Short Course
Introduction to Categorical Data Analysis

Alan Agresti
Distinguished Professor Emeritus
University of Florida, USA

Presented for ESALQ/USP, Piracicaba Brazil
March 8-10, 2016

© Alan Agresti, 2016
Course Outline

1. CONTINGENCY TABLE ANALYSIS
2. LOGISTIC REGRESSION
3. MODEL BUILDING
4. LOGLINEAR MODELS
5. LOGIT MODELS FOR MULTICATEGORY RESPONSES
6. MARGINAL AND RANDOM EFFECTS MULTIVARIATE MODELS
Focus of short course

- Overview of most important methods for analyzing categorical response data. Emphasis on concepts, examples of use, rather than theory, derivations, technical details.

- Course is based on *Categorical Data Analysis*, 3rd ed. (published by Wiley, 2013), referred to in notes by *CDA*.

- Examples of model fitting use R and SAS software. For more details and also Stata and SPSS, see website for text, [www.stat.ufl.edu/~aa/cda/cda.html](http://www.stat.ufl.edu/~aa/cda/cda.html)

- For R, useful tutorial by Laura Thompson to accompany 2nd ed. of *CDA* is at [www.stat.ufl.edu/~aa/cda/Thompson_manual.pdf](http://www.stat.ufl.edu/~aa/cda/Thompson_manual.pdf).

See also *Contingency Table Analysis: Methods and Implementation Using R* by Maria Kateri (2014).
CONTINGENCY TABLE ANALYSIS

- Measurement and distributions

- Three measures for comparing proportions

- Odds ratios and their properties

- Testing independence – Chi-squared test

- Loglinear model for two-way tables

- Testing independence – Fisher’s exact test for small samples
Choice of model, analysis, and interpretation depends on

- Response-explanatory variable distinction
- Measurement scales:
  - **binary** - (success, failure)
  - **nominal** - unordered categories
    choice of transport (car, bus, subway, bike, walk)
  - **ordinal** - ordered categories
    patient recovery, quality of life (excellent, good, fair, poor)
Contrasts with continuous scales for regression, ANOVA

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Probability Dist. for Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>←→ Normal</td>
</tr>
<tr>
<td>Categorical</td>
<td>←→ Binomial and multinominal</td>
</tr>
</tbody>
</table>

- **Binomial**: Binary response (0, 1)
- **Multinomial**: Multicategory (nominal or ordinal) response
  Cell counts in a contingency table
### Comparing Proportions (CDA Sec. 2.2)

**ex. Aspirin use and heart attacks**

Source: Harvard physicians health study

<table>
<thead>
<tr>
<th></th>
<th>Heart Attack</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Total</td>
</tr>
<tr>
<td>Placebo</td>
<td>189</td>
<td>10,845</td>
<td>11,034</td>
</tr>
<tr>
<td>Aspirin</td>
<td>104</td>
<td>10,933</td>
<td>11,037</td>
</tr>
</tbody>
</table>

The data are counts in a $2 \times 2$ contingency table

Sample proportions of heart attack

Placebo: $\frac{189}{11,034} = 0.0171$

Aspirin: $\frac{104}{11,037} = 0.0094$
Treat data as two independent binomial samples

Three descriptive measures:

1. **Difference of proportions**
   
   $$0.0171 - 0.0094 = 0.0077$$

2. **Relative risk**

   $$\frac{0.0171}{0.0094} = 1.82$$

3. **Odds ratio**

   $$\hat{\theta} = \frac{0.0171/0.9829}{0.0094/0.9906} = \frac{189 \times 10,933}{104 \times 10,845} = 1.83$$

   Also called *cross-product ratio* (Yule 1900, 1912)
Odds of attack were \((\text{number ‘yes’})/(\text{number ‘no’})\)

\[
\frac{189}{10,845} = 0.0174 \quad \text{for placebo}
\]
(i.e., 174 yes for every 10,000 no)

\[
\frac{104}{10,933} = 0.0095 \quad \text{for aspirin}
\]
(i.e., 95 yes for every 10,000 no)

Odds of attack for placebo group were \(174/95 = 1.83\) times odds of attack for aspirin group (i.e., 83\% higher for placebo group)

For conditional probabilities \(\{\pi_j|_i\}\), joint cell probabilities \(\{\pi_{ij}\}\), counts \(\{n_{ij}\}\)

Odds ratio \(\theta = \frac{\pi_{11}/\pi_{21}}{\pi_{12}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\)

Sample estimate \(\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}\)
Properties of odds ratio

- $0 \leq \theta \leq \infty$
- $\theta = 1 \leftrightarrow$ no effect
  
  $(\log \theta = 0)$  (“independence”)  

- Interchange rows (columns)

  $\theta \to 1/\theta$  $(\log \theta \to \log \theta)$

- $\theta = 4$  $(\log \theta = 1.39)$ same strength as

- $\theta = \frac{1}{4}$  $(\log \theta = -1.39)$. 
Denote

\( X = \) row variable

\( Y = \) column variable

\[
\theta = \frac{P(Y = 1|X = 1)/P(Y = 2|X = 1)}{P(Y = 1|X = 2)/P(Y = 2|X = 2)} = \frac{P(X = 1|Y = 1)/P(X = 2|Y = 1)}{P(X = 1|Y = 2)/P(X = 2|Y = 2)}
\]

\( \Rightarrow \) \( \theta \) applicable for \textit{prospective} (e.g., cohort) studies, or \textit{retrospective} (e.g., case-control) studies that observe past behavior on \( X \) for matched subjects at the levels of \( Y \).

Relative risk = \[
\frac{P(Y=1|X=1)}{P(Y=1|X=2)}
\]

not applicable for retrospective studies, but it may be approximated by odds ratio if \( P(Y = 1|X = x) \approx 0 \) for \( x = 1, 2 \).
Inference for odds ratio  \((CDA \text{ Sec. } 3.1)\)

- \(\hat{\theta} \to\) normal sampling distribution for large \(n\), but log \(\hat{\theta} \to\) normal much more quickly
  (conduct inferences such as CI for log \(\theta\), take antilogs for inferences about \(\theta\))

- Approximate large-sample standard error of log \(\hat{\theta}\) is
  \[
  SE = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}
  \]

**Example:** Aspirin and heart attacks

95% confidence interval for log \(\theta\) is

\[
\log 1.83 \pm 1.96 \sqrt{\frac{1}{189} + \frac{1}{10,845} + \frac{1}{104} + \frac{1}{10,983}}
\]

\[= 0.60 \pm 0.24, \text{ or } (0.36, 0.84)\]

\[\to (e^{0.36}, e^{0.84}) = (1.44, 2.33) \text{ for } \theta\]

Conclude \(\theta > 1\) (i.e., odds of heart attack higher for placebo)
R for confidence interval for odds ratio, based on logistic model:

---

```r
> placebo <- c(1,0)
> attack <- c(189, 104)
> n <- c(189 + 10845, 104 + 10933)

> fit <- glm(attack/n ~ placebo, weights=n, family=binomial)
> summary(fit)

Coefficients:  

              Estimate Std. Error z value Pr(>|z|)  
(Intercept)   -4.6552    0.09852 -47.249  < 2e-16 ***  
placebo       0.6054     0.12284  4.929    8.28e-07 ***  
---  
Null deviance: 2.5372e+01 on 1 degrees of freedom  
Residual deviance: -2.0983e-13 on 0 degrees of freedom

> exp(0.60544 - 1.96*(0.12284)); exp(0.60544 + 1.96*(0.12284))
[1] 1.440044
[1] 2.330788

> exp(confint(fit))  # to get profile likelihood CI for odds ratio

                     2.5 %     97.5 %  
(Intercept) 0.007792767 0.01147071
placebo  1.443579310 2.33786091

---
```
SAS for confidence intervals for odds ratio, relative risk:

```
data aspirin;
input group $ attack $ count @@;
datalines;
placebo yes 189 placebo no 10845 aspirin yes 104 aspirin no 10933
;
proc freq order=data; weight count;
   tables group*attack / measures ;
run;
```

Estimates of the Relative Risk (Row1/Row2)

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Value</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control (Odds Ratio)</td>
<td>1.8321</td>
<td>1.4400 2.3308</td>
</tr>
<tr>
<td>Cohort (Col1 Risk)</td>
<td>1.8178</td>
<td>1.4330 2.3059</td>
</tr>
<tr>
<td>Cohort (Col2 Risk)</td>
<td>0.9922</td>
<td>0.9892 0.9953</td>
</tr>
</tbody>
</table>
Context: $I \times J$ contingency table

Expected frequencies $\{\mu_{ij} = n\pi_{ij}\}$

ML estimates $\{\hat{\mu}_{ij} = n\hat{\pi}_{ij}\}$ are called *estimated expected frequencies*, or *fitted values*

Two responses $X$ and $Y$ are *independent* if

$$\pi_{ij} = \pi_i + \pi_j \quad \text{all } i \text{ and } j$$

i.e., $P(X = i, Y = j) = P(X = i)P(Y = j) \quad \text{all cells}$
Let \( p_{i+} = n_{i+}/n \), \( p_{+j} = n_{+j}/n \) (sample proportions)

Under \( H_0 \) : independence,

\[
\hat{\pi}_{ij} = p_{i+} + p_{+j} \\
\hat{\mu}_{ij} = n\hat{\pi}_{ij} = np_{i+}p_{+j} = \frac{n_{i+}n_{+j}}{n}
\]

Compare data \( \{n_{ij}\} \) to fit \( \{\hat{\mu}_{ij}\} \) using

\[
X^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \quad \text{(Pearson)}
\]

or

\[
G^2 = 2 \sum_i \sum_j n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right) \quad \text{(likelihood-ratio)}
\]

\( X^2 \) and \( G^2 \) measure discrepancy between data and the fit.
Properties

- Larger values provide stronger evidence against $H_0$: independence
- $X^2, G^2 \rightarrow \chi^2$ (chi-squared) as $\{\mu_{ij}\} \rightarrow \infty$
- $df = \text{difference in dimensions of parameter space under } H_0 U H_a \text{ and under } H_0$
  
  $$df = (IJ - 1) - [(I - 1) + (J - 1)] = (I - 1)(J - 1)$$

- To test $H_0$: model holds,
  
  $$P\text{-value} = P_{H_0}[X^2 \geq X^2_{obs}]$$
  or $$P_{H_0}[G^2 \geq G^2_{obs}]$$

- $X^2 - G^2 \xrightarrow{p} 0 \text{ as } n \rightarrow \infty, \text{ under } H_0$
- $X^2$ is a quadratic approximation for $G^2$ (CDA, p. 597)
- $X^2$ preferable for sparse data
  (sampling distribution close to chi-squared if most $\hat{\mu}_{ij} \geq 5$)
**Example:** Attained Education (Highest Degree) and Belief in God (*CDA*, p. 77)

<table>
<thead>
<tr>
<th>Highest Degree</th>
<th>Don't Believe</th>
<th>No Way to Find Out</th>
<th>Some Higher Power</th>
<th>Believe Sometimes</th>
<th>Believe but Doubts</th>
<th>Know God Exists</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than high school</td>
<td>9</td>
<td>8</td>
<td>27</td>
<td>8</td>
<td>47</td>
<td>236</td>
<td>335</td>
</tr>
<tr>
<td>High school</td>
<td>(10.0)$^1$</td>
<td>(15.9)</td>
<td>(34.2)</td>
<td>(12.7)</td>
<td>(55.3)</td>
<td>(206.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.4)$^2$</td>
<td>(-2.2)</td>
<td>(-1.4)</td>
<td>(-1.5)</td>
<td>(-1.3)</td>
<td>(3.6)</td>
<td></td>
</tr>
<tr>
<td>High school or junior college</td>
<td>23</td>
<td>39</td>
<td>88</td>
<td>49</td>
<td>179</td>
<td>706</td>
<td>1084</td>
</tr>
<tr>
<td></td>
<td>(32.5)</td>
<td>(51.5)</td>
<td>(110.6)</td>
<td>(41.2)</td>
<td>(178.9)</td>
<td>(669.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.5)</td>
<td>(-2.6)</td>
<td>(-3.3)</td>
<td>(1.8)</td>
<td>(0.0)</td>
<td>(3.4)</td>
<td></td>
</tr>
<tr>
<td>Bachelor or graduate</td>
<td>28</td>
<td>48</td>
<td>89</td>
<td>19</td>
<td>104</td>
<td>293</td>
<td>581</td>
</tr>
<tr>
<td></td>
<td>(17.4)</td>
<td>(27.6)</td>
<td>(59.3)</td>
<td>(22.1)</td>
<td>(95.9)</td>
<td>(358.8)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.1)</td>
<td>(4.7)</td>
<td>(4.8)</td>
<td>(-0.8)</td>
<td>(1.1)</td>
<td>(-6.7)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td>95</td>
<td>204</td>
<td>76</td>
<td>330</td>
<td>1235</td>
<td>2000</td>
</tr>
</tbody>
</table>

*Source: General Social Survey, National Opinion Research Center.*

$^1$Estimated expected frequency for testing independence

$^2$Standardized residual, $z = (n_{ij} - \hat{\mu}_{ij})/SE$. 

*e.g., $\hat{\mu}_{11} = \frac{335 \times 60}{2000} = 10.0$*

\[ X^2 = 76.15, \quad G^2 = 73.2, \quad df = (3 - 1)(6 - 1) = 10 \]

\[ P\text{-value} = P_{H_0}[X^2 \geq 76.15] = 0.000\ldots \]
R for data on education and belief in God

> data <- matrix(c(9,8,27,8,47,236,23,39,88,49,179,706,28,48,89,19,104,293),
   ncol=6,byrow=TRUE)
> chisq.test(data)

  Pearson's Chi-squared test
  X-squared = 76.1483, df = 10, p-value = 2.843e-12

> chisq.test(data)$stdres  # standardized residuals
[1,]  -0.368577 -2.227511 -1.418621 -1.481383 -1.3349600  3.590075
[2,]  -2.504627 -2.635335 -3.346628  1.832792  0.0169276  3.382637
[3,]   3.051857  4.724326  4.839597 -0.792912  1.0794638 -6.665195

To calculate $G^2$ and other summaries for two-way tables, see p. 10 of www.stat.ufl.edu/~presnell/Courses/sta4504-2000sp/R/R-CDA.pdf
SAS for testing independence between education and belief in God

```sas
data table;
  input degree belief $ count @@;
datalines;
1 1 9 1 2 8 1 3 27 1 4 8 1 5 47 1 6 236
2 1 23 2 2 39 2 3 88 2 4 49 2 5 179 2 6 706
3 1 28 3 2 48 3 3 89 3 4 19 3 5 104 3 6 293
;
proc freq order=data; weight count;
  tables degree*belief / chisq expected measures cmh1;
```

The FREQ Procedure

Statistics for Table of degree by belief

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>10</td>
<td>76.1483</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>10</td>
<td>73.1879</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>52.4861</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

“Chi-Square” is Pearson statistic.
“Mantel-Haenszel chi-square” is df=1 test treating rows and columns as ordinal, based on correlation for fixed row and column scores such as row number and column number (CDA, p. 87).
Independence can be expressed as a **loglinear model**

\[ \pi_{ij} = \pi_i + \pi_j \]

\[ \mu_{ij} = n\pi_i + \pi_j \]

\[ \log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y \]

for effect of being in row \( i \) of variable \( X \) and in column \( j \) of variable \( Y \).

Analogous to main effects model in 2-way ANOVA for two factor explanatory variables. We’ll present such models in Sec. 4 of this course.
For loglinear models,

- No distinction between response and explanatory classifications (express \( \{ \log \mu_{ij} \} \) in terms of levels of all variables)
- Dependence requires association parameters \( \{ \lambda_{ij}^{XY} \} \)

\[
\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}
\]

- Loglinear association parameters relate to odds ratios

2 x 2 case

\[
\log(\theta) = \log \left[ \begin{array}{c}
\mu_{11}
\mu_{22}
\mu_{12}
\mu_{21}
\end{array} \right]
\]

\[
= \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21}
\]

\[
= [\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}] + [\lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}]
\]

\[
- [\lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}] - [\lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}]
\]

\[
= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}
\]

(With identifiability constraints imposed, there’s actually only one extra parameter here.)
Fisher’s Exact Test \((CDA, \text{Sec. 3.5})\)

For small \(n\), can conduct exact test of independence \((H_0 : \pi_{ij} = \pi_i + \pi_j \text{ all } i, j)\) using the hypergeometric dist. This results from conditioning on margin totals, the “sufficient statistics” for unknown marginal probabilities.

