt{VGAM} library:

```r
> trauma <- read.table("trauma.dat", header=TRUE)
> trauma
dose y1 y2 y3 y4 y5
1   1 59 25 46 48 32
2   2 48 21 44 47 30
3   3 44 14 54 64 31
4   4 43  4 49 58 41
> library(VGAM)
> fit <- vglm(cbind(y1,y2,y3,y4,y5) ~ dose, family=cumulative(parallel=TRUE), data=trauma)
> summary(fit)
Coefficients:

Value Std. Error z value
(Intercept):1 -0.71917 0.15881 -4.5285
(Intercept):2 -0.31860 0.15642 -2.0368
(Intercept):3  0.69165 0.15793  4.3796
(Intercept):4  2.05701 0.17369 11.8429
dose     -0.17549 0.05632 -3.1159

Residual Deviance: 18.18245 on 11 degrees of freedom
Log-likelihood: -48.87282 on 11 degrees of freedom
> fitted(fit)  # estimated multinomial response prob's
    y1  y2  y3  y4  y5
1 0.2901506 0.08878053 0.2473198 0.2415349 0.1322142
2 0.2553767 0.08321565 0.2457635 0.2619656 0.1536786
3 0.2234585 0.07701184 0.2407347 0.2808818 0.1779132
4 0.1944876 0.07043366 0.2325060 0.2975291 0.15881

> vglm(cbind(y1,y2,y3,y4,y5) ~ 1, family=cumulative(parallel=TRUE), data=trauma)  # null model
Coefficients:
  -1.1423167  -0.7459897  0.2506811  1.6064484

Degrees of Freedom: 16 Total; 12 Residual
Residual Deviance: 27.79488 Log-likelihood: -53.67903
> 1 - pchisq(2*(53.67903 - 48.87282), df=1)
[1] 0.001932658 # P-value for likelihood-ratio test of no dose effect
```
R for modeling dose-response data using *polr* in MASS library, for which response must be an ordered factor:

```r
trauma2 <- read.table("trauma2.dat", header=TRUE)

trauma2
  dose response count
  1 1  1  59
  2 1  2  25
  3 1  3  46
  ...
  20 4  5  41

y <- factor(trauma2$response)

fit.clogit <- polr(y ~ dose, data=trauma2, weight=count)

summary(fit.clogit)
```

Coefficients:

<table>
<thead>
<tr>
<th>Value</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>dose</td>
<td>0.1754816</td>
<td>3.094245</td>
</tr>
</tbody>
</table>

Intercepts:

<table>
<thead>
<tr>
<th>Value</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>112</td>
<td>-0.7192</td>
<td>-4.5256</td>
</tr>
<tr>
<td>213</td>
<td>-0.3186</td>
<td>-2.0308</td>
</tr>
<tr>
<td>314</td>
<td>0.6917</td>
<td>4.3323</td>
</tr>
<tr>
<td>415</td>
<td>2.0570</td>
<td>11.7493</td>
</tr>
</tbody>
</table>

Residual Deviance: 2461.349

```r
fitted(fit.clogit)
```

1 0.2901467 0.08878330 0.2473217 0.2415357 0.1322126
2 0.2901467 0.08878330 0.2473217 0.2415357 0.1322126
...
20 0.1944866 0.07043618 0.2325084 0.2975294 0.2050394

Note: This uses the model formula \( \logit[P(y \leq j)] = \alpha_j - \beta x \) based on a latent variable model (below), for which \( \hat{\beta} \) has opposite sign.
SAS for cumulative logit modeling of dose-response data:

```sas
data trauma;
input dose outcome count @@;
datalines;
1 1 59 1 2 25 1 3 46 1 4 48 1 5 32
2 1 48 2 2 21 2 3 44 2 4 47 2 5 30
3 1 44 3 2 14 3 3 54 3 4 64 3 5 31
4 1 43 4 2 4 4 3 49 4 4 58 4 5 41
;
proc logistic; freq count; * proportional odds cumulative logit model;
  model outcome = dose / aggregate scale=none;
```

Deviance and Pearson Goodness-of-Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>18.1825</td>
<td>11</td>
<td>1.6530</td>
<td>0.0774</td>
</tr>
<tr>
<td>Pearson</td>
<td>15.8472</td>
<td>11</td>
<td>1.4407</td>
<td>0.1469</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log L</td>
<td>2470.961</td>
<td>2461.349</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>9.6124</td>
<td>1</td>
<td>0.0019</td>
</tr>
<tr>
<td>Score</td>
<td>9.4288</td>
<td>1</td>
<td>0.0021</td>
</tr>
<tr>
<td>Wald</td>
<td>9.7079</td>
<td>1</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