**ex.** \(2 \times 2\)

<table>
<thead>
<tr>
<th></th>
<th>(n_{11})</th>
<th>(n_{1+})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n_{+1})</td>
<td>(n_{+1})</td>
<td>(n_{+2})</td>
</tr>
<tr>
<td>(n)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
f_{H_0}(n_{11}) = \frac{\binom{n_{1+}}{n_{11}} \binom{n_{2+}}{n_{+1} - n_{11}}}{\binom{n}{n_{+1}}}
\]

e.g, for \(H_0 : \theta = 1, \ H_a : \theta > 1\) (odds ratio)

P-value = \(P_{H_0}[n_{11} \geq \text{observed } n_{11}]\)
Example: Fisher’s tea taster

<table>
<thead>
<tr>
<th>Poured First</th>
<th>Milk</th>
<th>Tea</th>
<th>( \text{Guess Poured First} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Tea</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

\[
P = \frac{\binom{4}{3} \binom{4}{1} + \binom{4}{4} \binom{4}{0}}{\binom{8}{4}} = 0.243
\]

The attainable \( P \)-values for these margins are 0.014, 0.243, 0.757, 0.986, 1.00, for \( n_{11} = 4, 3, 2, 1, 0 \).
R for Fisher’s exact test with tea tasting data:

```
> tea <- matrix(c(3,1,1,3), ncol=2, byrow=TRUE)

> fisher.test(tea, alternative="greater")

Fisher’s Exact Test for Count Data

data: tea
p-value = 0.2429
alternative hypothesis: true odds ratio is greater than 1

> fisher.test(tea)

Fisher’s Exact Test for Count Data

data: tea
p-value = 0.4857
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.2117329 621.9337505
sample estimates:
odds ratio  # This is the "conditional ML" estimate (CDA, p. 267-268)
 6.408309
```

---

A. Agresti (UF)  CDA  March 8-10, 2016  25 / 191
SAS for Fisher’s exact test:

```sas
data fisher;
  input poured guess count;
cards;
  1 1 3
  1 2 1
  2 1 1
  2 2 3
;
  proc freq;
    weight count;
  tables poured*guess / measures riskdiff;
  exact fisher or chisq trend / alpha=.05;
run;
```

Fisher’s Exact Test

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell (1,1) Frequency (F)</td>
<td>3</td>
</tr>
<tr>
<td>Left-sided Pr &lt;= F</td>
<td>0.9857</td>
</tr>
<tr>
<td>Right-sided Pr &gt;= F</td>
<td>0.2429</td>
</tr>
<tr>
<td>Table Probability (P)</td>
<td>0.2286</td>
</tr>
<tr>
<td>Two-sided Pr &lt;= P</td>
<td>0.4857</td>
</tr>
</tbody>
</table>
Notes

- *Yates continuity correction* is adjustment to Pearson $X^2$ yielding P-value (from $\chi^2$ distribution) close to exact $P$-value from Fisher test. (No reason ever to use it, given modern computation.)

- Fixing both margins makes sampling distribution highly discrete.

Most commonly, only $(n_{1+}, n_{2+})$ are naturally fixed (i.e., two independent binomials). Then, for testing

$$P(Y = 1|X = 1) = P(Y = 1|X = 2)$$

Fisher’s exact test is *conservative*; i.e., when $H_0$ true, may reject $H_0$ at 0.05 level much less than 5% of time.
Alternatives to reduce conservativeness in $2 \times 2$ case include

- “exact” tests that fix only row marginal totals \((CDA, \text{Sec. 3.5.6})\)
  These assume independent binomial sampling; see, e.g., Berger and Boos (1994)

- mid-P value \((CDA, \text{Sec. 1.4.4})\)
  \[ P = \frac{1}{2} P(n_{11} = n_{11} \text{ obs.}) + P(n_{11} > n_{11} \text{ obs.}) \]
  \[ \rightarrow \text{Satisfies } E_{H_0}(P) = 0.50 \]

- Fisher exact test generalizes to $I \times J$, $I \times J \times K$ tables, and to “exact” CI for odds ratios \((CDA, \text{Sec. 16.5, 16.6})\)
Categorical data: binary, nominal, ordinal

Binomial and multinomial sampling models for categorical data

Describing association:
- difference of proportions, relative risk, odds ratio

Inference for odds ratio

Independence for $I \times J$ table
- $G^2$ and $X^2$ chi-squared tests (large $n$)
- Fisher’s exact test (small $n$)

Independence is a loglinear model

Odds ratio is a natural parameter for a loglinear model (we’ll see this is also true for logistic regression)
Logistic regression as a generalized linear model

- Binary \( Y \), continuous \( x \)
- Binary \( Y \), categorical \( x \), multiple predictors
- Likelihood-based inference for model parameters
- Goodness of fit

Notation:

\( Y_i = \) response outcome for subject \( i \)

\( x_{ij} = \) value of explanatory variable \( j \) for subject \( i \)
Generalized Linear Models \((CDA, \text{Sec. 4.1})\)

1. Random component

   Choose response var. \(Y\). Let \(E(Y_i) = \mu_i\) for subject \(i\).

2. Systematic component

   Choose explanatory variables

   \[\text{Linear predictor } \sum_j \beta_j x_{ij}\]

3. Link function

   Monotone function \(g\) relating linear predictor to \(\mu_i\).

   \[g(\mu_i) = \sum_j \beta_j x_{ij}\]
Deviance

For GLMs, analysis of variance generalizes to analysis of likelihood functions, called analysis of \textit{deviance}.

Let maximized log likelihood $= L_M$ for the model $L_S = \text{maximized log likelihood for "saturated model" (perfect fit: parameter for each observation)}$.

Deviance $= -2[L_M - L_S]$, which equals $G^2$ for models for categorical data.

References on generalized linear models

Logistic regression as a GLM \((CDA, \text{Sec. 4.2})\)

- **Random component:** \(Y_i \sim \text{Bin}(n_i; \pi_i)\)

\[
p(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} = \binom{n}{y} (1 - \pi)^n \left(\frac{\pi}{1 - \pi}\right)^y
\]

\[
= \binom{n}{y} (1 - \pi)^n \exp \left[y \log \left(\frac{\pi}{1 - \pi}\right)\right]
\]

in the exponential family representation

- **Link:** \(\log\left(\frac{\pi_i}{1 - \pi_i}\right)\)

**Logit** link = natural parameter in expo. family representation (called the *canonical* link function)

Other popular link functions include

*Probit* (underlying normal latent variable)

**Complementary log-log:** \(\log[-\log(1 - \pi_i)] = \sum_j \beta_j x_{ij}\)

**Identity:** \(\pi_i = \sum_j \beta_j x_{ij}\)
Consider binary response $Y_i = 0$ or $1$.

Suppose subject $i$ has tolerance $T_i$, with

$$Y_i = \begin{cases} 1 & \text{if } x \geq T_i \\ 0 & \text{if } x < T_i \end{cases}$$

For population, suppose $P(T \leq t) = G(t)$.

Then,

$$P(Y = 1|x) = \pi(x) = P(T \leq x) = G(x) = F(\alpha + \beta x) \text{ for some “standard” cdf } F.$$ 

Suggests model of form

$$F^{-1}[\pi(x)] = \alpha + \beta x \text{ for some cdf } F.$$
Example: \( F = \text{standard normal cdf} \ \Phi \)

\( \pi(x) = \Phi(\alpha + \beta x) \) is called the **probit** model
(response curve looks like normal cdf)

The link \( \Phi^{-1} \) is the **probit link**

Example: \( F(x) = \frac{e^x}{1+e^x} = \text{standard logistic} \)

\[
\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}
\]

is the **logistic regression** model, for which

\[
1 - \pi(x) = \frac{1}{1 + e^{\alpha + \beta x}}
\]
For logistic regression with single explanatory variable,

\[ \text{odds} = \frac{\pi(x)}{1 - \pi(x)} = e^{\alpha + \beta x} \]

For two levels of \( x \), denoted \( x_1 \) and \( x_2 \),

\[ \text{odds ratio} = \frac{\pi(x_1)/[1 - \pi(x_1)]}{\pi(x_2)/[1 - \pi(x_2)]} = \frac{e^{\alpha + \beta x_1}}{e^{\alpha + \beta x_2}} = e^{\beta (x_1 - x_2)} \]

For \( x_1 - x_2 = 1 \), the odds of a success at \( x = x_1 \) are \( e^\beta \) times odds of success at \( x = x_2 \);
i.e., odds multiply by \( e^\beta \) for every 1-unit increase in \( x \)

\( \beta = 0 \leftrightarrow \text{odds ratio} = 1 \leftrightarrow \text{no effect} \)
Linearize form of model with \textit{logit} transform

$$\log \left[ \frac{\pi(x)}{1-\pi(x)} \right] = \alpha + \beta x$$

This is a GLM with binomial random component and logit link.

Generalizes to multiple logistic regression

$$\log \left[ \frac{\pi(x)}{1-\pi(x)} \right] = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p$$

Then, \(\exp(\beta_j)\) represents odds ratio between \(Y\) and two levels of \(x_j\) that are 1-unit apart, adjusting for other predictors in model.
Odds \(= \frac{\pi(x)}{1-\pi(x)} = e^{\alpha}(e^{\beta})^x\); i.e., multiplicative effect

Monotone

\[
\beta > 0: \quad \pi(x) \uparrow 1 \text{ as } x \to \infty \\
\beta < 0: \quad \pi(x) \downarrow 0 \text{ as } x \to \infty \\
\beta = 0: \quad \pi(x) \text{ constant}
\]

\[
\frac{\partial \pi(x)}{\partial x} = \beta \pi(x)(1 - \pi(x)),
\]

so slope is proportional to \(\beta\) and steepest \((\beta/4)\) at \(x\)-value where \(\pi(x) = 0.50\); this \(x\) value is \(x = -\alpha/\beta\).

Valid with retrospective studies, essentially because odds ratio for \(Y\) given \(x\) is the same as for \(X\) given \(y\). (CDA, Sec. 5.1.4)
**Example:** Predictors for cancer remission  
\(Y = \text{cancer remission (1 = yes, 0 = no)}\)  
\(x = \text{labeling index (LI)}\)

<table>
<thead>
<tr>
<th>LI</th>
<th>No. cases</th>
<th>No. remissions</th>
<th>← i.e, (Y = 0) for each of 2 observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

27  

With software, we can enter data as 27 Bernoulli outcomes or 14 binomials; same likelihood function and ML estimates either way.

\[
\text{Deviance} = -2[L_M - L_S] \text{ differs, because saturated model has more parameters for Bernoulli data file.}
\]
R for cancer remission data

> LI <- c(8,8,10,10,12,12,12,12,16,16,16,18,20,20,20,20,22,22,24,26, + 28,32,34,38,38,38)
> y <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,0,1,0,0,1,1,0,1,1,1,0)
> logit.fit <- glm(y ~ LI, family=binomial(link=logit))
> summary(logit.fit)

Coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | -3.77714 | 1.37862 | -2.740 | 0.00615 ** |
| LI | 0.14486 | 0.05934 | 2.441 | 0.01464 * |

---

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

Null deviance: 34.372 on 26 degrees of freedom
Residual deviance: 26.073 on 25 degrees of freedom
AIC: 30.073
Number of Fisher Scoring iterations: 4

> confint(logit.fit) # to get profile likelihood CI

2.5 % 97.5 %
(Intercept) -6.9951909 -1.4098443
LI 0.0425232 0.2846668

> exp(-3.77714+0.14486*20.1)/(1+exp(-3.77714+0.14486*20.1))
[1] 0.2962011
data remiss;
  input li success n @@;
cards;
  8 0 2 10 0 2 12 0 3 14 0 3 16 0 3
  18 1 1 20 2 3 22 1 2 24 0 1 26 1 1
  28 1 1 32 0 1 34 1 1 38 2 3
;
proc genmod; model success/n = li / dist=bin link=logit lrci type3;
proc logistic ; model success/n = li;
run;

PROC GENMOD

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.7771</td>
<td>1.3786</td>
<td>-6.9946 -1.4097</td>
<td>7.51</td>
<td>0.0061</td>
</tr>
<tr>
<td>li</td>
<td>0.1449</td>
<td>0.0593</td>
<td>0.0425 0.2846</td>
<td>5.96</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

LR Statistics For Type 3 Analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>li</td>
<td>1</td>
<td>8.30</td>
<td>0.0040</td>
</tr>
</tbody>
</table>

PROC LOGISTIC

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>8.2988</td>
<td>1</td>
<td>0.0040</td>
</tr>
<tr>
<td>Score</td>
<td>7.9311</td>
<td>1</td>
<td>0.0049</td>
</tr>
<tr>
<td>Wald</td>
<td>5.9594</td>
<td>1</td>
<td>0.0146</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<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</td>
<td>1</td>
<td>-3.7771</td>
<td>1.3786</td>
<td>7.5064</td>
<td>0.0061</td>
</tr>
<tr>
<td>li</td>
<td>1</td>
<td>0.1449</td>
<td>0.0593</td>
<td>5.9594</td>
<td>0.0146</td>
</tr>
</tbody>
</table>
ML fit of logistic regression model for \( \pi = \text{prob.}(\text{remission}) \) is

\[
\log \left[ \frac{\hat{\pi}}{1 - \hat{\pi}} \right] = -3.777 + 0.145x
\]

Prediction equation

\[ \hat{\pi} = \frac{\exp(\hat{\alpha} + \hat{\beta}x)}{1 + \exp(\hat{\alpha} + \hat{\beta}x)} = \frac{\exp(-3.777 + 0.145x)}{1 + \exp(-3.777 + 0.145x)} \]

e.g. at \( x = \bar{x} = 20.1 \),

\[ \hat{\pi} = \frac{\exp[-3.777 + 0.14486(20.1)]}{1 + \exp[-3.777 + 0.14486(20.1)]} = 0.296 \]

The incremental rate of change at \( x = 20.1 \) is

\[ \hat{\beta}\hat{\pi} (1 - \hat{\pi}) = 0.14486(0.296)(0.704) = 0.030 \]

\[ \hat{\pi} = 0.50 \leftrightarrow \log \left( \frac{\hat{\pi}}{1 - \hat{\pi}} \right) = 0 = \hat{\alpha} + \hat{\beta}x \]

\[ \leftrightarrow x = -\hat{\alpha}/\hat{\beta} = 26.0 \]
\( \hat{\beta} = 0.145 \) means that for each unit change in LI, estimated odds of remission are multiplied by \( \exp(0.145) = 1.16 \).

i.e., 16% increase when LI ↑ 1.

\[ \text{e.g., at } x = 26, \hat{\pi} = 0.498 \text{ (odds } = 0.990) \]
\[ \text{at } x = 27, \hat{\pi} = 0.534 \text{ (odds } = 1.145) \]
\[ = 0.990 \times 1.16 \]

i.e., odds ratio = \( \frac{0.534/(1-0.534)}{0.498/(1-0.498)} = 1.16 \)

Simpler effect measures:

Change in \( \hat{\pi} \) from minimum to maximum value of \( x \)
(Here, as LI goes from 8 to 38, \( \hat{\pi} \) increases from 0.07 to 0.85)

Change in \( \hat{\pi} \) between lower and upper quartiles of \( x \), which is more resistant to outliers
(Here, \( LQ = 13, UQ = 25, \hat{\pi} \) goes from 0.13 to 0.46)

With multiple predictors, can find these at means of other predictors.
Test of no effect \((H_0 : \beta = 0)\): \(\hat{\beta} = 0.145, \ SE = 0.059\)

- \(z = \hat{\beta}/SE = 2.45\)
  \((z^2 = 5.96 \sim \chi^2, \text{under } H_0, \text{called Wald statistic})\)
  Strong evidence of a positive association \((P = 0.015)\).

- **Likelihood-ratio** test stat. for \(H_0: \beta = 0\), equals 2 times decrease in maximized log-likelihood (= increase in deviance) when impose \(H_0\) constraint. Here, equals 8.30, \(df = 1\) \((P = 0.004)\).
  Likelihood-ratio test is often more powerful than Wald test, especially when \(|\beta|\) large, so prob. near 0 or near 1 \((CDA, \text{Sec. 5.2.6})\).
  In *lmtest* package in R, can get LR test comparing simpler model \(M_0\) to more complex model \(M_1\) using \(\text{lrtest}(M_1,M_0)\).

- **Score** test based on derivative of log likelihood at \(\beta = 0\).

- The three types of tests are asymptotically equivalent, when \(H_0\) true.
Similarly, there are three possible types of confidence intervals, based on inverting these three tests; e.g., 95% confidence interval is the set of $\beta_0$ not rejected at the 0.05 level in testing $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$.

- Best known is “Wald interval,” $\hat{\beta} \pm 1.96(SE)$.
- The likelihood-ratio test-based interval (profile likelihood) is the set of $\beta_0$ for which $-2 \log$ likelihood-ratio decreases by less than $(1.96)^2$, which is the 0.95 percentile of chi-squared with $df = 1$. For small to moderate $n$ or parameter taking value near boundary, this is preferable to Wald CI.

SAS: in PROC GENMOD, LRCI option
R: `confint` function

**ML Fitting:**

Likelihood equations are nonlinear.

$\hat{\beta}$ obtained iteratively using Newton-Raphson method. This provides a sequence of parabolic approximations to log likelihood $L$, which is concave.
With observed \((x_i, y_i)\) for subject \(i\), likelihood function is

\[
\ell(\alpha, \beta) = \prod_{i=1}^{n} \left[ \left( \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} \right)^{y_i} \left( \frac{1}{1 + e^{\alpha + \beta x_i}} \right)^{1-y_i} \right].
\]

Extension: For general model \(\logit(\pi) = X\beta\) with \(\beta = (\beta_1, \beta_2, \ldots, \beta_p)^T\),

\[
\sqrt{n}[\hat{\beta} - \beta] \rightarrow N(0, J^{-1})
\]

where information matrix \(J = -\left( \frac{\partial^2 L}{\partial \beta_a \partial \beta_b} \right)\) with \(L = \log \ell(\beta)\).

Estimate by \(\hat{J} = X^T \text{Diag}[\hat{\pi}_i(1 - \hat{\pi}_i)]X\)

Software reports standard errors from square roots of diagonal elements of estimated inverse information matrix.
As in ANOVA, qualitative factors can be explanatory variables in logistic regression models, using indicator (dummy) variables. (*CDA*, Sec. 5.3)

**Example:** Model for multi-center clinical trial

\[ Y = \text{response (success, failure)} \]
\[ X = \text{group (e.g., treatment)} \]
\[ Z = \text{center, clinic, or control var. such as age} \]

Main effects model is
\[
\log \left[ \frac{P(Y = 1 | X = i, Z = k)}{P(Y = 2 | X = i, Z = k)} \right] = \alpha + \beta_i^X + \beta_k^Z
\]

For each center, the odds of success for group 1 are
\[
\exp(\beta_1^X - \beta_2^X) \text{ times those for group 2.}
\]

For identifiability, software uses constraints (such as \( \beta_1^X = \beta_1^Z = 0 \) in R), when declare variable to be a factor.

For two groups, equivalent to set up indicator \( x = 0 \) for group 1, \( x = 1 \) for group 2, and use \( \beta x \) in linear predictor, where \( \beta = \beta_2^X \).
Example: AIDS symptoms and AZT use, by race  
(CDA, Sec. 5.4.2)

<table>
<thead>
<tr>
<th>Race</th>
<th>AZT Use</th>
<th>AIDS Symptoms</th>
<th>Proportion Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Yes</td>
<td>14</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>32</td>
<td>81</td>
</tr>
<tr>
<td>Black</td>
<td>Yes</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>12</td>
<td>43</td>
</tr>
</tbody>
</table>

\[ \pi_{ij} = \text{prob(AIDS symptoms = yes)} \] for AZT use \( i \) and race \( j \)

Logit model with categorical explanatory variables

\[
\log \left[ \frac{\pi_{ij}}{1 - \pi_{ij}} \right] = \alpha + \beta_i^A + \beta_j^R
\]

for effect \( \beta_i^A \) of category \( i \) of \( A = \) AZT use and effect \( \beta_j^R \) of category \( j \) of \( R = \) race.
Constraints $\beta_2^A = \beta_2^R = 0$ (e.g., as in SAS) correspond to expressing the model using indicator variables as

\[
\text{logit}[P(\text{AIDS symptoms} = \text{yes})] = \alpha + \beta_1 A + \beta_2 R
\]

where

\[
A = \begin{cases} 
1 & \text{"yes" AZT use} \\
0 & \text{"no" AZT use} 
\end{cases}
\]

\[
R = \begin{cases} 
1 & \text{white subject} \\
0 & \text{black subject} 
\end{cases}
\]

As in ordinary linear models, with $k$ categories for a predictor, need $k - 1$ indicator variables and $df = k - 1$ for testing its effect.

If you don’t set up indicators yourself,

SAS: PROC GENMOD and LOGISTIC have CLASS option

R: Can declare an explanatory variable to be a factor
R for AIDS and AZT use example

---------------
> race <- c(1,1,0,0); azt <- c(1,0,1,0)
> symptoms <- c(14,32,11,12); n <- c(107, 113, 63, 55)

> fit <- glm(symptoms/n ~ race + azt, weights=n, family=binomial(link=logit))

> summary(fit)
Call: glm(formula = symptoms/n ~ race + azt, data = n, family = binomial(link = logit))

Coefficients:  
                Estimate Std. Error z value Pr(>|z|)  
(Intercept)       -1.0736    0.2629  -4.083  4.45e-05 ***
race              0.0555    0.2886   0.192   0.84755 
azt               -0.7195    0.2789  -2.579  0.00991 **
---
Null deviance: 8.3499 on 3 degrees of freedom  
Residual deviance: 1.3835 on 1 degrees of freedom  
AIC: 24.86

> anova(fit)
Analysis of Deviance Table

Terms added sequentially (first to last)

                 Df Deviance Resid. Df Resid. Dev
NULL             3     8.3499
race             1   0.0955         2     8.2544
azt              1  6.8709          1   1.3835
---------------
> fit2 <- glm(symptoms/n ~ race, weights=n, family=binomial(link=logit))

> summary(fit2)
Coefficients:
  Estimate Std. Error   z value  Pr(>|z|)
(Intercept)  -1.41838    0.23239   -6.103 1.04e-09 ***
race          0.08797    0.28547    0.308  0.758
---
Null deviance:  8.3499 on 3 degrees of freedom
Residual deviance: 8.2544 on 2 degrees of freedom
AIC: 29.731

> 1 - pchisq(8.2544-1.3835,1)
[1] 0.008761052

> library("lmtest")

> lrtest(fit2,fit)
Likelihood ratio test
Model 1: symptoms/n ~ race
Model 2: symptoms/n ~ race + azt
  Df LogLik   Df Chisq Pr(>Chisq)
1  2 -12.8654
2  3 -9.4299   1  6.8709   0.008761 **
SAS for AIDS and AZT use example

```sas
data aids;
input race $ azt $ y n @@;
datalines;
  White Yes 14 107  White No 32 113  Black Yes 11 63  Black No 12 55 ;
proc genmod; class race azt;
  model y/n = azt race / dist=bin type3 lrci residuals obstats;
proc logistic; class race azt / param=reference;
  model y/n = azt race / aggregate scale=none clparm=both clodds=both;
  output out=predict p=pi_hat lower=lower upper=upper;
proc print data=predict;
proc logistic; class race azt (ref=first) / param=ref;
  model y/n = azt / aggregate=(azt race) scale=none;
run;
```

**PROC GENMOD:**

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>1</td>
<td>1.3835</td>
<td>1.3835</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>1</td>
<td>1.3910</td>
<td>1.3910</td>
</tr>
</tbody>
</table>

**Analysis Of Maximum Likelihood Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Confidence Limits</th>
<th>Wald 68.27%</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.0736</td>
<td>0.2629</td>
<td>-1.3365 0.8106</td>
<td>16.67</td>
<td></td>
</tr>
<tr>
<td>azt</td>
<td>1</td>
<td>-0.7195</td>
<td>0.2790</td>
<td>-0.9984 0.4405</td>
<td>6.65</td>
<td></td>
</tr>
<tr>
<td>race</td>
<td>1</td>
<td>0.0555</td>
<td>0.2886</td>
<td>-0.2331 0.3441</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

**LR Statistics For Type 3 Analysis**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>azt</td>
<td>1</td>
<td>6.87</td>
<td>0.0088</td>
</tr>
<tr>
<td>race</td>
<td>1</td>
<td>0.04</td>
<td>0.8473</td>
</tr>
</tbody>
</table>
Estimated effects with constraints $\beta^A_2 = \beta^R_2 = 0$:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ML Estimate</th>
<th>Standard Error (SE)</th>
<th>Wald Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta^A_1$</td>
<td>$-0.720$</td>
<td>$0.279$</td>
<td>$6.65$</td>
</tr>
<tr>
<td>$\beta^R_1$</td>
<td>$0.055$</td>
<td>$0.289$</td>
<td>$0.04$</td>
</tr>
</tbody>
</table>

$$\hat{\beta}^A_1 - \hat{\beta}^A_2 = -0.720 - (0) = -0.720$$

$$= \log \left[ \frac{\hat{\pi}_{1j}}{1-\hat{\pi}_{1j}} \right] - \log \left[ \frac{\hat{\pi}_{2j}}{1-\hat{\pi}_{2j}} \right] = \log \left[ \frac{\hat{\pi}_{1j}/(1-\hat{\pi}_{1j})}{\hat{\pi}_{2j}/(1-\hat{\pi}_{2j})} \right]$$

$$= \text{log odds ratio between } A \text{ and response, at fixed level of } R$$
Thus, adjusting for race, odds of AIDS symptoms for those using AZT estimated to be $e^{-0.72} = 0.49$ times the odds for those not using it.

95% Wald CI for log odds ratio is $-0.720 \pm 1.96 (0.279)$, or $(-1.27, -0.17)$, which translates to $(e^{-1.27}, e^{-0.17}) = (0.28, 0.84)$ for odds ratio.

Note: Does not contain “no effect” value of 1.0

Similarly, for any choice of constraints,

$$\exp(\hat{\beta}_1^R - \hat{\beta}_2^R) = e^{0.055} = 1.06$$

i.e, estimated odds of symptoms when subject was white are 1.06 times estimated odds when subject was black, given AZT status.

(Not significantly different from 1.0)
For categorical $x$ with most counts $\geq 5$, we can use $X^2$ or $G^2$ to test fit of model (i.e., $H_0$: model holds)

If $n_i$ subjects at $i^{th}$ setting of $x$,
$\hat{\pi}_i = $ estimated $P(Y = 1)$ at that setting,
$\hat{\mu}_{i1} = n_i\hat{\pi}_i =$ predicted number of $(Y = 1)$ observations
$\hat{\mu}_{i2} = n_i(1 - \hat{\pi}_i) =$ predicted number of $(Y = 0)$ observations

Substitute counts of $(Y = 1)$ and $(Y = 0)$ and the fitted values into $X^2$ or $G^2$ (the deviance) to get statistic for testing fit

e.g., $X^2 = \sum$(observed - fitted)$^2$/fitted

When $\{\mu_i\}$ large, $X^2$ and $G^2 \rightarrow \chi^2$
$df =$ no. binomial observations $-$ no. parameters in model
Example: AIDS symptoms

\[ \hat{\alpha} = -1.074, \; \hat{\beta}_1^A = -0.720, \; \hat{\beta}_1^R = 0.055 \]

When \( A = 1 \) (yes), \( R = 1 \) (white), estimated probability of AIDS symptoms is

\[
\frac{e^{-1.074-0.720+0.055}}{1 + e^{-1.074-0.720+0.055}} = 0.150
\]

For 107 cases where \( A=1, \; R=1 \),

fitted number of “yes” responses on AIDS symptoms

\[ = 107 (0.150) = 16.0 \; (\text{observed 14}) \]

Fitted number of “no” responses

\[ = 107 (0.850) = 91.0 \; (\text{observed 93}) \]
Doing this for all 4 AZT–Race combinations yields sum over 8 cells,

\[ G^2 = 2 \sum \text{observed log} \left( \frac{\text{observed}}{\text{fitted}} \right) = 1.38. \]

For the 2 \( \times \) 2 \( \times \) 2 table, there are 2 \( \times \) 2 = 4 logits,
3 parameters in model \( \text{logit}(\pi) = \alpha + \beta_i^A + \beta_j^R \)
\( df = 4 - 3 = 1 \), so the fit is adequate.

\( G^2 \) (or \( X^2 = 1.39 \)) tests \( H_0 \): model holds; i.e., logit model with additive A and R effects holds, meaning there is no interaction between A and R in their effects on \( P(\text{Symptoms} = \text{yes}) \).

In GLM literature and R output, \( G^2 \) is the (residual) \textit{deviance}. 
Note

- Likelihood-ratio stat. $-2(L_0 - L_1)$ for testing $H_0: \beta_1^A = 0$, where $-2(L_0 - L_1) = -2(L_0 - L_S) - [−2(L_1 - L_S)] = \text{difference in deviances for models with and without AZT term}$. Note this is same for Bernoulli (ungrouped) or binomial (grouped) data file, because although $L_S$ differs for them, $L_S$ cancels in finding $-2(L_0 - L_1)$.

$(8.25 - 1.38 = 6.9, \ df = 2 - 1 = 1)$

- $X^2$ and $G^2 \not\rightarrow \chi^2$ when data sparse
  - small $n$
  - large $n$, lots of cells, small cell counts
  - continuous $x$

Can then judge goodness of fit by comparing model to more complex models (e.g., with interaction terms), or create pseudo chi-squared statistic for observed and fitted counts for partition of predictor space (e.g., “Hosmer-Lemeshow test,” CDA Sec. 5.2.5).

SAS: Hosmer-Lemeshow — LACKFIT option in PROC LOGISTIC

Generalized linear models useful for discrete or continuous data. Select $Y$ and its distribution (*random component*), predictors for linear predictor (*systematic component*), function of mean to model (*link function*). Same ML algorithm applies for all such models.

- Logit link useful for *binary* response, continuous or categorical $x$ or both (GLM with binomial response and logit link function).
- Logit effects interpreted multiplicatively using odds, odds ratios.
- ML model fitting requires iterative estimation, but simple with modern software.
- Inference (tests, CI’s) use likelihood in various ways, with likelihood-ratio methods preferable to Wald methods when $n$ small or parameters near boundary.
MODEL BUILDING

\((CDA, \text{ Chap. 6})\)

- Using likelihood-ratio \(G^2\) (deviance) to compare models
- Selecting explanatory variables for a model
- Residual analysis
- Testing conditional independence (using models or ‘Mantel-Haenszel’ methods)
- Sample size and power
- Infinite estimates and remedies
Nested models: $M_1$, $M_0$, with $M_0$ simpler

Likelihood-ratio stat. comparing models (twice the difference in maximized log-likelihoods) equals

$$G^2(M_0|M_1) = G^2(M_0) - G^2(M_1),$$

difference between deviance statistics.

**Example:** AIDS symptoms and AZT use, by race

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-linear model</th>
<th>$G^2$</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>$\log \left( \frac{\pi_{ij}}{1-\pi_{ij}} \right) = \alpha + \beta_i^A + \beta_j^R$</td>
<td>1.38</td>
<td>1</td>
</tr>
<tr>
<td>$M_0$</td>
<td>$\log \left( \frac{\pi_{ij}}{1-\pi_{ij}} \right) = \alpha + \beta_j^R$</td>
<td>8.25</td>
<td>2</td>
</tr>
</tbody>
</table>

$$G^2(M_0|M_1) = 8.25 - 1.38 = 6.87, \ df = 2 - 1 = 1$$

Tests $H_0 : \beta_i^A = 0$ in $M_1$

Simpler model fits significantly worse
Selecting Predictors for a Logistic Model (CDA, Sec. 6.1)

- Analogs exist of stepwise automatic selection procedures. e.g., backward elimination starts with all predictors, eliminates according to which has the largest $P$-value for a test of its effect.

- Usual regression considerations apply, such as (1) when some predictors are highly correlated, simpler models may achieve nearly as high a maximized log likelihood, (2) for very large $n$, an effect (e.g., a complex interaction) may be statistically significant without being practically useful (i.e., model parsimony is a sensible objective).

- Rough guideline – number of outcomes of each type (S, F) should exceed $5 \times$ (number of predictors), or you may have problems with severe bias in $\hat{\beta}_j$, infinite estimates.

- Analogs of $R^2$ for binary or count data are only partially successful (CDA, Sec. 6.3), but correlation between $Y$ and the predicted value (e.g., $\hat{P}(Y = 1)$ for logistic regression) is useful for comparing models.

- Alternative criterion: Minimize AIC = $-2$(maximized log likelihood – no. model parameters) to find model with best fit to population.
Example: Horseshoe crab data (CDA, Sec. 4.3.2, Sec. 6.1)
Full data at www.stat.ufl.edu/~aa/cda/data.html

$n = 173$ female crabs in data set

$Y =$ whether female horseshoe crab has ‘satellites’ (1 = yes, 0 = no)
$C =$ color (4 categories, light to dark)
$S =$ spine condition (3 categories)
$W =$ width of carapace shell (cm)
$WT =$ weight of crab (kg)

Weight and width are very highly correlated ($r = 0.89$), and we do not consider weight in this analysis.

Backward elimination results in a model with main effects for width and color.

Further simplification results from reducing color to two categories (dark, other).
### Results of fitting several logistic regression models to horseshoe crab data

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Deviance $G^2$</th>
<th>df</th>
<th>AIC</th>
<th>Models Compared</th>
<th>Deviance Difference</th>
<th>Correlation $R(y, \hat{\mu})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C<em>S + C</em>W + S*W</td>
<td>173.68</td>
<td>155</td>
<td>209.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>C + S + W</td>
<td>186.61</td>
<td>166</td>
<td>200.6</td>
<td>(2)–(1)</td>
<td>12.9 (df = 11)</td>
<td>0.456</td>
</tr>
<tr>
<td>3a</td>
<td>C + S</td>
<td>208.83</td>
<td>167</td>
<td>220.8</td>
<td>(3a)–(2)</td>
<td>22.2 (df = 1)</td>
<td>0.314</td>
</tr>
<tr>
<td>3b</td>
<td>S + W</td>
<td>194.42</td>
<td>169</td>
<td>202.4</td>
<td>(3b)–(2)</td>
<td>7.8 (df = 3)</td>
<td>0.402</td>
</tr>
<tr>
<td>3c</td>
<td>C + W</td>
<td>187.46</td>
<td>168</td>
<td>197.5</td>
<td>(3c)–(2)</td>
<td>0.8 (df = 2)</td>
<td>0.452</td>
</tr>
<tr>
<td>4a</td>
<td>C</td>
<td>212.06</td>
<td>169</td>
<td>220.1</td>
<td>(4a)–(3c)</td>
<td>24.5 (df = 1)</td>
<td>0.285</td>
</tr>
<tr>
<td>4b</td>
<td>W</td>
<td>194.45</td>
<td>171</td>
<td>198.5</td>
<td>(4b)–(3c)</td>
<td>7.0 (df = 3)</td>
<td>0.402</td>
</tr>
<tr>
<td>5</td>
<td>(C = dark) + W</td>
<td>187.96</td>
<td>170</td>
<td>194.0</td>
<td>(5)–(3c)</td>
<td>0.5 (df = 2)</td>
<td>0.447</td>
</tr>
<tr>
<td>6</td>
<td>None</td>
<td>225.76</td>
<td>172</td>
<td>227.8</td>
<td>(6)–(5)</td>
<td>37.8 (df = 2)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<sup>a</sup>C, color; S, spine condition; W, width.