Standard Wald

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-0.7192</td>
<td>0.1588</td>
<td>20.5080</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>-0.3186</td>
<td>0.1564</td>
<td>4.1490</td>
<td>0.0417</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>0.6916</td>
<td>0.1579</td>
<td>19.1795</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 4</td>
<td>1</td>
<td>2.0570</td>
<td>0.1737</td>
<td>140.2518</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dose</td>
<td>1</td>
<td>-0.1755</td>
<td>0.0563</td>
<td>9.7079</td>
<td>0.0018</td>
</tr>
</tbody>
</table>
```
Odds ratio interpretation: For dose $i + 1$, estimated odds of outcome $\leq j$ (instead of $> j$) equal $\exp(-0.176) = 0.84$ times estimated odds for dose $i$; equivalently, for dose $i + 1$, estimated odds of outcome $\geq j$ (instead of $< j$) equal $\exp(0.176) = 1.19$ times estimated odds for dose $i$.

95% confidence interval for $\exp(-\beta)$ is

$$\exp^{0.176 \pm 1.96(0.056)} = (1.07, 1.33).$$

Cumulative odds ratio for dose levels (rows) 1 and 4 equals

$$\exp^{(4-1)0.176} = 1.69,$$

95% confidence interval $\exp^{3[0.176 \pm 1.96(0.056)]} = (1.22, 2.36)$.

Goodness-of-fit statistics:

Pearson $X^2 = \cdot \frac{(\text{observed} - \text{fitted})^2}{\text{fitted}} = 15.8$, $P$-value = 0.15

deviance $G^2 = 18.2$ (formula later), $P$-value = 0.18

$df = 16 - 5 = 11$, since 16 multinomial para’s and 5 model para’s.

Model seems to fit adequately.
Any equally-spaced scores (e.g. 0, 10, 20, 30) for dose provide same fitted values and same test statistics (different $\hat{\beta}$, $SE$).

Unequally-spaced scores more natural in many cases (e.g., doses may be 0, 125, 250, 500). “Sensitivity analysis” usually shows substantive results don’t depend much on that choice, unless data highly unbalanced (e.g., Graubard and Korn 1987 show mid-rank scores may not be appropriate).

The cumulative logit model uses ordinality of $y$ without assigning category scores.

Alternative analysis treats dose as factor, using indicator variables. But double the log-likelihood increases only 0.13, $df = 2$. With $\beta_1 = 0$:

$\hat{\beta}_2 = -0.12$, $\hat{\beta}_3 = -0.32$, $\hat{\beta}_4 = -0.52$ ($SE = 0.18$ each).

Testing $H_0: \beta_2 = \beta_3 = \beta_4 = 0$ gives likelihood-ratio (LR) stat. = 9.8 ($df = 3$, $P = 0.02$).

Using ordinality often increases power (focused on $df = 1$).
R for modeling dose-response data, with dose as a qualitative factor, using the \texttt{vglm} function in the \textit{VGAM} library:

\begin{verbatim}
> attach(trauma)
> library(VGAM)

> fit2 <- vglm(cbind(y1,y2,y3,y4,y5) ~ factor(dose),
  +  family=cumulative(parallel=TRUE), data=trauma)

> summary(fit2)
Coefficients:
  Estimate Std. Error z value
(Intercept):1 -0.91880  0.13204 -6.95875
(Intercept):2 -0.51826  0.12856 -4.03122
(Intercept):3  0.49215  0.12841  3.83255
(Intercept):4  1.85785  0.14527 12.78927
factor(dose):2 -0.11756  0.17843 -0.65885
factor(dose):3 -0.31740  0.17473 -1.81649
factor(dose):4 -0.52077  0.17795 -2.92657

Residual deviance: 18.04959 on 9 degrees of freedom
Log-likelihood: -48.80638 on 9 degrees of freedom
Number of iterations: 4

> 1 - pchisq(2*(53.679 - 48.806), df=3)
[1] 0.02085  # P-value for likelihood-ratio test of no dose effect
\end{verbatim}
SAS for modeling dose-response data, with dose as a factor using a CLASS statement to create indicator predictors for first three categories:

```sas
data trauma;
input dose outcome count @@;
datalines;
  1 1 59 1 2 25 1 3 46 1 4 48 1 5 32
  2 1 48 2 2 21 2 3 44 2 4 47 2 5 30
  3 1 44 3 2 14 3 3 54 3 4 64 3 5 31
  4 1 43 4 2 4 4 3 49 4 4 58 4 5 41
;proc logistic; freq count; class dose / param=ref; * treat dose as factor;
  model outcome = dose / aggregate scale=none;

Deviance and Pearson Goodness-of-Fit Statistics Criterion

<table>
<thead>
<tr>
<th>Value</th>
<th>DF</th>
<th>Value/DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>18.0496</td>
<td>9</td>
<td>2.0055</td>
</tr>
<tr>
<td>Pearson</td>
<td>15.7881</td>
<td>9</td>
<td>1.7542</td>
</tr>
</tbody>
</table>

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 Log L</td>
<td>2470.961</td>
<td>2461.216</td>
</tr>
</tbody>
</table>

Testing Global Null Hypothesis: BETA=0

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>9.7453</td>
<td>3</td>
<td>0.0209</td>
</tr>
<tr>
<td>Score</td>
<td>9.5583</td>
<td>3</td>
<td>0.0227</td>
</tr>
<tr>
<td>Wald</td>
<td>9.8440</td>
<td>3</td>
<td>0.0199</td>
</tr>
</tbody>
</table>

Parameter Estimate Std Error Wald Chi-Square Pr > ChiSq

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Wald</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-1.4396</td>
<td>0.1416</td>
<td>103.3943</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>-1.0390</td>
<td>0.1369</td>
<td>57.6363</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>-0.0286</td>
<td>0.1317</td>
<td>0.0472</td>
<td>0.8280</td>
<td></td>
</tr>
<tr>
<td>Intercept 4</td>
<td>1</td>
<td>1.3371</td>
<td>0.1428</td>
<td>87.7207</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>dose</td>
<td>1</td>
<td>0.5208</td>
<td>0.1779</td>
<td>8.5641</td>
<td>0.0034</td>
<td></td>
</tr>
<tr>
<td>dose</td>
<td>2</td>
<td>0.4032</td>
<td>0.1820</td>
<td>4.9072</td>
<td>0.0267</td>
<td></td>
</tr>
<tr>
<td>dose</td>
<td>3</td>
<td>0.2034</td>
<td>0.1779</td>
<td>1.3071</td>
<td>0.2529</td>
<td></td>
</tr>
</tbody>
</table>
```
Other properties of cumulative logit models

Model extends to multiple explanatory variables,

$$\text{logit}[P(y \leq j)] = \alpha_j + \beta_1x_1 + \cdots + \beta_p x_p$$

that can be qualitative (i.e., factors) or quantitative (use indicator variables for factors).

For subject $i$ with values $x_i$ on a set of explanatory variables, estimated conditional distribution function is

$$\hat{P}(y_i \leq j) = \frac{\exp(\hat{\alpha}_j + \hat{\beta} x_i)}{1 + \exp(\hat{\alpha}_j + \hat{\beta} x_i)}.$$

Estimated probability of outcome $j$ is

$$\hat{P}(y_i = j) = \hat{P}(y_i \leq j) - \hat{P}(y_i \leq j - 1).$$

\[ y^* = \beta' x + \varphi \text{ where } \varphi \text{ has cdf } G \text{ with mean 0.} \]

Thresholds (cutpoints)
\[-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_c = \infty \text{ such that} \]
\[ y = j \text{ if } \alpha_{j-1} < y^* \leq \alpha_j. \]

Ex. earlier, p. 6. Then, at fixed \( x \) (see figure on next page),
\[
P(y \leq j) = P(y^* \leq \alpha_j) = P(y^* - \beta x \leq \alpha_j - \beta x)
\]
\[
= P(\varphi \leq \alpha_j - \beta x) = G(\alpha_j - \beta x).
\]

\[ \rightarrow \text{Model } \quad G^{-1}[P(y \leq j \mid x)] = \alpha_j - \beta' x \]

with \( G^{-1} \) a link function. Get cumulative logit model when \( G = \text{logistic cdf} \) \((G^{-1} = \text{logit})\). So, cumulative logit model fits well when regression model holds for underlying logistic response.

Note: The model is often expressed as
\[ \logit[P(y \leq j)] = \alpha_j - \beta' x. \]

Then, \( \beta_j > 0 \) has usual interpretation of ‘positive’ effect.
Note: This derivation suggests such models are designed to detect shifts in *location* (center), not dispersion (spread), at different settings of explanatory variables.
Can use similar model with alternative “cumulative link”

\[ \text{link}[P(y_i \leq j)] = \alpha_j - \beta x_i \]

of cumulative prob.'s (McCullagh 1980).

e.g., *cumulative probit* model (link function = inverse of standard normal cdf) applies naturally when underlying regression model has normal \( y^* \).

Effects \( \beta \) invariant to choice and number of response categories (If model holds for given response categories, holds with same \( \beta \) when response scale collapsed in any way).
Likelihood function for the model

For subject $i$, let $(y_{i1}, \ldots, y_{ic})$ be binary indicators of the response, where $y_{ij} = 1$ when response in category $j$.

For independent multinomial observations at values $x_i$ of the explanatory variables for subject $i$, the likelihood function is

\[
\sum_{j=1}^{n_c} \left( \prod_{i=1}^{n_c} \left( \frac{\exp(\alpha_j + \beta x_i)}{1 + \exp(\alpha_j + \beta x_i)} \right)^{y_{ij}} - \frac{\exp(\alpha_{j-1} + \beta x_i)}{1 + \exp(\alpha_{j-1} + \beta x_i)} \right) \cdot \left( \prod_{i=1}^{n_c} \left( \frac{\exp(\alpha_j + \beta x_i)}{1 + \exp(\alpha_j + \beta x_i)} \right)^{y_{ij}} - \frac{\exp(\alpha_{j-1} + \beta x_i)}{1 + \exp(\alpha_{j-1} + \beta x_i)} \right).
\]

A. Agresti (UF)
Model fitting requires iterative methods. Log likelihood is concave (Pratt 1981). To get standard errors, Newton-Raphson inverts observed information matrix 
\[-\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b}.
\]
Fisher scoring inverts expected information matrix 
\[E \left( -\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b} \right),\] such as with vglm function in R.


Inference uses standard methods for testing \(H_0: \beta_j = 0\) (likelihood-ratio, Wald, score tests) and inverting tests of \(H_0: \beta_j = \beta_{j0}\) to get confidence intervals for \(\beta_j\). 
\[
\hat{\beta}_j - \beta_{j0}
\]
Wald: \(z = \frac{\hat{\beta}_j - \beta_{j0}}{SE}\), or \(z^2 \sim \chi^2\), can behave poorly with extremely large estimates (and useless when \(\hat{\beta}_j = \infty\)).

Likelihood-ratio: 
\[-2[L(\hat{\beta}_0) - L(\hat{\beta})] \sim \chi^2\]
In general, explanatory variables can be categorical and/or quantitative; the data file typically has observations at the subject level rather than counts from a contingency table.

Example: Study of effects associated with mental health

\[ y = \text{mental impairment} \]
(1=well, 2=mild impairment, 3=moderate impairment, 4=impaired)

\[ x_1 = \text{number of “life events”} \]

\[ x_2 = \text{socioeconomic status (1 = high, 0 = low)} \]

Entire data file (\( n = 40 \)) is available at

www.stat.ufl.edu/˜aa/glm/data/Mental.dat

<table>
<thead>
<tr>
<th>subject</th>
<th>impair</th>
<th>SES</th>
<th>life</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>39</td>
<td>4</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
R (vglm in VGAM library) for modeling mental impairment with subject-level data file.

```r
> Mental <- read.table("Mental.dat", header=TRUE)

> Mental
  impair  ses  life
  1     1     1
  2     1     1
  3     1     1
  4     1     1
  5     1     0
...
40     4     0     9

> library(VGAM)

> fit <- vglm(impair ~ ses + life,
  family=cumulative(parallel=TRUE), data=Mental)

> summary(fit)

Coefficients:

  Estimate Std. Error   z value
(Intercept):1 -0.28176  0.62304  -0.45223
(Intercept):2  1.21291  0.65119   1.86260
(Intercept):3  2.20947  0.71719   3.08075
ses      1.11112  0.61427   1.80884
life     -0.31888  0.11944  -2.66973