C + S + W denotes the model with main effects of color, spine condition, width.

C*S + C*W + S*W denotes the model that has all the two-factor interactions as well as the “main effects.”

Note: In model 3c, with indicator (dummy) variables for the first three colors, logit[$\hat{P}(Y = 1)$] = $-12.715 + 1.33c_1 + 1.40c_2 + 1.11c_3 + 0.468$(width), the color estimates (1.33, 1.40, 1.11, 0) suggest it may be adequate to replace color as a factor by an indicator of whether a crab is dark.
Analysis of Residuals  (*CDA*, Sec. 6.2)

If $Y_i \sim \text{Bin}(n_i, \pi_i)$, with $\mu_i = n_i\pi_i$, **Pearson residual**

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{n_i\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

is commonly used for binomial models (e.g., logistic regression)

When summed over success and failure counts, these satisfy $X^2 = \sum_i e_i^2$.

However, substituting $\hat{\mu}_i$ for $\mu_i$ reduces variance of numerator, considerably if model has many parameters.

Better to *standardize* the Pearson residual:

$$r_i = \frac{y_i - \hat{\mu}_i}{SE(y_i - \hat{\mu}_i)} = \frac{\text{Pearson residual}}{\sqrt{1 - \text{leverage}_i}}$$

When model holds, these are asymptotically $N(0,1)$.

(We did this to follow up the test of independence on page 18.)

There are analogous *deviance* and *standardized deviance* residuals.
**Example:** Berkeley admissions data (CDA, p. 63)

<table>
<thead>
<tr>
<th>Department</th>
<th>Admitted, male</th>
<th></th>
<th>Admitted, female</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>A</td>
<td>512 (62%)</td>
<td>313</td>
<td>89 (82%)</td>
<td>19</td>
</tr>
<tr>
<td>B</td>
<td>353 (63%)</td>
<td>207</td>
<td>17 (68%)</td>
<td>8</td>
</tr>
<tr>
<td>C</td>
<td>120 (37%)</td>
<td>205</td>
<td>202 (34%)</td>
<td>391</td>
</tr>
<tr>
<td>D</td>
<td>138 (33%)</td>
<td>279</td>
<td>131 (35%)</td>
<td>244</td>
</tr>
<tr>
<td>E</td>
<td>53 (28%)</td>
<td>138</td>
<td>94 (24%)</td>
<td>299</td>
</tr>
<tr>
<td>F</td>
<td>22 (6%)</td>
<td>351</td>
<td>24 (7%)</td>
<td>317</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1198 (45%)</td>
<td>1493</td>
<td>557 (30%)</td>
<td>1278</td>
</tr>
</tbody>
</table>

Percents in table are percent admitted by gender and department.

Data nearly satisfy “Simpson’s paradox,” whereby marginal association has different direction than in each partial table.
A = Admitted (yes, no)
G = Gender (male, female)
D = Department (A, B, C, D, E, F)

2 × 2 × 6 table

Use logit model with A as response variable

<table>
<thead>
<tr>
<th>Model</th>
<th>$G^2$</th>
<th>df</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G + D$</td>
<td>20.2</td>
<td>5</td>
<td>no interaction</td>
</tr>
<tr>
<td>$D$</td>
<td>21.7</td>
<td>6</td>
<td>no gender effect</td>
</tr>
</tbody>
</table>

Logit model with $P(A = yes)$ dependent on department but independent of gender (given department) fits poorly, as does more complex model also including gender effect but assuming lack of interaction.

(Note: Similar example with different data set in *CDA*, Sec. 6.2.3.)
Conditional indep. model fits well for all departments except first.

<table>
<thead>
<tr>
<th>Dept.</th>
<th>Contribution</th>
<th>df</th>
<th>Sample odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19.05</td>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>B</td>
<td>0.26</td>
<td>1</td>
<td>0.80</td>
</tr>
<tr>
<td>C</td>
<td>0.75</td>
<td>1</td>
<td>1.13</td>
</tr>
<tr>
<td>D</td>
<td>0.30</td>
<td>1</td>
<td>0.92</td>
</tr>
<tr>
<td>E</td>
<td>0.99</td>
<td>1</td>
<td>1.22</td>
</tr>
<tr>
<td>F</td>
<td>0.38</td>
<td>1</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>21.7</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
R for logit model and standardized residuals with graduate admissions data, assuming no effect for gender:

```r
> Berkeley <- read.table("berkeley.dat", header=TRUE)
> Berkeley
         dept gender yes no
1         A  male   512 313
2         A female   89  19
3         B  male   353 207
4         B female   17   8
... 
11        F  male   22  351
12        F female   24  317
> attach(Berkeley)
> n <- yes + no

> fit <- glm(yes/n ~ factor(dept), weights=n, family=binomial)
> summary(fit)

Call:  glm(formula = yes/n ~ factor(dept), family = binomial, weights = n)

Coefficients:  
               Estimate Std. Error z value Pr(>|z|)  
(Intercept)     0.59346    0.06838  8.679  <2e-16 ***
factor(dept)B   -0.05059    0.10968 -0.461  0.645    
factor(dept)C   -1.20915    0.09726 -12.432 <2e-16 ***
factor(dept)D   -1.25833    0.10152 -12.395 <2e-16 ***
factor(dept)E   -1.68296    0.11733 -14.343 <2e-16 ***
factor(dept)F   -3.26911    0.16707 -19.567 <2e-16 ***

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 877.056 on 11 degrees of freedom
  Residual deviance: 21.736 on 6 degrees of freedom

> rstandard(fit, type="pearson")
          1          2          3          4          5          6          7
-4.1530728  4.1530728 -0.5037077  0.5037077  0.8680662 -0.8680662 -0.5458732
          8          9         10         11         12
  0.5458732  1.0005342 -1.0005342 -0.6197526  0.6197526
```

A. Agresti (UF)
SAS for the same logit model and standardized residuals:

```sas
data admit;
  input depart $ gender $ yes no @@; n = yes + no;
datalines;
  A m 512 313 A f 89 19
  B m 353 207 B f 17 8
  C m 120 205 C f 202 391
  D m 138 279 D f 131 244
  E m 53 138 E f 94 299
  F m 22 351 F f 24 317;
proc genmod; class depart gender;
  model yes/n = depart / dist=bin link=logit residuals obstats;
proc genmod; class depart gender;
  model yes/n = gender depart / dist=bin link=logit residuals obstats;
run;
```

Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>6</td>
<td>21.7355</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>6</td>
<td>19.9384</td>
</tr>
</tbody>
</table>

Observation Statistics

<table>
<thead>
<tr>
<th>YES</th>
<th>N</th>
<th>Pred</th>
<th>StReschi</th>
</tr>
</thead>
<tbody>
<tr>
<td>512</td>
<td>825</td>
<td>0.6442</td>
<td>-4.1531</td>
</tr>
<tr>
<td>89</td>
<td>108</td>
<td>0.6442</td>
<td>4.1531</td>
</tr>
<tr>
<td>353</td>
<td>560</td>
<td>0.6325</td>
<td>-0.5037</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
<td>0.6325</td>
<td>0.5037</td>
</tr>
<tr>
<td>120</td>
<td>325</td>
<td>0.3508</td>
<td>0.8681</td>
</tr>
<tr>
<td>202</td>
<td>593</td>
<td>0.3508</td>
<td>-0.8681</td>
</tr>
<tr>
<td>138</td>
<td>417</td>
<td>0.3396</td>
<td>-0.5459</td>
</tr>
<tr>
<td>131</td>
<td>375</td>
<td>0.3396</td>
<td>0.5459</td>
</tr>
<tr>
<td>53</td>
<td>191</td>
<td>0.2517</td>
<td>1.0005</td>
</tr>
<tr>
<td>94</td>
<td>393</td>
<td>0.2517</td>
<td>-1.0005</td>
</tr>
<tr>
<td>22</td>
<td>373</td>
<td>0.0644</td>
<td>-0.6198</td>
</tr>
<tr>
<td>24</td>
<td>341</td>
<td>0.0644</td>
<td>0.6198</td>
</tr>
</tbody>
</table>
Testing Conditional Independence (CDA, Sec. 6.4)

$H_0$: $X$ and $Y$ independent, given $Z$

Cochran-Mantel-Haenszel (CMH) test designed for $2 \times 2 \times K$ contingency tables.

A common application is multi-center clinical trials, such as $2 \times 2 \times 8$ table on p. 226 of CDA.

<table>
<thead>
<tr>
<th>Center</th>
<th>$X$</th>
<th>Response</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Drug</td>
<td>Success</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Placebo</td>
<td>Failure</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Drug</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$K$</td>
<td>Drug</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conditional on margins in each $2 \times 2$ layer, data determined by

$$\{n_{11k}, \ k = 1, \cdots, K\}$$

Under $H_0$,

$$\mu_{11k} = E(n_{11k}) = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

$$\text{var}(n_{11k}) = \left( n_{1+k}n_{2+k}n_{+1k}n_{+2k} \right) / n_{++k}(n_{++k} - 1)$$

Test statistic

$$M^2 = \frac{\left( \sum_k n_{11k} - \sum_k \mu_{11k} \right)^2}{\sum_k \text{var}(n_{11k})}$$

builds power by pooling results across strata.

(Note: Different from testing independence in marginal table!)
Properties

- $M^2 \rightarrow \chi^2_1$ under $H_0$, for large $n$
- Inappropriate when association “positive” in some strata, “negative” in others (Most powerful when odds ratio $\theta_{XY|Z}$ same for each $2 \times 2$)
- $H_0$ corresponds to $H_0 : \beta^X_1 = \beta^X_2 ( = 0)$ in model
  \[
  \log \left( \frac{P(Y=1)}{P(Y=2)} \right) = \alpha + \beta^X_i + \beta^Z_j
  \]
  (Equivalently, with two levels for $X$, $H_0 : \beta = 0$ in model
  logit[$P(Y = 1)$] = $\alpha + \beta x + \beta^Z_j$ with $x = 1$ for drug and 0 for placebo)

  CMH statistic can be derived as “score” test for this model, based on $\partial (\log \text{likelihood})/\partial \beta$ evaluated at $\beta = 0$.
- Alternatively, likelihood-ratio tests apply directly for the $XY$ association parameter in logit models
  e.g., test $\beta = 0$ by comparing log likelihoods for
  \[
  \text{logit}[P(Y = 1)] = \alpha + \beta x + \beta^Z_j
  \]
  and \[
  \text{logit}[P(Y = 1)] = \alpha + \beta^Z_j
  \]
Example: Treating fungal infections (*CDA*, Sec. 6.5.2)

<table>
<thead>
<tr>
<th>Center</th>
<th>Group</th>
<th>Success</th>
<th>Failure</th>
<th>$\mu_{11k}$</th>
<th>Sample $\hat{\theta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drug</td>
<td>0</td>
<td>5</td>
<td>0.0</td>
<td>undefined</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Drug</td>
<td>1</td>
<td>12</td>
<td>0.57</td>
<td>$\infty$</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Drug</td>
<td>0</td>
<td>7</td>
<td>0.0</td>
<td>undefined</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Drug</td>
<td>6</td>
<td>3</td>
<td>4.24</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Drug</td>
<td>5</td>
<td>9</td>
<td>3.50</td>
<td>3.33</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>2</td>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

“Positive” sample association for centers 2, 4, 5; tables 1 and 3 do not contribute to CMH statistic

$M^2 = 5.02, \; df = 1 \quad (P = 0.025)$
R for CMH test with fungal infection data:

> data <- c(0,0,5,9,1,0,12,10,0,0,7,5,6,2,3,6,5,2,9,12)
> data <- array(data, dim=c(2,2,5))
> mantelhaen.test(data, correct=FALSE)

Mantel-Haenszel chi-squared test without continuity correction

data:  data
Mantel-Haenszel X-squared = 5.017, df = 1, p-value = 0.0251
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
  1.184021 18.776825
sample estimates:
common odds ratio
  4.715098

> mantelhaen.test(data, correct=FALSE, exact=TRUE)

Exact conditional test of independence in 2 x 2 x k tables

data:  data  # Generalizes Fisher’s exact test to stratified data
S = 12, p-value = 0.03334
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
  1.01079 23.56849
sample estimates:
common odds ratio
  4.351693

---------------------------------------------------------------------
SAS for CMH test with fungal infection data:

```sas
data cmh;
input center group $ response count;
datalines;
  1 drug 1 0
  1 drug 2 5
  1 placebo 1 0
  1 placebo 2 9
...
  5 drug 1 5
  5 drug 2 9
  5 placebo 1 2
  5 placebo 2 12
;
proc freq;
  weight count;
  tables center*group*response / cmh ;
```

SUMMARY STATISTICS FOR GROUP BY RESPONSE CONTROLLING FOR CENTER

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Alternative Hypothesis</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nonzero Correlation</td>
<td>1</td>
<td>5.017</td>
<td>0.025</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>1</td>
<td>5.017</td>
<td>0.025</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>1</td>
<td>5.017</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Estimates of the Common Relative Risk (Row1/Row2)

<table>
<thead>
<tr>
<th>Type of Study</th>
<th>Method</th>
<th>Value</th>
<th>95% Confidence Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control</td>
<td>Mantel-Haenszel</td>
<td>4.715</td>
<td>1.214 18.316</td>
</tr>
<tr>
<td>(Odds Ratio)</td>
<td>Logit *</td>
<td>3.968</td>
<td>1.098 14.339</td>
</tr>
</tbody>
</table>
$M^2$ strictly for hypothesis testing. Should follow up with “Mantel-Haenszel odds ratio estimate” of a common or average odds ratio,

$$\hat{\theta}_{MH} = \frac{\sum_k (n_{11k} n_{22k} / n_{++k})}{\sum_k (n_{12k} n_{21k} / n_{++k})}$$

(ex. $\hat{\theta}_{MH} = 4.715$).

Or, fit logit model, and use ML estimate $\exp(\hat{\beta}^X_1 - \hat{\beta}^X_2)$

(ex. $\exp(\hat{\beta}^X_1 - \hat{\beta}^X_2) = \exp(1.546) = 4.693$).

When $n$ small, use exact test of conditional independence (In R output, two-sided $P$-value $= 0.033$).

Alternative approach treats centers as levels of random effect (see Sec. 6 of short course).

CMH test generalizes to $I \times J \times K$, for nominal and ordinal responses. (CDA, Sec. 8.4.3)
Sample Size and Power for Logistic Regression

\textit{(CDA, Sec. 6.6)}


Model \( \text{logit}(\pi) = \beta_0 + \beta_1 x \), with \( X \sim \text{approx. normal} \)

\( H_0 : \beta_1 = 0, \quad H_a : \beta_1 > 0 \)

Desire \( P(\text{Type I error}) = \alpha, P(\text{Type II error}) = \beta \).

Let \( \pi = \pi \) at \( x = \overline{x} \)

\[ \lambda = \log \theta \text{ comparing } \pi \text{ at } \overline{x} \text{ to } \pi \text{ at } \overline{x} + s_X \]

\[ \delta = [1 + (1 + \lambda^2)e^{5\lambda^2/4}][1 + e^{-\lambda^2/4}]^{-1} \]

Then,

\[ n = [z_\alpha + z_\beta e^{-\lambda^2/4}]^2 (1 + 2\pi\delta) / \pi \lambda^2 \]

For multiple predictors with \( \rho = \) multiple correlation between \( X \) and other predictors,

\[ n^* = n / (1 - \rho^2) \]
Example: $\pi = \text{prob(heart attack)}$

$x = \text{cholesterol (elderly population)}$

Suppose $\pi = 0.08$, and we want test to be sensitive to 50% increase (i.e., to 0.12) for a std. dev. increase in cholesterol.

$$\theta = \frac{0.12/0.88}{0.08/0.92} = 1.57, \quad \lambda = \log(1.57) = 0.450.$$  

For $\alpha = 0.05, \beta = 0.10, z_\alpha = 1.645, z_\beta = 1.28$

$\rightarrow \delta = 1.3, \quad n = 612$

For $x_1 = \text{cholesterol}, \ x_2 = \text{blood pressure}$, if $\rho = 0.40$, need about $n = 612/(1 - 0.4^2) = 729$ to test partial effect of cholesterol.
For ML estimation of logistic model parameters:

- At least one parameter estimate is infinite if can separate with a plane the \( x \) values where \( y = 1 \) and where \( y = 0 \).