Residual deviance: 99.0979 on 115 degrees of freedom

Log-likelihood: -49.54895 on 115 degrees of freedom
```
Alternative ways of summarizing effects

Some researchers find odds ratios difficult to interpret.

Can compare probabilities or cumulative prob’s for $y$ directly, such as comparing $\hat{P}(y = 1)$ or $\hat{P}(y = c)$ at maximum and minimum values of a predictor (at means of other predictors).

ex.: At mean life events of 4.3, $\hat{P}(y = 1) = 0.37$ at high SES and $\hat{P}(y = 1) = 0.16$ at low SES.

For high SES, $\hat{P}(y = 1) = 0.70$ at $x_1 = \text{min} = 0$ and $\hat{P}(y = 1) = 0.12$ at $x_1 = \text{max} = 9$.

For low SES, $\hat{P}(y = 1) = 0.43$ at $x_1 = \text{min}$ and $\hat{P}(y = 1) = 0.04$ at $x_1 = \text{max}$.
Summary measures of predictive power include

1. **concordance index** (probability that observations with different outcomes are concordant with predictions).

2. **correlation** between $y$ and estimated mean of conditional distribution of $y$ from model fit, based on scores $\{v_j\}$ for $y$ (mimics multiple correlation, Sec. 3.4.6 of *OrdCDA*).

3. $R^2$ for regression model for underlying latent response variable (McKelvey and Zavoina 1975, provided by Stata).

4. To compare two groups with responses represented by independent r.v.’s $y_1$ and $y_2$, at fixed setting of other explanatory variables, the effect size measure

   $$\gamma = P(y_1 > y_2) + 2^{-1}P(y_1 = y_2)$$ (Agresti and Kateri 2016).
Alternative ways of summarizing effects (continued)

For the cumulative link model, a useful measure is $\gamma = P(y_1^* > y_2^*)$ for the underlying latent variables for the two groups.

For latent normal linear model, $\gamma = \Phi(\beta / 2)$ for $\beta$ coefficient of indicator for group variable in cumulative probit model with latent variable parameterization ($-\beta$), at every fixed setting for explanatory var’s.

For cumulative logit model (underlying logistic latent variables),

$\gamma \approx \exp(\beta / \sqrt{2}) / [1 + \exp(\beta / 2)]$ for the $\beta$ group coefficient.

Example: For cumulative logit model predicting mental impairment using SES (low, high) and life events, $\hat{\beta} = -1.111$ for effect of SES.

The value $\hat{\gamma} = 0.317 \approx \exp(\hat{\beta} / 2) / [1 + \exp(\hat{\beta} / 2)] = 0.313$. At any particular value for life events, there is about a 1/3 chance of lower mental impairment at low SES than at high SES.

A 95% confidence interval for $\beta$ generates a confidence interval for $\gamma$ of (0.16, 0.51).
Sample size for comparing two groups

Want power $1 - \beta$ in $\alpha$-level test for effect of size $\beta_0$ in proportional odds version of cumulative logit model. Let $\{\pi_j\} = \text{marginal probabilities for } y$. For two-sided test with equal group sample sizes, need approximately (Whitehead 1993, Sec. 3.7.2 of OrdCDA)

$$n = \frac{12(z_{\alpha/2} + z_\beta)^2}{\beta^2 \left(1 - \sum_j \pi_j^3\right)}.$$ 

When $\{\pi_j = 1/c\}$, sample size $n(c)$ needed for $c$ categories satisfies

$$n(c)/n(2) = 0.75/[1 - 1/c^2].$$

Relative to continuous response ($c = \infty$), using $c$ categories has efficiency $(1 - 1/c^2)$. Substantial loss of information from collapsing to binary response (often done in medical research, but results in measurement error and loss of power), but little gain with $c$ more than about 5.

McCullagh (1980): Score test of no effect in comparing two groups with this model is equivalent to grouped-data version of Wilcoxon test.
Checking goodness of fit with non-sparse contingency tables

With non-sparse contingency table data, can check goodness of fit using Pearson $X^2$, deviance $G^2$ comparing observed cell counts to expected frequency estimates.

At setting $x_i$ of predictors with $n_i = \sum_{j=1}^{c} n_{ij}$ multinomial observations, expected frequency estimates equal

$$\hat{\mu}_{ij} = n_i \hat{P}(y = j), \quad j = 1, \ldots, c.$$ 

Pearson test statistic is

$$X^2 = \sum_{i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}.$$ 

Deviance, which is likelihood-ratio test statistic for testing that model holds against unrestricted alternative, is

$$G^2 = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{\mu}_{ij}}.$$ 

$df =$ No. multinomial parameters $-$ no. model parameters.
Checking goodness of fit with sparse data

With sparse data and/or continuous explanatory variables, these statistics are not valid for tests of fit (not chi-squared), but can use $G^2$ to compare nested models.

Can check particular aspects of fit using likelihood-ratio test to compare to more complex models. Test statistic $= 2 \times \text{increase in maximized log likelihood}$.

Some software provides score test of proportional odds assumption, by comparing model to more general “non-proportional odds model” with effects $\{\beta_j\}$. But this test is liberal (i.e., $P($Type I error$)$ too high).

LR test also possible comparing model to non-proportional odds model, except when more general model has cumulative probabilities out-of-order.
Lack of fit may result from omitted predictors (e.g., interaction between predictors) or violation of proportional odds assumption (such as when there are dispersion as well as location effects).

When model with proportional odds structure fails, we can use estimated effects in non-proportional odds model (e.g., after fitting binary logistic to each collapsing) to describe effects more fully.

Even if the model has lack of fit, it may usefully summarize “first-order effects” and have good power for testing $H_0$: no effect, because of its parsimony (e.g., later examples).
Other criteria besides significance tests can help select a good model, such as by minimizing

\[ \text{AIC} = -2(\text{log likelihood} - \text{number of parameters in model}) \]

which penalizes a model for having many parameters. This attempts to find model for which fit is “closest” to reality, and overfitting (too many parameters) can hurt this.

Advantages of utilizing ordinality of response include:

No interval-scale assumption about distances between response categories (i.e., no scores needed for \( y \)).

Greater variety of models, including ones that are more parsimonious than models that ignore ordering (such as baseline-category logit models, which have a separate set of parameters for each logit).

Greater statistical power for testing effects (compared to treating categories as nominal), because of focusing effect on smaller \( df \).
Cumulative logit models without proportional odds

Generalized model permits effects of explanatory variables to differ for different cumulative logits (Sec. 3.6 of *OrdCDA*),

\[
\text{logit}[P(y_i \leq j)] = \alpha_j + \beta' x_i, \quad j = 1, \ldots, c - 1.
\]

Each predictor has \(c - 1\) parameters, allowing different effects for logit\([P(y_i \leq 1), \text{logit}[P(y_i \leq 2)], \ldots, \text{logit}[P(y_i \leq c - 1)]\].

Even if this model fits better, for reasons of parsimony a simple model with proportional odds structure is sometimes preferable.

- Effects of explanatory variables easier to summarize and interpret, and still describe overall trends. Is variability in effects great enough to make it worthwhile to use more complex model?
- With large \(n\), small \(P\)-value in test of proportional odds assumption may reflect statistical, not practical, significance.
- Effect estimators using simple model are biased but may have smaller MSE, and tests may have greater power, especially when more complex model has many more parameters.
R for modeling *dose-response* data without proportional odds, using *vglm* in VGAM library without parallel=TRUE option:

```r
> trauma <- read.table("trauma.dat", header=TRUE)
> trauma
dose y1 y2 y3 y4 y5
1 1 59 25 46 48 32
2 2 48 21 44 47 30
3 3 44 14 54 64 31
4 4 43 4 49 58 41
> library(VGAM)
> fit2 <- vglm(cbind(y1,y2,y3,y4,y5) ~ dose, family=cumulative, data=trauma)
> summary(fit2)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-0.864585</td>
<td>0.194230</td>
<td>-4.45133</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>-0.093747</td>
<td>0.178494</td>
<td>-0.52521</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>0.706251</td>
<td>0.175576</td>
<td>4.02248</td>
</tr>
<tr>
<td>(Intercept):4</td>
<td>1.908668</td>
<td>0.238380</td>
<td>8.00684</td>
</tr>
<tr>
<td>dose:1</td>
<td>-0.112912</td>
<td>0.072881</td>
<td>-1.54926</td>
</tr>
<tr>
<td>dose:2</td>
<td>-0.268895</td>
<td>0.068319</td>
<td>-3.93585</td>
</tr>
<tr>
<td>dose:3</td>
<td>-0.182341</td>
<td>0.063855</td>
<td>-2.85555</td>
</tr>
<tr>
<td>dose:4</td>
<td>-0.119255</td>
<td>0.084702</td>
<td>-1.40793</td>
</tr>
</tbody>
</table>

Residual Deviance: 3.85163 on 8 degrees of freedom
Log-likelihood: -41.70741 on 8 degrees of freedom

> 1 - pchisq(deviance(fit)-deviance(fit2),
  df=df.residual(fit)-df.residual(fit2))
[1] 0.002487748
```

The improvement in fit is statistically significant, but perhaps not substantively significant; effect of dose is moderately negative for each cumulative probability.
Example: Religious fundamentalism by region

(2006 GSS data)

<table>
<thead>
<tr>
<th>x = Region</th>
<th>Fundamentalist</th>
<th>Moderate</th>
<th>Liberal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>92 (14%)</td>
<td>352 (52%)</td>
<td>234 (34%)</td>
</tr>
<tr>
<td>Midwest</td>
<td>274 (27%)</td>
<td>399 (40%)</td>
<td>326 (33%)</td>
</tr>
<tr>
<td>South</td>
<td>739 (44%)</td>
<td>536 (32%)</td>
<td>412 (24%)</td>
</tr>
<tr>
<td>West/Mountain</td>
<td>192 (20%)</td>
<td>423 (44%)</td>
<td>355 (37%)</td>
</tr>
</tbody>
</table>

We create indicator variables \( \{ r_i \} \) for region and fit the model

\[
\text{logit}[P(y \leq j)] = \alpha_j + \beta_1 r_1 + \beta_2 r_2 + \beta_3 r_3.
\]
R for religion and region data, using \texttt{vglm} for cumulative logit modeling with proportional odds structure:

```r
> religion <- read.table("religion_region.dat", header=TRUE)
> summary(religion)

   region  y1  y2  y3
1       1  92 352 234
2       2 274  399 326
3       3 739  536 412
4       4 192  423 355

> r1 <- ifelse(region==1,1,0); r2 <- ifelse(region==2,1,0); r3 <- ifelse(region==3,1,0)
> cbind(r1,r2,r3)
     r1 r2 r3
[1,]  1  0  0
[2,]  0  1  0
[3,]  0  0  1
[4,]  0  0  0

> library(VGAM)

> fit.po <- vglm(cbind(y1,y2,y3) ~ r1+r2+r3, family=cumulative(parallel=TRUE), data=religion)

> summary(fit.po)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-1.261818</td>
<td>0.064033</td>
<td>-19.70584</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>0.472851</td>
<td>0.061096</td>
<td>7.73948</td>
</tr>
<tr>
<td>r1</td>
<td>-0.069842</td>
<td>0.093035</td>
<td>-0.75071</td>
</tr>
<tr>
<td>r2</td>
<td>0.268777</td>
<td>0.083536</td>
<td>3.21750</td>
</tr>
<tr>
<td>r3</td>
<td>0.889677</td>
<td>0.075704</td>
<td>11.75211</td>
</tr>
</tbody>
</table>

Residual Deviance: 98.0238 on 3 degrees of freedom
Log-likelihood: -77.1583 on 3 degrees of freedom
```
R for religion and region data, using \textit{vglm} for cumulative logit modeling without proportional odds structure:

\begin{verbatim}
> fit.npo <- vglm(cbind(y1,y2,y3) ~ r1+r2+r3, family=cumulative,religion)
> summary(fit.npo)

Coefficients:

\begin{verbatim}
Value Std. Error  z value
(Intercept):1 -1.399231  0.080583 -17.36377
(Intercept):2  0.549504  0.066655  8.24398
r1:1    -0.452300  0.138093  -3.27532
r1:2     0.090999  0.104731   0.86888
r2:1     0.426188  0.107343   3.97032
r2:2     0.175343  0.094849   1.84866
r3:1     1.150175  0.094349  12.19065
r3:2     0.580174  0.087490   6.63135
\end{verbatim}

Residual Deviance: -5.168e-13 on 0 degrees of freedom
Log-likelihood:  -28.1464 on 0 degrees of freedom
\end{verbatim}

\begin{verbatim}
> 1 - pchisq(deviance(fit.po)-deviance(fit.npo),
            df=df.residual(fit.po)-df.residual(fit.npo))
[1] 4.134028e-21
\end{verbatim}

The more complex model is better, in terms of statistical significance.
Model assuming proportional odds has (with $\beta_4 = 0$)
$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-0.07, 0.27, 0.89).$$

For more general model,
$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (-0.45, 0.43, 1.15) \text{ for } \text{logit}[P(Y \leq 1)],$$
$$(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (0.09, 0.18, 0.58) \text{ for } \text{logit}[P(Y \leq 2)].$$

Change in sign of $\hat{\beta}_1$ reflects lack of stochastic ordering of regions 1 and 4; their cdf's don't always have same order.

Compared to resident of West, a Northeast resident is less likely to be fundamentalist (see $\hat{\beta}_1 = -0.45 < 0$ for $\text{logit}[P(Y \leq 1)]$) but slightly more likely to be fundamentalist or moderate and slightly less likely to be liberal (see $\hat{\beta}_1 = 0.09 > 0$ for $\text{logit}[P(Y \leq 2)]$).

Peterson and Harrell (1990) proposed partial proportional odds model falling between proportional odds model and more general model (Sec. 3.6.4 of OrdCDA),

$$\text{logit}[P(y_i \leq j)] = \alpha_j + \beta x_i + \gamma_j u_i, \quad j = 1, \ldots, c - 1.$$
R for modeling *mental impairment* data with partial proportional odds (life events but not SES), using `vglm` in VGAM library:

```r
> fit3 <- vglm(impair ~ ses + life, family=cumulative(parallel=FALSE~ses))
> summary(fit3)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-0.17660</td>
<td>0.69506</td>
<td>-0.25408</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>1.00567</td>
<td>0.66327</td>
<td>1.51623</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>2.39555</td>
<td>0.77894</td>
<td>3.07539</td>
</tr>
<tr>
<td>ses:1</td>
<td>0.98237</td>
<td>0.76430</td>
<td>1.28531</td>
</tr>
<tr>
<td>ses:2</td>
<td>1.54149</td>
<td>0.73732</td>
<td>2.09066</td>
</tr>
<tr>
<td>ses:3</td>
<td>0.73623</td>
<td>0.81213</td>
<td>0.90655</td>
</tr>
<tr>
<td>life</td>
<td>-0.32413</td>
<td>0.12017</td>
<td>-2.69736</td>
</tr>
</tbody>
</table>

Names of linear predictors: `logit(P[Y<=1]), logit(P[Y<=2]), logit(P[Y<=3])`

Residual deviance: 97.36467 on 113 degrees of freedom
Log-likelihood: -48.68234 on 113 degrees of freedom

Deviance (97.36, df=113) similar to obtained with more complex non-proportional odds model (96.75, df=111) and the simpler proportional odds model (99.10, df=115), which seems adequate.
Cumulative Logit Marginal Models for Ordinal Data

Example: Randomized clinical trial for comparing hypnotic drug with placebo in patients with insomnia problems. (Data file is file number 23 at www.stat.ufl.edu/~aa/cda/data.html)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Initial</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;20</td>
<td>20–30</td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>20–30</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>30–60</td>
<td>13</td>
<td>23</td>
</tr>
<tr>
<td>&gt;60</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>20–30</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>30–60</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>&gt;60</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>
$Y_t$ = time to fall asleep,
$x$ = treatment (0 = placebo, 1 = active),
$t$ = occasion (0 = initial, 1 = follow-up after 2 weeks).

Model: \( \logit[P(Y_t \leq j)] = \alpha_j + \beta_1 t + \beta_2 x + \beta_3 (t \times x), \quad j = 1, 2, 3. \)

Generalized estimating equations (GEE) estimates, with independence working equations and robust SE:
\( \hat{\beta}_1 = 1.04 \ (SE = 0.17), \) occasion effect only for placebo,
\( \hat{\beta}_2 = 0.03 \ (SE = 0.24), \) treatment effect only initially, \( \hat{\beta}_3 = 0.71 \ (SE = 0.24), \) interaction (significant).

Considerable evidence that distribution of time to fall asleep decreased more for treatment than placebo group.