**Complete separation:** No observations on that plane

**Quasi-complete separation:** On the plane boundary, both outcomes occur (common in contingency tables)

- Most software does not adequately detect this, as convergence of iterative model-fitting occurs at very large estimate, where log-likelihood looks flat.

- Reported standard errors (se) values also useless, because based on curvature at ML estimate (and log-likelihood essentially flat at convergence).

- Wald inference can be very poor whenever a \( \beta \) is very large, as \( \beta \) increases, se of \( \hat{\beta} \) grows faster than \( \beta \) (Hauck and Donner 1977).
Example: R for data where ML estimate is actually infinite

```r
> x <- c(10, 20, 30, 40, 60, 70, 80, 90)
> y <- c(0, 0, 0, 0, 1, 1, 1, 1)  # complete separation
> fit <- glm(y ~ x, family=binomial)
> summary(fit)

Coefficients:
               Estimate Std. Error z value Pr(>|z|) 
(Intercept)   -118.158   296046.187   0    1
           x          2.363       5805.939   0    1

Residual deviance: 2.1827e-10 on 6 degrees of freedom
```

Wald test says absolutely no evidence of effect!!
(Yet data suggest strong evidence by other criteria.)

We’d have quasi-complete separation if we add two observations at $x = 50$, one with $y = 1$ and one with $y = 0$ (still $\hat{\beta} = \infty$).
How can this happen with a categorical explanatory variable?

Complete separation:

<table>
<thead>
<tr>
<th>Group</th>
<th>Success</th>
<th>Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 ($x = 0$)</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Group 2 ($x = 1$)</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

Quasi-complete separation:

<table>
<thead>
<tr>
<th>Group</th>
<th>Success</th>
<th>Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 ($x = 0$)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Group 2 ($x = 1$)</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

If model contains interaction of $x$ with other variable, quasi-complete separation occurs if this happens at any particular value of other variable.
• With complete separation, model fit should give perfect predictions for \( Y \). So, warning sign is log-likelihood = 0, deviance = 0.

• Warning sign of complete or quasi-complete separation: Enormous se values (because of flat log-likelihood function, at convergence).

• For contingency table data, ML estimates necessarily exist if all counts positive. If at least one count = 0, the ML estimates may or may not exist.

• In principle, if estimated odds ratio equals 0 or \( \infty \), Wald test or CI is useless, but one can construct profile likelihood CI of form \([0, u]\) or \([l, \infty]\). e.g., use the R package \texttt{cond}\ for higher-order likelihood-based conditional inference (Alessandra Brazzale) or the R \texttt{ProfileLikelihood} package.

• Often, estimates of relevant parameters may be fine, but \( X^2, G^2 \) goodness-of-fit statistics unreliable if many \( \hat{\mu}_i \) small (\( X^2 \) better, usually fine if most \( \hat{\mu}_i \geq 5 \))

• \( G^2(M_0|M_1) \) more robust than \( G^2(M) \)

\( X^2(M_0|M_1) \) more robust than \( X^2(M) \)
**Example**: Effects of empty cells (quasi-complete separation), illustrated with fungal infection data

<table>
<thead>
<tr>
<th>Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Success</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

| Placebo     |          |          |
|            | 0        | 9        |

| 2         |

| Treatment   | Success  | Failure  |
|            | 1        | 12       |

| Placebo     |          |          |
|            | 0        | 10       |

| 3         |

| Treatment   | Success  | Failure  |
|            | 0        | 7        |

| Placebo     |          |          |
|            | 0        | 5        |

| 4         |

| Treatment   | Success  | Failure  |
|            | 6        | 3        |

| Placebo     |          |          |
|            | 2        | 6        |

| 5         |

| Treatment   | Success  | Failure  |
|            | 5        | 9        |

| Placebo     |          |          |
|            | 2        | 12       |
Model: \( \text{logit}[\text{Prob}(\text{Success})] = \alpha + \beta_i^T + \beta_j^C, \) with \( \beta_1^T = \beta_1^C = 0 \) for identifiability.

Equivalently, \( \text{logit}[P(S)] = \alpha_j^C + \beta x, \) where treatment \( x = 1 \) for drug and \( 0 \) for control, for which all parameters identifiable.

- Zero margins for centers 1,3, cause \( \hat{\alpha}_1^C = \hat{\alpha}_3^C = -\infty \)
- To estimate treatment effect \( \beta \), strategies include
  - remove “uninformative” centers 1,3
  - add small constant (e.g. \( 10^{-8} \)) to zero cells, so all estimates exist
  - combine some centers
- CMH test or exact test about treatment effect ignores centers 1,3

Example: Quasi-complete separation affects \{\( \hat{\alpha}_j^C \)\}, but not \( \hat{\beta} \).

If add any constant \( \leq 0.001 \) to empty cells, or if delete centers 1 and 3, \( \hat{\beta} = 1.55 \) (odds ratio = \( e^{1.55} = 4.7 \)), \( SE = 0.70, \ G^2 = 0.50. \)

If combine centers 1-3, \( \hat{\beta} = 1.56 \) \( (SE = 0.70). \)
R for sparse clinical trial data, where ML estimates for centers 1 and 3 are actually $-\infty$:

```r
> data <- read.table("fungal.dat",header=TRUE)
> data
  center treatment y  n
1       1         1 0  5
2       1         0 0  9
3       2         1 1 13
4       2         0 0 10
5       3         1 0  7
6       3         0 0  5
7       4         1 6  9
8       4         0 2  8
9       5         1 5 14
10      5         0 2 14
> attach(data)
> fit <- glm(y/n ~ treatment + factor(center), weights=n, family=binomial)
> summary(fit)
Coefficients:
         Estimate Std. Error z value Pr(>|z|) 
(Intercept) -2.459e+01 2.330e+04  -0.001 0.9992 
treatment    1.546e+00 7.017e-01  2.203  0.0276 *
factor(center)2  2.039e+01 2.330e+04  0.001 0.9993 
factor(center)3  4.809e-03 3.172e+04  0.000 1.0000 
factor(center)4  2.363e+01 2.330e+04  0.001 0.9992 
factor(center)5  2.257e+01 2.330e+04  0.001 0.9992 
---
Residual deviance: 0.50214 on 4 degrees of freedom
Number of Fisher Scoring iterations: 21
```

A. Agresti (UF) CDA March 8-10, 2016 86 / 191
Here are R results for alternate parameterization where center effects refer to them alone rather than to contrasts:

> fit2 <- glm(y/n ~ treatment + factor(center) -1, weights=n, family=binomial)
> summary(fit2)
Coefficients:

|                  | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------|----------|------------|---------|----------|
| treatment        | 1.5460   | 0.7017     | 2.203   | 0.027569 |
| factor(center)1  | -24.5922 | 23296.3959 | -0.001  | 0.999158 |
| factor(center)2  | -4.2025  | 1.1891     | -3.534  | 0.000409 |
| factor(center)3  | -24.5874 | 21523.6453 | -0.001  | 0.999089 |
| factor(center)4  | -0.9592  | 0.6548     | -1.465  | 0.142956 |
| factor(center)5  | -2.0223  | 0.6700     | -3.019  | 0.002540 |

---

Residual deviance: 0.50214 on 4 degrees of freedom

Again, \( P \)-values essentially equal 1 for estimates that are actually infinite!
SAS for sparse clinical trial data:

```sas
data clinical;
input center treat y n ;
datalines;
1 1 0 5
1 0 0 9
2 1 1 13
2 0 0 10
3 1 0 7
3 0 0 5
4 1 6 9
4 0 2 8
5 1 5 14
5 0 2 14
;
proc genmod data = clinical;
  class center;
  model y/n = treat center / dist=bin link=logit noint;
run;
```

### Criteria For Assessing Goodness Of Fit

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>4</td>
<td>0.5021</td>
<td>0.1255</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>4</td>
<td>0.3602</td>
<td>0.0900</td>
</tr>
</tbody>
</table>

### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 - 0.0000</td>
<td>0.00</td>
</tr>
<tr>
<td>treat</td>
<td>1</td>
<td>1.5460</td>
<td>0.7017</td>
<td>0.1708 - 2.9212</td>
<td>4.85</td>
</tr>
<tr>
<td>center</td>
<td>1</td>
<td>-28.0221</td>
<td>213410.4</td>
<td>-418305 - 418248.7</td>
<td>0.00</td>
</tr>
<tr>
<td>center</td>
<td>2</td>
<td>-4.2025</td>
<td>1.1891</td>
<td>-6.5331 - -1.8720</td>
<td>12.49</td>
</tr>
<tr>
<td>center</td>
<td>3</td>
<td>-27.9293</td>
<td>188688.5</td>
<td>-369851 - 369794.7</td>
<td>0.00</td>
</tr>
<tr>
<td>center</td>
<td>4</td>
<td>-0.9592</td>
<td>0.6548</td>
<td>-2.2426 - 0.3242</td>
<td>2.15</td>
</tr>
<tr>
<td>center</td>
<td>5</td>
<td>-2.0223</td>
<td>0.6700</td>
<td>-3.3354 - -0.7092</td>
<td>9.11</td>
</tr>
</tbody>
</table>

A. Agresti (UF)
Remedies for infinite estimates?

- **Bayesian approach**: Influence of prior distribution smooths data and results in finite posterior mean estimates. *(CDA, Sec. 7.2)*

- **Penalized likelihood approach**: Add a penalty term to the likelihood function. Maximizing the penalized likelihood results in shrinking estimates toward 0. *(Ref: Firth 1993, CDA, Sec. 6.5.3)*

R: Package *logistf* can do the Firth approach:

```r
> fit3 <- logistf(y/n ~ treatment + factor(center)-1, weights=n, family=binomial)
> summary(fit3)
```

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Chisq</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment</td>
<td>1.3678143</td>
<td>0.6436197</td>
<td>-3.125353</td>
<td>5.9101373</td>
<td>0.3426</td>
</tr>
<tr>
<td>factor(center)1</td>
<td>-4.00367</td>
<td>1.5193002</td>
<td>-8.998902</td>
<td>-1.6994870</td>
<td>17.897</td>
</tr>
<tr>
<td>factor(center)2</td>
<td>-3.63515</td>
<td>1.0063781</td>
<td>-8.204822</td>
<td>-0.995359</td>
<td>11.199</td>
</tr>
<tr>
<td>factor(center)3</td>
<td>-4.17072</td>
<td>1.5811491</td>
<td>-9.187891</td>
<td>-1.61435</td>
<td>14.201</td>
</tr>
<tr>
<td>factor(center)4</td>
<td>-0.84871</td>
<td>0.6264638</td>
<td>-5.897048</td>
<td>4.253802</td>
<td>0.0315</td>
</tr>
<tr>
<td>factor(center)5</td>
<td>-1.83285</td>
<td>0.6200202</td>
<td>-6.599538</td>
<td>2.995656</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Likelihood ratio test=40.72184 on 6 df, p=3.28493e-07, n=10
Wald test = 26.05109 on 6 df, p = 0.000217816
Example: Risk factors for endometrial cancer \((CDA, \text{Sec. 7.2.2, 7.4.8})\)

\(y = \text{histology (HG: 1 = high, 0 = low)}\)

\(x_1 = \text{neovasculuation (NV: 1 = present, 0 = absent)}\)

\(x_2 = \text{pulsatibility index of arteria uterina (PI: 0 to 49)}\)

\(x_3 = \text{endometrium height (EH: 0.27 to 3.61)}\).

\[
\begin{array}{ccccc|ccccc}
\text{HG} & \text{NV} & \text{PI} & \text{EH} & \text{HG} & \text{NV} & \text{PI} & \text{EH} \\
0 & 0 & 13 & 1.64 & 0 & 0 & 16 & 2.26 \\
\ldots & & & & \ldots & & & \\
1 & 1 & 21 & 0.98 & 1 & 0 & 5 & 0.35 \\
\end{array}
\]

Main effects model with \(n = 79\) patients,

\[
\logit[P(Y = 1)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3
\]

For all 13 patients having \(x_1 = 1\), we observe \(y = 1\).

There is quasi-complete separation, and ML est. \(\hat{\beta}_1 = \infty\).

The 95% profile likelihood confidence interval for \(\beta_1\) is \((1.28, \infty)\) and LR stat. for \(H_0: \beta_1 = 0\) is 9.36 (A. Brazzale \textit{cond} R package).
SAS for logistic and penalized logistic analyses for endometrial cancer data, with standardized quantitative predictors

```sas
data endometrial;
input nv pi eh hg ;
p2 = (pi-17.3797)/9.9978; eh2 = (eh-1.6616)/0.6621;
datalines;
   0 13 1.64 0
   0 16 2.26 0
... 0 33 0.85 1
;
proc logistic descending;
   model hg = nv p2 eh2 / clparm=pl;
proc logistic descending;
   model hg = nv p2 eh2 / firth clparm=pl;
run;
```

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>49.5092</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>37.5923</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>11.8109</td>
<td>3</td>
<td>0.0081</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<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</td>
<td>1</td>
<td>-1.2516</td>
<td>0.3688</td>
<td>11.5201</td>
<td>0.0007</td>
</tr>
<tr>
<td>nv</td>
<td>1</td>
<td>12.9546</td>
<td>206.9</td>
<td>0.0039</td>
<td>0.9501</td>
</tr>
<tr>
<td>p2</td>
<td>1</td>
<td>-0.4217</td>
<td>0.4432</td>
<td>0.9054</td>
<td>0.3413</td>
</tr>
<tr>
<td>eh2</td>
<td>1</td>
<td>-1.9218</td>
<td>0.5598</td>
<td>11.7841</td>
<td>0.0006</td>
</tr>
</tbody>
</table>
SAS output continued:

Profile Likelihood Confidence Interval for Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.2257</td>
<td>-0.7028</td>
</tr>
<tr>
<td>nv2</td>
<td>12.9546</td>
<td>1.2841</td>
</tr>
<tr>
<td>pi2</td>
<td>-0.4217</td>
<td>-1.3705 0.3818</td>
</tr>
<tr>
<td>eh2</td>
<td>-1.9218</td>
<td>-3.1688 -0.9510</td>
</tr>
</tbody>
</table>

FIRTH penalized likelihood:

Analysis of Maximum Likelihood Estimates

<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</td>
<td>1</td>
<td>-1.1565</td>
<td>0.3477</td>
<td>11.0635</td>
<td>0.0009</td>
</tr>
<tr>
<td>nv</td>
<td>1</td>
<td>2.9292</td>
<td>1.5507</td>
<td>3.5680</td>
<td>0.0589</td>
</tr>
<tr>
<td>pi2</td>
<td>1</td>
<td>-0.3474</td>
<td>0.3957</td>
<td>0.7708</td>
<td>0.3800</td>
</tr>
<tr>
<td>eh2</td>
<td>1</td>
<td>-1.7242</td>
<td>0.5138</td>
<td>11.2613</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

Profile Likelihood Confidence Interval for Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.1565</td>
<td>-1.9165 -0.5206</td>
</tr>
<tr>
<td>nv</td>
<td>2.9292</td>
<td>-1.2443 0.4045</td>
</tr>
<tr>
<td>pi2</td>
<td>-0.3474</td>
<td>-2.8902 -0.8162</td>
</tr>
<tr>
<td>eh2</td>
<td>-1.7242</td>
<td>-3.1688 -0.9510</td>
</tr>
</tbody>
</table>
R for logistic and penalized logistic analyses for endometrial cancer data, with standardized quantitative predictors

```r
> Endometrial <- read.table("endometrial.dat",header=TRUE)
> Endometrial
   nv pi eh hg
 1  1  13 1.64  0
 2  0  16 2.26  0
...  
79  0  33 0.85  1
> attach(Endometrial)
> pi2 <- (pi - mean(pi))/sd(pi)
> eh2 <- (eh - mean(eh))/sd(eh)
> fitML <- glm(hg ~ nv + pi2 + eh2, family=binomial)
> summary(fitML)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2517    0.3688   -3.394  0.00069
   nv          18.1856   1715.7509   0.011  0.99154
  pi2         -0.4217     0.4432   -0.952  0.34133
  eh2         -1.9219     0.5599   -3.433  0.00059

> install.packages("logistf")
> library("logistf")
> fitpenalized <- logistf(hg ~ nv + pi2 + eh2, family=binomial)
> summary(fitpenalized)

Confidence intervals and p-values by Profile Likelihood

       coef  se(coef) lower 0.95 upper 0.95  Chisq p
(Intercept) -1.1566    0.3477  -1.9167  -0.5207 13.82  2.01e-04
   nv         2.9293    1.5508   0.6097  7.8546  6.79  9.12e-03
  pi2        -0.3474    0.3957  -1.2443  0.4045  0.75  3.87e-01
  eh2        -1.7243    0.5138  -2.8903  0.8162 17.76  5.21e-05

Likelihood ratio test=43.65582 on 3 df, p=1.78586e-09, n=79
```
Compare nested models using $G^2$ (deviance) differences

Residuals may indicate how to amend model to improve fit.

Test conditional independence using CMH test or analogous likelihood-ratio test for logit or loglinear models.