Occasion effect = 1.04 for placebo, 1.04 + 0.71 = 1.75 for active.
Occasion odds ratios \( e^{1.04} = 2.8 \) for placebo, \( e^{1.75} = 5.7 \) for active.
Treatment odds ratio \( e^{0.03} = 1.03 \) initially, \( e^{0.03+0.71} = 2.1 \) follow-up.
R for GEE analysis of insomnia data (using repolr):

---

> insomnia<-read.table("insomnia.dat",header=TRUE)
> insomnia<-as.data.frame(insomnia)
> insomnia

<table>
<thead>
<tr>
<th>case</th>
<th>treat</th>
<th>occasion</th>
<th>outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>239</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

> library(repolr)
> fit <- repolr(formula = outcome ~ treat + occasion + treat * occasion, 
+                        + subjects="case", data=insomnia, times=c(1,2), categories=4,
+                        + corstr = "independence")
> summary(fit$gee)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Naive S.E.</th>
<th>Naive z</th>
<th>Robust S.E.</th>
<th>Robust z</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor(cuts)1</td>
<td>-2.26708899</td>
<td>0.2027367</td>
<td>-11.1824294</td>
<td>0.2187606</td>
<td>-10.3633343</td>
</tr>
<tr>
<td>factor(cuts)2</td>
<td>-0.95146176</td>
<td>0.1784822</td>
<td>-5.3308499</td>
<td>0.1809172</td>
<td>-5.2591017</td>
</tr>
<tr>
<td>factor(cuts)3</td>
<td>0.35173977</td>
<td>0.1726860</td>
<td>2.0368745</td>
<td>0.1784232</td>
<td>1.9713794</td>
</tr>
<tr>
<td>treat</td>
<td>0.03361002</td>
<td>0.2368973</td>
<td>0.1418759</td>
<td>0.2384374</td>
<td>0.1409595</td>
</tr>
<tr>
<td>occasion</td>
<td>1.03807641</td>
<td>0.2375992</td>
<td>4.3690229</td>
<td>0.2384374</td>
<td>4.1943093</td>
</tr>
<tr>
<td>treat:occasion</td>
<td>0.70775891</td>
<td>0.3341759</td>
<td>2.1179234</td>
<td>0.2435197</td>
<td>2.9063728</td>
</tr>
</tbody>
</table>

---

Note that because of positive correlation between two responses, naive SE based on independence may be very poor.
**R for GEE analysis of insomnia data (using multgee package):**

```r
insomnia <- read.table("C:/ ... /insomnia.dat", header=TRUE)
insomnia
case treat y1 y2
1 1 1 1 1
2 2 1 1 1
...
238 238 0 4 4
239 239 0 4 4
## data transformation:
id <- rep(insomnia$case,2)
treat <- rep(insomnia$treat,2)
resp <- c(insomnia$y1, insomnia$y2)
time <- c(rep(1, 239), rep(2, 239))
insomnia <- data.frame(id, treat, resp, time)
rep <- factor(time)
tr <- factor(treat)

fitord2 <- ordLORgee(resp~rep+tr+rep:tr, data=insomnia, id=id, repeated=time)
summary(fitord2)
```

**GEE FOR ORDINAL MULTINOMIAL RESPONSES**

**Link**: Cumulative logit

**Local Odds Ratios:**

**Structure**: uniform

**Coefficients:**

|     | Estimate | san.se | san.z | Pr>|san.z| |
|-----|----------|--------|-------|-----|------|
| beta01 | -2.27671  | 0.21952 | -10.3715 | < 2e-16 *** |
| beta02 | -0.95768  | 0.18119 | -5.2855  | < 2e-16 *** |
| beta03 | 0.34525   | 0.17857 | 1.9334   | 0.05319 |
| rep2   | 1.03834   | 0.16915 | 6.1387   | < 2e-16 *** |
| tr1    | 0.03683   | 0.23871 | 0.1543   | 0.87738 |
| rep2:tr1 | 0.71287  | 0.24489 | 2.9110   | 0.00360 ** |
SAS: GEE analysis of insomnia data

data sleep;
  input case treat occasion outcome;
datalines;
  1 1 0 1
  1 1 1 1
  2 1 0 1
  2 1 1 1
...
  239 0 0 4
  239 0 1 4
;
proc genmod data=sleep;
  class case;
  model outcome = treat occasion treat*occasion /
    dist=multinomial link=cumlogit;
  repeated subject=case / type=indep corrw;
run;

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>1Z1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept1</td>
<td>-2.2671</td>
<td>0.2188</td>
<td>-2.6959 -1.8383</td>
<td>-10.36</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Intercept2</td>
<td>-0.9515</td>
<td>0.1809</td>
<td>-1.3061 -0.5969</td>
<td>-5.26</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>Intercept3</td>
<td>0.3517</td>
<td>0.1784</td>
<td>0.0020 0.7014</td>
<td>1.97</td>
<td>0.0487</td>
<td></td>
</tr>
<tr>
<td>treat</td>
<td>0.0336</td>
<td>0.2384</td>
<td>-0.4337 0.5009</td>
<td>0.14</td>
<td>0.8879</td>
<td></td>
</tr>
<tr>
<td>occasion</td>
<td>1.0381</td>
<td>0.1676</td>
<td>0.7096 1.3665</td>
<td>6.19</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>treat*occasion</td>
<td>0.7078</td>
<td>0.2435</td>
<td>0.2305 1.1850</td>
<td>2.91</td>
<td>0.0037</td>
<td></td>
</tr>
</tbody>
</table>
Cumulative Logit Random Effects Models for Ordinal Data

Insomnia example: $Y_t =$ time to fall asleep with treatment $x$ ($0 =$ placebo, $1 =$ active) at occasion $t$ ($0 =$ initial, $1 =$ follow-up). Marginal model

$$\text{logit}[P(Y_t \leq j)] = \alpha_j + \beta_1 t + \beta_2 x + \beta_3 tx, \quad j = 1, 2, 3$$

By contrast, random effects model with common random intercept for each logit is

$$\text{logit}[P(Y_{it} \leq j)] = u_i + \alpha_j + \beta_1 t + \beta_2 x + \beta_3 tx.$$  

<table>
<thead>
<tr>
<th>Effect</th>
<th>GEE</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occasion ($t$)</td>
<td>1.04 (.17)</td>
<td>1.60 (.28)</td>
</tr>
<tr>
<td>Treatment ($x$)</td>
<td>0.03 (.24)</td>
<td>0.06 (.37)</td>
</tr>
<tr>
<td>Treatment×Occasion ($t \times x$)</td>
<td>0.71 (.24)</td>
<td>1.08 (.38)</td>
</tr>
</tbody>
</table>

Results substantively similar, but random-effects model estimates and SE values about 50% larger. Reflects heterogeneity ($\hat{\sigma} = 1.90$ in SAS) and resultant strong $(Y_{i1}, Y_{i2})$ association.
Why are marginal effects weaker? For a particular cumulative probability (e.g., $j = 1$) in a cumulative logit random-intercept model, the conditional subject-specific curves yield a marginal (population-averaged) curve (averaging over these probabilities at each $x$) with attenuated effect.
For R, here we use the *clmm* function in the *ordinal* package, which employs a Laplace approximation for the likelihood function.

```r
> insomnia<- read.table(file="insomnia.txt", header=TRUE)
> attach(insomnia)
> insomnia[1:6,] # part of input data set (each case has been duplicated count times)

<table>
<thead>
<tr>
<th>case</th>
<th>treat</th>
<th>occasion</th>
<th>outcome</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

> outcome<-factor(outcome, levels=1:4, labels=c("20","20-30","30-60","60"), ordered=TRUE)
> treat<-factor(treat, levels=0:1, labels=c("Placebo","Active"))
> occasion<-factor(occasion, levels=0:1, labels=c("Initial","Follow-Up"))
> require("ordinal")
> fit<-clmm(outcome ~ treat*occasion + (1|case)) # random intercept
> summary(fit)

Cumulative Link Mixed Model fitted with the Laplace approximation

Random effects:

|           | Estimate | Std. Error | z value | Pr(>|z|) |
|-----------|----------|------------|---------|---------|
| case      | 3.043    | 1.744      |         |         |

|                |          | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|---------|
| treatActive    | -0.04831 | 0.34390    | -0.140  | 0.88828 |
| occasionFollow-Up | -1.57116 | 0.28404    | -5.531  | 3.18e-08 *** |
| treatActive:occasionFollow-Up | -1.05796 | 0.37812    | -2.798  | 0.00514 ** |

---

Threshold coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20-30</td>
<td>-3.3986</td>
<td>0.3501</td>
</tr>
<tr>
<td>20-30</td>
<td>30-60</td>
<td>-1.4451</td>
<td>0.2780</td>
</tr>
<tr>
<td>30-60</td>
<td>60</td>
<td>0.5457</td>
<td>0.2569</td>
</tr>
</tbody>
</table>
```
SAS (PROC GLIMMIX) for ordinal random effects modeling of insomnia data, fitted using Gauss-Hermite quadrature, which better approximates the likelihood function:

```
data insomnia;
  input case treat occasion outcome;
datalines;
  1 1 0 1
  1 1 1 1
  2 1 0 1
  2 1 1 1
  3 1 0 1
  3 1 1 1
  4 1 0 1
  4 1 1 1
...  239 0 0 4
  239 0 1 4
;
proc glimmix method=quad(qpoints=50) data = insomnia;
  class case;
  model outcome = treat occasion treat*occasion / link=cumlogit dist=multinomial solution;
  random int / sub=case;
run;
```

A. Agresti (UF)
Fit Statistics for Conditional Distribution

-2 log L(outcome l r. effects) 789.00

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>case</td>
<td>3.6280</td>
<td>0.8815</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect         | Outcome | Estimate | Error   | DF   | t Value | Pr > |t|l |
|----------------|---------|----------|---------|------|---------|------|---|
| Intercept      | 1       | -3.4896  | 0.3588  | 237  | -9.73   | <.0001|    |
| Intercept      | 2       | -1.4846  | 0.2903  | 237  | -5.11   | <.0001|    |
| Intercept      | 3       | 0.5613   | 0.2702  | 237  | 2.08    | 0.0388|    |
| treat          |         | 0.05786  | 0.3663  | 235  | 0.16    | 0.8746|    |
| occasion       |         | 1.6016   | 0.2834  | 235  | 5.65    | <.0001|    |
| treat*occasion |         | 1.0813   | 0.3805  | 235  | 2.84    | 0.0049|    |

Estimated standard deviation of random effects is \( \sqrt{3.628} = 1.90 \).
Alternative \textit{(transitional)} model (Sec. 9.3 of \textit{OrdCDA}):\[
\text{logit}[P(y_2 \leq j)] = \alpha_j + \beta_1 x + \beta_2 y_1.
\]

This is an ordinary univariate model that can be fitted easily by ML, e.g. in R with \textit{vglm} function in \textit{VGAM} library, treating $y_2$ as the response variable and $y_1$ as a covariate.

$\hat{\beta}_1 = 0.885$ ($SE = 0.246$) provides strong evidence that follow-up time to fall asleep is lower for the active drug group.

For any given value for the initial response, the estimated odds of falling asleep by a particular time for the active treatment are $\exp(0.885) = 2.4$ times those for the placebo group.

See p. 276 of \textit{ordCDA} for advantages of this approach over marginal models when the initial marginal distributions differ.
Summary, and Alternative Approaches

- Logistic regression for binary responses extends in various ways to handle ordinal responses: We used logits for cumulative probabilities.

- Alternative ordinal logit models exist, such as proportional odds structure with logits for adjacent-response categories, for which effects apply to individual categories.

- Models can also use alternative links, such as cumulative probit models, which correspond to normal latent variable models.

- Models also extend to handle correlated responses, with marginal models (using GEE) and generalized linear mixed models containing random effects.

- Ordinal loglinear models can describe joint distributions of multiple ordinal response variables. See Chapter 6 of *OrdCDA*.

- The Bayesian approach is also possible for these models. See Chapter 11 of *OrdCDA* and also P. Hoff (2009) *First Course in Bayesian Statistical Methods*, Chap. 12.