Sample size and power approximations require strong assumptions, including normal dist. for $x$ and prediction of effect size.

With complete or quasi-complete separation in space of explanatory variable values, infinite estimates occur. Remedies include penalized likelihood and Bayesian approaches.
Association structure in multi-way contingency tables

Types of independence/dependence

Loglinear model formulas for various patterns of association

Goodness of fit

Correspondence with logit models
Illustrate with models for 3-way tables: (CDA, Sec. 9.2)

$I \times J \times K$ table for responses $X$, $Y$, $Z$

Multinomial probabilities $\{\pi_{ijk}\}$

Expected frequencies $\{\mu_{ijk} = n\pi_{ijk}\}$

Cell counts $\{n_{ijk}\}$

ML fitting treats counts as independent Poisson variates (GLM with Poisson random component, log link).

Conditional on $n$, Poisson $\rightarrow$ multinomial, get same ML parameter estimates for each.
Types of Independence / Dependence

(a) **Mutual independence**

\[ \pi_{ijk} = \pi_{i++} \pi_{+j+} \pi_{++k} \quad \text{all} \quad i, j, k \]

- Says \( P(X = i, Y = j, Z = k) = P(X = i)P(Y = j)P(Z = k) \) for all \( i, j, k \).

- Corresponds to loglinear model

\[ \log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \]

- Denote model by \((X, Y, Z)\)

- Almost always too simple to be of much practical use.
Example: Drug use in survey of high school seniors \((CDA, \text{Sec. 9.2.4})\)

- Alcohol use \((A)\)
- Cigarette use \((C)\)
- Marijuana use \((M)\)

<table>
<thead>
<tr>
<th>Alcohol Use</th>
<th>Cigarette Use</th>
<th>Marijuana Use</th>
<th>(A)</th>
<th>(C)</th>
<th>(M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>911</td>
<td>538</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>44</td>
<td>456</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>3</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>2</td>
<td>279</td>
<td></td>
</tr>
</tbody>
</table>

*Source: Prof. Harry Khamis, Wright State Univ.*

Loglinear models describe the association structure, treating all three variables as response variables.
R for loglinear modeling of high school survey data: Mutual independence model

> drugs <- read.table("drugs.dat",header=TRUE)
> drugs
   a  c  m  count
 1 yes yes yes  911
 2 yes yes no  538
 3 yes no yes  44
 4 yes no no  456
 5 no yes yes  3
 6 no yes no  43
 7 no no yes  2
 8 no no no  279
> alc <- factor(a); cig <- factor(c); mar <- factor(m)
> indep <- glm(count ~ alc + cig + mar, family=poisson(link=log), data=drugs)
> summary(indep) # loglinear model (A, C, M)

Coefficients:

                     Estimate Std. Error z value Pr(>|z|)
(Intercept)    6.2915405  0.0366669 171.558  < 2e-16 ***
alc2          -1.7851072  0.0597565 -29.872  < 2e-16 ***
cig2          -0.6493068  0.0441481 -14.707  < 2e-16 ***
mar2           0.3154228  0.0424352   7.431 1.08e-13 ***
---
Null deviance: 2851.5 on 7 degrees of freedom
Residual deviance: 1286.0 on 4 degrees of freedom
AIC: 1343.1
Number of Fisher Scoring iterations: 6

A. Agresti (UF)
CDA
March 8-10, 2016 99 / 191
(b) \( Y \) jointly independent of \( X \) and \( Z \)

\[
\begin{align*}
\pi_{ijk} &= \pi_{i+k}\pi_{j+} \quad \text{all } i, j, k \quad (IK \times J) \\
\log \mu_{ijk} &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}
\end{align*}
\]

- \( X \) and \( Z \) may be associated
- Denote by \((XZ, Y)\)
- Corresponds to ordinary independence in two-way table cross classifying \( Y \) with all combinations of values of \( X \) and \( Z \).
- \((X, Y, Z) \Rightarrow (XZ, Y), (YZ, X), (XY, Z)\)
  e.g, \((X, Y, Z)\) is special case of \((XZ, Y)\) with \( \lambda_{ik}^{XZ} = 0 \).
- Too simplistic to be very useful.
(c) $X$ and $Y$ **conditionally independent**, given $Z$

$$P(X = i, Y = j | Z = k) =$$

$$P(X = i | Z = k)P(Y = j | Z = k)$$

$$\pi_{ij} \mid k = \pi_{i+} \mid k \pi_{+j} \mid k \quad \text{all } i, j, k$$

$$\leftrightarrow \pi_{ijk} = \pi_{i+k} \pi_{+jk} / \pi_{++k}$$

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

- Denote by $(XZ, YZ)$

- $(X, Y, Z) \Rightarrow (Y, XZ) \Rightarrow (XZ, YZ)$

- $XY$ conditional independence $\not\Rightarrow XY$ marginal independence.
<table>
<thead>
<tr>
<th>Major</th>
<th>Gender</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal Arts</td>
<td>Female</td>
<td>0.18</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>Science or Engineering</td>
<td>Female</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.08</td>
<td>0.32</td>
</tr>
<tr>
<td>Total</td>
<td>Female</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

\[
\theta_{Lib. Arts} = \frac{0.18 \times 0.08}{0.12 \times 0.12} = 1.0, \quad \theta_{Sci. Eng.} = 1.0,
\]

Marginal \( \theta = \frac{0.2 \times 0.4}{0.2 \times 0.2} = 2.0 \). Why the discrepancy?

Note that conditional \( \theta = 6.0 \) between Major, Gender, and \( \theta = 6.0 \) between Major, Income.
(d) **No three-factor interaction (homogeneous association)**

\[
\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}
\]

- Denote by \((XY, XZ, YZ)\)
- Each pair of variables may be associated, conditionally and marginally.
- Implies odds ratio between two variables identical at each level of third variable (homogeneous association).
- Independence models are special cases \(\{\lambda_{ij}^{XY} = 0\}\) is \(XY\) conditional independence.

(e) **Three-factor interaction:** Adds \(\lambda_{ijk}^{XYZ}\)

**Goodness of Fit for Loglinear Models**

Once fitted values are obtained (e.g., by Newton-Raphson), one can test fit of model by comparing \(\{\hat{\mu}_{ijk}\}\) to \(\{n_{ijk}\}\) with \(X^2\) or \(G^2\).

\(df = \text{number of Poisson counts} - \text{number of parameters}\)
Goodness of fit of some models for student survey data:

<table>
<thead>
<tr>
<th>Model</th>
<th>$G^2$</th>
<th>$df$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AC, AM)</td>
<td>497.4</td>
<td>2</td>
</tr>
<tr>
<td>(AC, CM)</td>
<td>92.0</td>
<td>2</td>
</tr>
<tr>
<td>(AM, CM)</td>
<td>187.8</td>
<td>2</td>
</tr>
<tr>
<td>(AC, AM, CM)</td>
<td>0.37</td>
<td>1</td>
</tr>
</tbody>
</table>

Very poor fit when any association term excluded.

Loglinear model two-factor parameters relate to odds ratios.

E.g., when $(XY, XZ, YZ)$ or simpler model holds for $2 \times 2 \times k$ table, there is *homogeneous association*.

\[
\theta_{XY|Z=1} = \theta_{XY|Z=2} = \cdots = \theta_{XY|Z=k}
\]

\[
\hat{\theta}_{XY|Z} = \exp(\hat{\lambda}_{11}^{XY} + \hat{\lambda}_{22}^{XY} - \hat{\lambda}_{12}^{XY} - \hat{\lambda}_{21}^{XY})
\]

Software output shows estimated conditional odds ratios are:

\[
e^{2.05} = 7.8 \text{ for AC association}
\]

\[
e^{2.99} = 19.8 \text{ for AM association}
\]

\[
e^{2.85} = 17.3 \text{ for CM association}
\]

(95% Wald CI is $\exp[2.85 \pm 1.96(0.164)] = (12.5, 23.8)$)
R continued for loglinear modeling of high school survey data

```r
> homo.assoc <- update(indep, ~. + alc:cig + alc:mar + cig:mar)
> summary(homo.assoc) # loglinear model (AC, AM, CM)

            Estimate Std. Error  z value Pr(>|z|)  
(Intercept)   6.81387    0.03313 205.699 < 2e-16 ***
alc2         -5.52827    0.45221 -12.225 < 2e-16 ***
cig2         -3.01575    0.15162 -19.891 < 2e-16 ***
mar2          -0.52486    0.05428  -9.669 < 2e-16 ***
alc2:cig2      2.05453    0.17406  11.803 < 2e-16 ***
alc2:mar2      2.98601    0.46468   6.426  1.31e-10 ***
cig2:mar2      2.84789    0.16384  17.382 < 2e-16 ***

---
Residual deviance: 0.37399 on 1 degrees of freedom

> pearson <- summary.lm(homo.assoc)$residuals # Pearson residuals
> sum(pearson^2) # Pearson goodness-of-fit statistic
[1] 0.4011006
> std.resid <- rstandard(homo.assoc, type="pearson") # standardized
> expected <- fitted(homo.assoc) # estimated expected frequencies
> cbind(count, expected, pearson, std.resid)

    count expected      pearson    std.resid
 1     911  910.3832   0.02044342  0.6333249
 2     538  538.6168  -0.02657821 -0.6333249
 3      44   44.6168  -0.09234564 -0.6333251
 4     456  455.3832   0.02890528  0.6333249
 5      3    3.6168  -0.32434089 -0.6333251
 6     43   42.3832   0.09474777  0.6333249
 7      2    1.3832   0.52447895  0.6333251
 8     279  279.6168 -0.03688791 -0.6333249

---
Note the Pearson residuals take 8 different values, even though df = 1. Better to use standardized residuals, which reflect same difference between observed and fitted in each cell.
```
SAS for loglinear modeling of high school survey data:

```sas
data drugs;
input a c m count @@;
datalines;
  1 1 1 911 1 1 2 538 1 2 1 44 1 2 2 456
  2 1 1 3 2 1 2 43 2 2 1 2 2 2 2 279
;
proc genmod; class a c m;
  model count = a c m a*m a*c c*m / dist=poi link=log lrci type3 obstats;
run;
```

<table>
<thead>
<tr>
<th>Criterion</th>
<th>DF</th>
<th>Value</th>
<th>Value/DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>1</td>
<td>0.3740</td>
<td>0.3740</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>1</td>
<td>0.4011</td>
<td>0.4011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>5.6334</td>
<td>0.0597</td>
<td>5.5164 5.7504</td>
</tr>
<tr>
<td>a*m</td>
<td>1</td>
<td>2.9860</td>
<td>0.4647</td>
<td>2.0753 3.8968</td>
</tr>
<tr>
<td>a*m</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>a*m</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>a*m</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>a*c</td>
<td>1</td>
<td>2.0545</td>
<td>0.1741</td>
<td>1.7134 2.3957</td>
</tr>
<tr>
<td>a*c</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>a*c</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>a*c</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>c*m</td>
<td>1</td>
<td>2.8479</td>
<td>0.1638</td>
<td>2.5268 3.1690</td>
</tr>
<tr>
<td>c*m</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>c*m</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
<tr>
<td>c*m</td>
<td>2</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
</tr>
</tbody>
</table>

LR Statistics For Type 3 Analysis

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>a*m</td>
<td>1</td>
<td>91.64</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>a*c</td>
<td>1</td>
<td>187.38</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c*m</td>
<td>1</td>
<td>497.00</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Correspondence with logit models

Model \((AC, AM, CM)\) for drug data

\[
\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^C + \lambda_k^M + \lambda_{ij}^{AC} + \lambda_{ik}^{AM} + \lambda_{jk}^{CM}
\]

Treat \(M\) as response, \(A\) and \(C\) as explanatory

\[
\log \left( \frac{P(M = 1)}{P(M = 2)} \right) = \log \left( \frac{\mu_{ij1}}{\mu_{ij2}} \right) = \log \mu_{ij1} - \log \mu_{ij2}
\]

\[
= (\lambda_1^M - \lambda_2^M) + (\lambda_{i1}^{AM} - \lambda_{i2}^{AM}) + (\lambda_{j1}^{CM} - \lambda_{j2}^{CM})
\]

\[
= \alpha + \beta_i^A + \beta_j^C
\]

Residual \(df = 8 - 7\) (no. Poisson counts – no. loglinear para.)

\(= 4 - 3\) (no. binomial logits – no. logit parameters) \(= 1\)
i.e., we get same results for the association between $M$ and each of $A$ and $C$ if we treat the data as four binomials (instead of eight Poissons) and model the logit for marijuana use in terms of additive effects for alcohol use and cigarette use.

Illustration using R for the corresponding logistic model:

```r
> drugs2 <- read.table("drugs_binomial.dat", header=TRUE)
> drugs2
     A   C M_yes M_no  n
 1 yes yes   911  538 1449
 2 yes no    44  456  500
 3 no yes    3   43   46
 4 no no     2  279  281
> attach(drugs2)
> alc <- factor(A); cig <- factor(C)

> fit.logistic <- glm(M_yes/n ~ alc + cig, weights=n,
                      family=binomial(link=logit))

> summary(fit.logistic)
Coefficients:
             Estimate  Std. Error   z value  Pr(>|z|)
(Intercept) -5.309011  0.4752387 -11.17178  9.35e-26
alcyes      2.985974  0.4647313   6.42636  1.31e-10
 cigyes     2.847864  0.1638436  17.38187  7.47e-09
---
Null deviance: 843.82664 on 3 degrees of freedom
Residual deviance: 0.37399 on 1 degrees of freedom
```

A. Agresti (UF)  CDA  March 8-10, 2016  108 / 191
Loglinear models corresponding to logit models, with $M$ as response

<table>
<thead>
<tr>
<th>Loglinear</th>
<th>Logit =</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(M, AC)$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$(AM, AC)$</td>
<td>$\alpha + \beta_i^A$</td>
</tr>
<tr>
<td>$(CM, AC)$</td>
<td>$\alpha + \beta_j^C$</td>
</tr>
<tr>
<td>$(AM, CM, AC)$</td>
<td>$\alpha + \beta_i^A + \beta_j^C$</td>
</tr>
<tr>
<td>$(ACM)$</td>
<td>$\alpha + \beta_i^A + \beta_j^C + \beta_{ij}^{AC}$</td>
</tr>
</tbody>
</table>

Logit form of model more natural if there is a distinction between a response variable and a set of explanatory variables, and focus on effects of explanatory variables on the response.

Loglinear model more natural for several response variables (e.g., drug example) with focus on their association and interaction structure.

Equivalent loglinear model has most possible general term relating variables that are explanatory in logit model (e.g., with four predictors in logit, would have 4-factor interaction among them in loglinear model).
Hierarchy of loglinear models describes association patterns.

Conditional independence vs. marginal independence

$X^2$ and $G^2$ goodness of fit for non-sparse tables

Interpret conditional associations using odds ratios.

Loglinear-logit connection

Iterative methods required for ML fitting simple, since special case of GLM for Poisson random component with log link.
Nominal response models

e.g., model choice of product brand, or where choose to shop, or how get to work.
Standard modeling: Apply logit to pairs of categories.

Ordinal response models

e.g., patient quality of life (excellent, good, fair, poor),
political philosophy (very liberal, slightly liberal, moderate, slightly conservative, very conservative)
Standard modeling: Apply logits to cumulative probabilities.

In both cases, focus is on modeling how, for subject $i$,

$$
\pi_{ij} = P(Y_i = j), \quad j = 1, 2, \ldots, c,
$$

depends on explanatory variables $x$ (categorical and/or quantitative).
The models treat observations on $y$ at fixed $x$ as multinomial.
Nominal response: Baseline-category logits

Baseline category logits (CDA, Sec. 8.1) are

$$\log \left( \frac{\pi_{i1}}{\pi_{ic}} \right), \log \left( \frac{\pi_{i2}}{\pi_{ic}} \right), \ldots, \log \left( \frac{\pi_{i,c-1}}{\pi_{ic}} \right).$$

Model is

$$\log \left( \frac{\pi_{ij}}{\pi_{ic}} \right) = \beta_j^T x_i, \quad j = 1, \ldots, c - 1.$$

These $c - 1$ equations determine parameters for logits with other pairs of response categories, since

$$\log \left( \frac{\pi_{ia}}{\pi_{ib}} \right) = \log \left( \frac{\pi_{ia}}{\pi_{ic}} \right) - \log \left( \frac{\pi_{ib}}{\pi_{ic}} \right) = (\beta_a - \beta_b)^T x_i.$$

Equation that expresses multinomial logistic models directly in terms of response probabilities $\{\pi_{ij}\}$ is

$$\pi_{ij} = \frac{\exp(\beta_j^T x_i)}{1 + \sum_{h=1}^{c-1} \exp(\beta_h^T x_i)} \quad \text{with} \quad \beta_c = 0.$$
**Example:** Alligator food choice \((CDA, \text{Sec. 8.1.2})\)

<table>
<thead>
<tr>
<th>Lake</th>
<th>Size (meters)</th>
<th>Fish</th>
<th>Invertebrate</th>
<th>Reptile</th>
<th>Bird</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hancock</td>
<td>≤ 2.3</td>
<td>23</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>&gt; 2.3</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Oklawaha</td>
<td>≤ 2.3</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>&gt; 2.3</td>
<td>13</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Trafford</td>
<td>≤ 2.3</td>
<td>5</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>&gt; 2.3</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>George</td>
<td>≤ 2.3</td>
<td>16</td>
<td>19</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>&gt; 2.3</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

We’ll model primary food choice using Fish as the baseline category.
R for baseline-category logit model with alligator data

```r
> alligators <- read.table("alligators.dat",header=TRUE)
> alligators
  lake size y1 y2 y3 y4 y5
  1  1   1  23  4  2  2  8
  2  1   0  7  0  1  3  5
  3  2   1  5 11  1  0  3
  4  2   0 13  8  6  1  0
  5  3   1  5 11  2  1  5
  6  3   0  8  7  6  3  5
  7  4   1 16 19  1  2  3
  8  4   0 17  1  0  1  3
> attach(alligators)

> library(VGAM)

> fit <- vglm(formula = cbind(y2,y3,y4,y5,y1) ~ size + factor(lake),
       +   family=multinomial, data=alligators) # fish=1 is baseline category
```
R continued for baseline-category logit model

```r
> summary(fit)

<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>-3.20738</td>
<td>0.63873</td>
<td>-5.02147</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>-2.07176</td>
<td>0.70672</td>
<td>-2.93150</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>-1.39796</td>
<td>0.60852</td>
<td>-2.29731</td>
</tr>
<tr>
<td>(Intercept):4</td>
<td>-1.07808</td>
<td>0.47091</td>
<td>-2.28932</td>
</tr>
<tr>
<td>size:1</td>
<td>1.45820</td>
<td>0.39595</td>
<td>3.68285</td>
</tr>
<tr>
<td>size:2</td>
<td>-0.35126</td>
<td>0.58003</td>
<td>-0.60559</td>
</tr>
<tr>
<td>size:3</td>
<td>-0.63066</td>
<td>0.64248</td>
<td>-0.98160</td>
</tr>
<tr>
<td>size:4</td>
<td>0.33155</td>
<td>0.44825</td>
<td>0.73966</td>
</tr>
<tr>
<td>factor(lake):1</td>
<td>2.59558</td>
<td>0.65971</td>
<td>3.93442</td>
</tr>
<tr>
<td>factor(lake):2</td>
<td>1.21610</td>
<td>0.78045</td>
<td>1.54716</td>
</tr>
<tr>
<td>factor(lake):3</td>
<td>-1.34833</td>
<td>1.16353</td>
<td>-1.15882</td>
</tr>
<tr>
<td>factor(lake):4</td>
<td>-0.82054</td>
<td>0.72956</td>
<td>-1.12471</td>
</tr>
<tr>
<td>factor(lake):3:1</td>
<td>2.78034</td>
<td>0.67122</td>
<td>4.14220</td>
</tr>
<tr>
<td>factor(lake):3:2</td>
<td>1.69248</td>
<td>0.78045</td>
<td>2.16860</td>
</tr>
<tr>
<td>factor(lake):3:3</td>
<td>0.39265</td>
<td>0.78177</td>
<td>0.50226</td>
</tr>
<tr>
<td>factor(lake):3:4</td>
<td>0.69017</td>
<td>0.55967</td>
<td>1.23317</td>
</tr>
<tr>
<td>factor(lake):4:1</td>
<td>1.65836</td>
<td>0.61288</td>
<td>2.70586</td>
</tr>
<tr>
<td>factor(lake):4:2</td>
<td>-1.24278</td>
<td>1.18543</td>
<td>-1.04837</td>
</tr>
<tr>
<td>factor(lake):4:3</td>
<td>-0.69512</td>
<td>0.78126</td>
<td>-0.88974</td>
</tr>
<tr>
<td>factor(lake):4:4</td>
<td>-0.82620</td>
<td>0.55754</td>
<td>-1.48186</td>
</tr>
</tbody>
</table>

Residual deviance: 17.07983 on 12 degrees of freedom
Log-likelihood: -47.5138 on 12 degrees of freedom

> 1 - pchisq(17.07983,df=12)
[1] 0.146619```

A. Agresti (UF)
CDA
March 8-10, 2016
115 / 191
SAS for baseline-category logit model with alligator data

---------------
data gator;
input lake gender size food count;
datalines;
1 1 1 1 7
1 1 1 2 1
1 1 1 3 0
1 1 1 4 0
1 1 1 5 5
...
4 2 2 4 0
4 2 2 5 1
;
proc logistic; freq count;
class lake size / param=ref;
model food(ref='1') = lake size / link=glogit aggregate scale=none;
run;
---------------
## Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>food</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</td>
<td>2</td>
<td>1</td>
<td>-1.5490</td>
<td>0.4249</td>
<td>13.2890</td>
<td>0.0003</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>1</td>
<td>-3.3139</td>
<td>1.0528</td>
<td>9.9081</td>
<td>0.0016</td>
</tr>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1</td>
<td>-2.0931</td>
<td>0.6622</td>
<td>9.9894</td>
<td>0.0016</td>
</tr>
<tr>
<td>Intercept</td>
<td>5</td>
<td>1</td>
<td>-1.9043</td>
<td>0.5258</td>
<td>13.1150</td>
<td>0.0003</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>2</td>
<td>-1.6583</td>
<td>0.6129</td>
<td>7.3216</td>
<td>0.0068</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>3</td>
<td>1.2422</td>
<td>1.1852</td>
<td>1.0985</td>
<td>0.2946</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>4</td>
<td>0.6951</td>
<td>0.7813</td>
<td>0.7916</td>
<td>0.3736</td>
</tr>
<tr>
<td>lake</td>
<td>1</td>
<td>5</td>
<td>0.8262</td>
<td>0.5575</td>
<td>2.1959</td>
<td>0.1384</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>1</td>
<td>0.9372</td>
<td>1.1179</td>
<td>3.9443</td>
<td>0.0279</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>2</td>
<td>2.4583</td>
<td>1.1179</td>
<td>4.8360</td>
<td>0.0016</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>4</td>
<td>-0.6532</td>
<td>1.2021</td>
<td>0.2953</td>
<td>0.5869</td>
</tr>
<tr>
<td>lake</td>
<td>2</td>
<td>5</td>
<td>0.00565</td>
<td>0.7766</td>
<td>0.0001</td>
<td>0.9942</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>2</td>
<td>1.1220</td>
<td>0.4905</td>
<td>5.2321</td>
<td>0.0222</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>3</td>
<td>2.9347</td>
<td>1.1161</td>
<td>6.9131</td>
<td>0.0086</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>4</td>
<td>1.0878</td>
<td>0.8417</td>
<td>1.6703</td>
<td>0.1962</td>
</tr>
<tr>
<td>lake</td>
<td>3</td>
<td>5</td>
<td>1.5164</td>
<td>0.6214</td>
<td>5.9541</td>
<td>0.0147</td>
</tr>
<tr>
<td>size</td>
<td>2</td>
<td>1</td>
<td>1.4582</td>
<td>0.3959</td>
<td>13.5634</td>
<td>0.0002</td>
</tr>
<tr>
<td>size</td>
<td>3</td>
<td>1</td>
<td>-0.3513</td>
<td>0.5800</td>
<td>0.3668</td>
<td>0.5448</td>
</tr>
<tr>
<td>size</td>
<td>4</td>
<td>1</td>
<td>-0.6307</td>
<td>0.6425</td>
<td>0.9635</td>
<td>0.3263</td>
</tr>
<tr>
<td>size</td>
<td>5</td>
<td>1</td>
<td>0.3316</td>
<td>0.4483</td>
<td>0.5471</td>
<td>0.4595</td>
</tr>
</tbody>
</table>
Size of alligator has a noticeable effect.

E.g., prediction equation for log odds of selecting invertebrates instead of fish is

$$\log(\hat{\pi}_I/\hat{\pi}_F) = -3.207 + 1.458s + 2.596z_O + 2.780z_T + 1.658z_G.$$ 

For a given lake, for small alligators \((s = 1)\), estimated odds primary food choice was invertebrates instead of fish are \(\exp(1.458) = 4.30\) times estimated odds for large alligators \((s = 0)\).

Estimated effect imprecise, as Wald 95% CI is \(\exp[1.458 \pm 1.96(0.396)] = (1.98, 9.34)\).

Lake effects indicate that estimated odds that primary food choice was invertebrates instead of fish are relatively higher at Lakes Oklawaha, Trafford and George than at Lake Hancock.

Viewing all equations, size has greatest impact in terms of whether invertebrates rather than fish are primary food choice.
Y an ordinal response (c categories)

\( x \) an explanatory variable

Model \( P(Y_i \leq j), \ j = 1, 2, \ldots, c - 1, \) using logits

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

This is called a \textit{cumulative logit} model.

As in ordinary logistic regression, effects described by odds ratios. Here, we compare odds of being below vs. above any point on the response scale (\textit{cumulative odds ratios}).

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

$$\log \left[ \frac{P(Y \leq j \mid x_1)/P(Y > j \mid x_1)}{P(Y \leq j \mid x_2)/P(Y > j \mid x_2)} \right] = \beta(x_1 - x_2)$$

for all $j$ (called proportional odds property).

- $\beta = \textit{cumulative log odds ratio}$ for 1-unit increase in predictor.
- Model assumes effect $\beta$ is identical for every “cutpoint” for cumulative probability, $j = 1, \ldots, c - 1$.
- Logistic regression is special case $c = 2$.
- Software for maximum likelihood (ML) fitting includes R ($\textit{vglm}$ in VGAM library, $\textit{polr}$ (proportional odds logistic regression) in MASS library) and Stata ($\textit{ologit}, \textit{oglm}$).
Properties of cumulative logit models

- Model extends to multiple explanatory variables,

\[
\text{logit}[P(Y_i \leq j)] = \alpha_j + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}
\]

that can be qualitative or quantitative (use indicator variables for qualitative explanatory var’s).

- For subject \(i\) with values \(x_i\) on a set of explanatory variables, estimated conditional distribution function is

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

Estimated probability of outcome \(j\) is

\[
\hat{P}(Y_i = j) = \hat{P}(Y_i \leq j) - \hat{P}(Y_i \leq j - 1).
\]

- Can motivate proportional odds structure by a regression model for underlying continuous latent variable.
\( Y = \text{observed ordinal response} \)

\( Y^* = \text{underlying continuous latent variable,} \)

\[
Y^* = \beta^T x + \epsilon \quad \text{where } \epsilon \text{ has cdf } G \text{ with mean } 0. \text{ Thresholds (cutpoints)}
\]

\[-\infty = \alpha_0 < \alpha_1 < \ldots < \alpha_c = \infty \text{ such that} \]

\[
Y = j \quad \text{if } \alpha_{j-1} < Y^* \leq \alpha_j
\]

Then, at fixed \( x \) (see figure on next page)

\[
P(Y \leq j) = P(Y^* \leq \alpha_j) = P(Y^* - \beta^T x \leq \alpha_j - \beta^T x)
\]

\[
= P(\epsilon \leq \alpha_j - \beta^T x) = G(\alpha_j - \beta^T x)
\]

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

with \( G^{-1} \) a \textit{link function}.

Get cumulative logit model when \( G = \text{logistic cdf} \) \((G^{-1} = \logit)\).

**Note:** The model is often expressed (e.g., in Stata, SPSS) as

\[
\logit[P(Y \leq j)] = \alpha_j - \beta^T x.
\]

Then, \( \beta_j > 0 \) has usual interpretation of ‘positive’ effect.

Same fit, estimates, as using \( \alpha_j + \beta^T x \), except sign of \( \hat{\beta} \).
Note: This derivation suggests such models detect *location* effects (i.e., shifts in center), not dispersion effects (spread).

This model implies that dist. of $y$ at different settings of $x$ are *stochastically ordered*; i.e., the cdf at one setting is always above or always below the cdf at another level.
Example: Effect of intravenous medication doses on patients with subarachnoid hemorrhage trauma

<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>

Some indication that chance of death decreases as dose increases.

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

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

has ML estimate $\hat{\beta} = -0.176$ ($SE = 0.056$).

Likelihood-ratio test of $H_0 \beta = 0$ has test statistic $= 9.6$ ($df = 1, P = 0.002$).
R for modeling dose-response data, using `vglm` in VGAM library

```
> 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:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-0.72</td>
<td>0.159</td>
<td>-4.53</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>-0.32</td>
<td>0.156</td>
<td>-2.04</td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>0.69</td>
<td>0.158</td>
<td>4.38</td>
</tr>
<tr>
<td>(Intercept):4</td>
<td>2.06</td>
<td>0.174</td>
<td>11.84</td>
</tr>
<tr>
<td>dose</td>
<td>-0.18</td>
<td>0.056</td>
<td>-3.12</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>y1</th>
<th>y2</th>
<th>y3</th>
<th>y4</th>
<th>y5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.29</td>
<td>0.09</td>
<td>0.25</td>
<td>0.24</td>
<td>0.13</td>
</tr>
<tr>
<td>0.26</td>
<td>0.08</td>
<td>0.25</td>
<td>0.26</td>
<td>0.15</td>
</tr>
<tr>
<td>0.22</td>
<td>0.08</td>
<td>0.24</td>
<td>0.28</td>
<td>0.18</td>
</tr>
<tr>
<td>0.19</td>
<td>0.08</td>
<td>0.23</td>
<td>0.29</td>
<td>0.20</td>
</tr>
</tbody>
</table>

> vglm(cbind(y1,y2,y3,y4,y5) ~ 1,  # null model
      family=cumulative(parallel=TRUE), data=trauma)

Coefficients:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept):1</td>
<td>-1.14</td>
<td>0.75</td>
<td>0.25</td>
<td>1.60</td>
</tr>
<tr>
<td>(Intercept):2</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept):3</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept):4</td>
<td>1.60</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 dose effect
```
R for modeling dose-response data using \texttt{polr} in MASS library, for which response must be an ordered factor:

```
> trauma2 <- read.table("trauma2.dat", header=TRUE)
> trauma2
dose response count
  1 1 1 59
  2 1 2 25
...  20 4 5 41
> y <- factor(trauma2$response)
> fit.clogit <- polr(y ~ dose, data=trauma2, weight=count)
> summary(fit.clogit)
Coefficients:
               Value Std. Error t value
dose    0.1754816  0.05671224  3.094245
Intercepts:
   Value Std. Error t value
 1|2 -0.7192  0.1589 -4.5256
 2|3 -0.3186  0.1569 -2.0308
 3|4  0.6917  0.1597  4.3323
 4|5  2.0570  0.1751 11.7493
Residual Deviance: 2461.349
> fitted(fit.clogit)
   1    2    3    4    5
 1 0.2901 0.089 0.247 0.241 0.132
 2 0.2901 0.089 0.247 0.241 0.132
... 20 0.1944 0.070 0.232 0.297 0.205
```

Note: \texttt{polr} uses the model formula
\[
\text{logit}[P(y \leq j)] = \alpha_j - \beta^T x.
\]
PROC LOGISTIC; 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>

Parameter Estimates

<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</td>
<td>1</td>
<td>-0.7192</td>
<td>0.1588</td>
<td>20.5080</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.3186</td>
<td>0.1564</td>
<td>4.1490</td>
<td>0.0417</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.6916</td>
<td>0.1579</td>
<td>19.1795</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</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>
Goodness-of-fit statistics:

\[ \text{Pearson } X^2 = 15.8, \text{ deviance } G^2 = 18.2 \]

\( (df = 16 - 5 = 11) \)

\( P \)-values = 0.15 and 0.18

Model seems to fit adequately

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

\[ e^{0.176 \pm 1.96(0.056)} = (1.07, 1.33). \]

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

\[ e^{(4-1)0.176} = 1.69 \]
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, CDA p. 89).

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

Alternative analysis treats dose as factor, using indicator variables. Double the log-likelihood increases by only 0.13, $df = 2$.

With $\beta_1 = 0$ (R coding):
$\hat{\beta}_2 = -0.12, \hat{\beta}_3 = -0.32, \hat{\beta}_4 = -0.52$ ($SE = 0.18$ each)

With $\beta_4 = 0$ (SAS coding):
$\hat{\beta}_1 = 0.52, \hat{\beta}_2 = 0.40, \hat{\beta}_3 = 0.20$ ($SE = 0.18$ each)

Testing $H_0$: $\beta_1 = \beta_2 = \beta_3 = \beta_4$ 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 factor, using the `vglm` function in the `VGAM` library:

```r
> 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
```

Note that the factor effects (0, −0.12, −0.32, −0.52) are monotone decreasing, not far from linear in scores (1,2,3,4).
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

<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>18.0496</td>
<td>9</td>
<td>2.0055</td>
<td>0.0346</td>
</tr>
<tr>
<td>Pearson</td>
<td>15.7881</td>
<td>9</td>
<td>1.7542</td>
<td>0.0714</td>
</tr>
</tbody>
</table>

Model Fit Statistics

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</td>
<td>1</td>
<td>-1.4396</td>
<td>0.1416</td>
<td></td>
<td>103.3943</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>-1.0390</td>
<td>0.1369</td>
<td></td>
<td>57.6363</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept</td>
<td>3</td>
<td>-0.0286</td>
<td>0.1317</td>
<td></td>
<td>0.0472</td>
<td>0.8280</td>
</tr>
<tr>
<td>Intercept</td>
<td>4</td>
<td>1.3371</td>
<td>0.1428</td>
<td></td>
<td>87.7207</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>dose</td>
<td>1</td>
<td>0.5208</td>
<td>0.1779</td>
<td></td>
<td>8.5641</td>
<td>0.0034</td>
</tr>
<tr>
<td>dose</td>
<td>2</td>
<td>0.4032</td>
<td>0.1820</td>
<td></td>
<td>4.9072</td>
<td>0.0267</td>
</tr>
<tr>
<td>dose</td>
<td>3</td>
<td>0.2034</td>
<td>0.1779</td>
<td></td>
<td>1.3071</td>
<td>0.2529</td>
</tr>
</tbody>
</table>
Checking goodness of fit for contingency tables

- With nonsparse contingency table data, can check goodness of fit using Pearson $X^2$, deviance $G^2$.

- At setting $i$ of predictor 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 test statistic is $G^2 = 2 \sum_{i,j} n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right)$. 

  $df = \text{No. multinomial parameters} - \text{no. model parameters}$

- With sparse data, continuous predictors, can use such measures to compare nested models.
Other properties of cumulative logit models

- Can use similar model with alternative “cumulative link”
  \[
  \text{link}[P(Y_i \leq j)] = \alpha_j - \beta^T x_i \quad \text{(McCullagh 1980)}
  \]
e.g., *cumulative probit* model results from underlying normal \( Y^* \).

- Effects \( \beta \) invariant to choice and number of response categories (If model holds, same \( \beta \) when response scale collapsed in any way).

- 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

\[
\prod_{i=1}^{n} \left\{ \prod_{j=1}^{c} [P(Y_i = j \mid x_i)]^{y_{ij}} \right\} = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{c} [P(Y_i \leq j \mid x_i) - P(Y_i \leq j-1 \mid x_i)]^{y_{ij}} \right\}
\]

\[
= \prod_{i=1}^{n} \left\{ \prod_{j=1}^{c} \left[ \frac{\exp(\alpha_j + \beta^T x_i)}{1 + \exp(\alpha_j + \beta^T x_i)} - \frac{\exp(\alpha_{j-1} + \beta^T x_i)}{1 + \exp(\alpha_{j-1} + \beta^T x_i)} \right]^{y_{ij}} \right\}
\]
Model fitting requires iterative methods. Log likelihood is concave. To get standard errors, Newton-Raphson inverts observed information matrix 

$$-\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b}$$  (e.g., SAS PROC GENMOD)

Fisher scoring inverts expected information matrix 

$$E(-\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b})$$  (e.g., R vglm function, SAS PROC LOGISTIC).

McCullagh (1980) provided Fisher scoring algorithm for cumulative link models.

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 $\beta_j$. 

Wald: $z = \frac{\hat{\beta}_j - \beta_{j0}}{SE}$, or $z^2 \sim \chi^2$

Likelihood-ratio: $-2([L(\hat{\beta}_0) - L(\hat{\beta})] \sim \chi^2$
Alternative ways of summarizing effects

- 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).

- Summary measures of predictive power include
  
  (1) $R^2$ for regression model for underlying latent response variable (McKelvey and Zavoina 1975, provided by Stata)

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

  (3) To compare two groups with responses represented by independent r.v.’s $y_1$ and $y_2$, effect size measures

  \[ \Delta = P(y_1 > y_2) - P(y_2 > y_1), \quad \gamma = P(y_1 > y_2) + \frac{1}{2}P(y_1 = y_2) \]

  (Agresti and Kateri 2015). For latent normal linear model, $\gamma = \Phi(\beta/\sqrt{2})$ for $\beta$ coefficient of indicator for group variable.
Checking fit (general case) and selecting a model

- Lack of fit may result from omitted predictors (e.g., interaction), non-proportional odds effects, wrong link function. Often, lack of fit reflects effects of dispersion as well as location.
- Can check particular aspects of fit using likelihood-ratio test (change in deviance) to compare to more complex models.
- Some software provides score test of proportional odds assumption, but test is liberal (i.e., P(Type I error) too high). Like ratio test compares model to more general “non-proportional odds model” with effects $\{\beta_j\}$, but fitting of more general model fails when cumulative probabilities out-of-order.
- When model with proportional odds structure fits poorly, can use $\hat{\beta}_j$ in non-proportional odds model (e.g., after fitting binary logistic to each collapsing) to describe effects more fully.
- Even if proportional odds 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
R for modeling *dose-response* data without proportional odds, using `vglm()` in VGAM library without parallel=TRUE option

```r
> 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)

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

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
```

This test compares models with and without common $\beta_j$. The improvement in fit here is statistically significant ($P = 0.002$), but perhaps not substantively significant. Effect of dose is moderately negative for each cumulative probability, which simpler model summarizes by $\hat{\beta} = -0.175.$
Logit models for nominal response pair each category with a baseline, imply logits for each pair of categories.

Logit models for ordinal responses use logits of cumulative probabilities, with same effect for each logit (proportional odds). Interpret using odds ratios for binary collapsings of response (cumul. prob. and its complement).

We have not considered non-model based inferences (e.g., generalized CMH or measures of assoc.) also available for fixed or rank scores, which can often be viewed as “score tests” for models we presented (e.g., CDA, Sec. 8.4).

Handling dependence due to repeated measurement and other forms of clustering

Matched pairs: Comparing two marginal dependent proportions; connections between models and classical McNemar, CMH analyses

Marginal logit models: Comparing several dependent proportions, while adjusting for covariates; Generalized estimating equations (quasi-likelihood) method of estimating model parameters (GEE)

Generalized linear mixed models: Using random effects to model the dependence
**Comparing proportions for matched pairs**  
(*CDA*, Sec. 11.1)

**Example: President’s performance rated in successive months**

<table>
<thead>
<tr>
<th></th>
<th>Survey 1</th>
<th></th>
<th>Survey 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Approve</td>
<td>Disapprove</td>
<td></td>
</tr>
<tr>
<td>Approve</td>
<td>794</td>
<td>150</td>
<td>944</td>
</tr>
<tr>
<td>Disapprove</td>
<td>86</td>
<td>570</td>
<td>656</td>
</tr>
<tr>
<td></td>
<td>880</td>
<td>720</td>
<td>1600</td>
</tr>
</tbody>
</table>

The standard methods for comparing proportions are based on *independent* samples, but the row and column margins here are *dependent* samples.

Comparisons of $p_{1+}$ and $p_{+1}$ with test or CI must take into account that they’re calculated from same sample.
Consider $H_0: \pi_{1+} = \pi_{+1}$ for binary case

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>$\pi_{11}$</td>
<td>$\pi_{12}$</td>
</tr>
<tr>
<td>F</td>
<td>$\pi_{21}$</td>
<td>$\pi_{22}$</td>
</tr>
<tr>
<td></td>
<td>$\pi_{+1}$</td>
<td></td>
</tr>
</tbody>
</table>

Note $\pi_{1+} = \pi_{+1} \iff \pi_{12} = \pi_{21}$

Marginal homogeneity $\iff$ Symmetry (but only when $l = 2$)

For large $n$, can test $H_0$ using $z = \frac{p_{1+} - p_{+1}}{\sqrt{\text{Var}(p_{1+} - p_{+1})}}$,

and construct a CI, $(p_{1+} - p_{+1}) \pm 1.96 \sqrt{\text{Var}(p_{1+} - p_{+1})}$
Treating \( \{n_{ij}\} \) as multinomial \( \{\pi_{ij}\} \), then

\[
\text{Var}(p_{1+} - p_{+1}) = \frac{\pi_{1+}(1 - \pi_{1+}) + \pi_{+1}(1 - \pi_{+1}) - 2(\pi_{11}\pi_{22} - \pi_{12}\pi_{21})}{n}
\]

For matched samples, usually

\[
\pi_{11}\pi_{22} \gg \pi_{12}\pi_{21}
\]

\(\rightarrow\) variance is smaller than for indep. samples.

(Recall Var has form \( \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \) for indep. samples.)
\[
\begin{array}{c|cc|c}
& \text{Survey 1} & \text{Survey 2} \\
\hline
\text{Approve} & 794 & 150 & 944 \\
\text{Disapprove} & 86 & 570 & 656 \\
\hline
& 880 & 720 & 1600 \\
\end{array}
\]

Marginal proportions are \( \frac{944}{1600} = 0.59 \), \( \frac{880}{1600} = 0.55 \). Standard error of difference of marginal proportions is 0.0095.

\[
z = \frac{0.59 - 0.55}{0.0095} = 4.2 \text{ for } H_0 : \pi_1+ = \pi_+1
\]

95% CI for true difference is \( 0.04 \pm 1.96(0.0095) \), or (0.02, 0.06).

\[SE = 0.0175 \text{ for independent samples of sizes 1600 each having proportions 0.59 and 0.55.}\]
Small $n$

To test $H_0 : \pi_{1+} = \pi_{+1}$ ($\pi_{12} = \pi_{21}$), condition on $n_{12} + n_{21}$ and use $n_{12} \sim \text{binomial} \left( n_{12} + n_{21}, \frac{1}{2} \right)$ under $H_0$.

<table>
<thead>
<tr>
<th>$X$</th>
<th>$n_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{21}$</td>
<td>$X$</td>
</tr>
</tbody>
</table>

Test ignores main-diagonal counts. Cox (1958) justified this using logistic model for the matched pair $(Y_{s1}, Y_{s2})$ for subject $s$.

Model

\[
\log \left( \frac{P(Y_{st} = 1)}{P(Y_{st} = 0)} \right) = \alpha_s, \quad \text{time 1}
\]

\[
= \alpha_s + \beta, \quad \text{time 2}
\]

Here, $\beta$ compares margins as $\exp(\beta) = \frac{P(Y_{s2}=1)/P(Y_{s2}=0)}{P(Y_{s1}=1)/P(Y_{s1}=0)}$

\{$\alpha_s$\} are nuisance parameters.

After conditioning to eliminate \{$\alpha_s$\}, conditional ML estimate

$\exp(\hat{\beta}) = \frac{n_{12}}{n_{21}}$ (CDA, Sec. 11.2.2, 11.2.3).
To test marginal homogeneity ($\beta = 0$), normal approximation to binomial gives

$$z = \frac{n_{12} - \left(\frac{1}{2}\right)(n_{12} + n_{21})}{\sqrt{(n_{12} + n_{21})\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}} = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}}$$

$$z^2 \rightarrow \chi_1^2 \text{ (McNemar's test)}$$

ex. $z = \frac{150 - 86}{\sqrt{150 + 86}} = 4.16$

McNemar test is special case of Cochran's Q test, which is a $\chi^2_{T-1}$ test for testing equality of correlated proportions at $T$ occasions.
**McNemar stat.** = *Cochran-Mantel-Haenszel* stat. applied to *n* $2 \times 2$ tables, where $k^{th}$ table gives responses for pair $k$.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time</th>
<th>Approve</th>
<th>Disapp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

This is useful representation for suggesting tests for case-control data with several controls per case, or for ordinal responses or multiple times.
A model for marginal logits is a multivariate model that simultaneously forms logits for each margin.

Awkward to apply ML, especially for large tables, because model applies to marginal distributions but likelihood is in terms of joint multinomial probabilities.

ML requires special software, such as R function `mph.fit` available from Joseph Lang at Univ. of Iowa, and even that can handle only relatively simple cases (e.g., a few categorical predictors).

Also possible is `hmm` R package for hierarchical multinomial marginal models (Colombi, Giordano, Cazzaro).
Generalized Estimating Equations (GEE) Approach

(CDA, Sec. 12.2, 12.3)

Multivariate generalization of quasi-likelihood methods. Estimates are solutions of equations similar to likelihood equations, without fully specifying distribution.

GEE useful when primary interest is modeling marginal dist. of $Y_t$ as fn. of $x$’s, rather than modeling association among $(Y_1, Y_2, \ldots, Y_T)$ as loglinear models do. Not necessary to assume full multivariate distribution.

Steps of GEE Methodology:

- Assume marginal regression model, “working” covariance structure (e.g., exchangeable, autoregressive, independence).
- Estimates are consistent even if covariance structure misspecified (if marginal model correct).
- Method generates robust estimates of std. errors that are valid even if covariance structure misspecified. Uses “sandwich” covariance matrix incorporating empirical variability.
Some motivation, first for univariate QL:

- The quasi-likelihood $\hat{\beta}$ are solutions of quasi-score “estimating equations”
  \[
  u(\beta) = \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \frac{(y_i - \mu_i)}{v(\mu_i)} = 0.
  \]

- The QL estimators $\hat{\beta}$ are asymptotically normal with model-based covariance matrix approximated by
  \[
  V = \left[ \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \left[ v(\mu_i) \right]^{-1} \left( \frac{\partial \mu_i}{\partial \beta} \right) \right]^{-1}.
  \]

- To find actual $\text{var}(\hat{\beta})$, we use $u(\hat{\beta}) \approx u(\beta) + \frac{\partial u(\beta)}{\partial \beta} (\hat{\beta} - \beta)$.

  Since $u(\hat{\beta}) = 0$, $(\hat{\beta} - \beta) \approx -\left( \frac{\partial u(\beta)}{\partial \beta} \right)^{-1} u(\beta)$.

- But $[\partial u(\beta)/\partial \beta]$ is Hessian matrix for quasi log-likelihood, so $-\left[ \partial u(\beta)/\partial \beta \right]^{-1}$ is analog of inverse observed information matrix for specified model and approximates model-based $V$. Thus,
  \[
  \text{var}(\hat{\beta}) \approx V \{ \text{var}[u(\beta)] \} V
  \]
Also,

\[
\text{var}[u(\beta)] = \text{var} \left[ \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \frac{(y_i - \mu_i)}{v(\mu_i)} \right] = \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \frac{\text{var}(y_i)}{[v(\mu_i)]^2} \left( \frac{\partial \mu_i}{\partial \beta} \right).
\]

In summary, actual asymptotic covariance matrix of \( \hat{\beta} \) is

\[
\text{var}(\hat{\beta}) \approx V \left[ \sum_{i=1}^{n} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T \frac{\text{var}(y_i)}{[v(\mu_i)]^2} \left( \frac{\partial \mu_i}{\partial \beta} \right) \right] V.
\]

In practice, true \( \text{var}(y_i) \) is unknown. Can estimate \( \text{var}(\hat{\beta}) \) by sample analog (sandwich estimator), replacing \( \mu_i \) by \( \hat{\mu}_i \) and \( \text{var}(y_i) \) by \( (y_i - \hat{\mu}_i)^2 \).

For multivariate QL (GEE), with \( D_i = \frac{\partial \mu_i}{\partial \beta} \), analog is

\[
\text{var}(\hat{\beta}) \approx n \left[ \sum_{i=1}^{n} D_i^T V_i^{-1} D_i \right]^{-1} \left[ \sum_{i=1}^{n} D_i^T V_i^{-1} \text{var}(y_i) V_i^{-1} D_i \right] \left[ \sum_{i=1}^{n} D_i^T V_i^{-1} D_i \right]^{-1}
\]

where \( V_i \) is working covariance matrix.
Originally specified (Liang and Zeger 1986) for univariate \( y_t \) (e.g., binomial, Poisson), but extensions exist for cumulative logit models.

Some versions base working correlation on structure for odds ratios (e.g., exchangeable), more natural for categorical responses.

No likelihood function, since do not fully specify joint dist of \((y_1, y_2, \ldots, y_T)\).

In GEE methodology, cannot use standard likelihood-ratio tests, model comparison, tests of fit.

R: gee package has gee function that handles binary data. Package repolr contains repolr (repeated proportional odds logistic regression) for ordinal GEE analyses. Touloumis (2012) uses odds ratios for multinomial data (ordinal or nominal) with multgee R package. SAS: PROC GENMOD, with Lipsitz et al. (1994) approach for ordinal data available only with independence working equations.
Example (binary data): Support for legalizing abortion in three situations

<table>
<thead>
<tr>
<th>Gender</th>
<th>General Social Survey Sequence of Responses in Three Situations</th>
<th>(1,1,1)</th>
<th>(1,1,0)</th>
<th>(0,1,1)</th>
<th>(0,1,0)</th>
<th>(1,0,1)</th>
<th>(1,0,0)</th>
<th>(0,0,1)</th>
<th>(0,0,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td>342</td>
<td>26</td>
<td>6</td>
<td>21</td>
<td>11</td>
<td>32</td>
<td>19</td>
<td>356</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>440</td>
<td>25</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>47</td>
<td>22</td>
<td>457</td>
</tr>
</tbody>
</table>

Situations are (1) if the family has a very low income and cannot afford any more children, (2) when the woman is not married and does not want to marry the man, and (3) when the woman wants it for any reason. 1, yes; 0, no.

Note: Overwhelming majority of responses are (0,0,0) and (1,1,1), suggesting strong pairwise associations.

Marginal model:

$$\text{logit}[P(Y_{ij} = 1)] = \alpha + \beta_j + \gamma x_i.$$  

where $x_i = 1$ for females and 0 for males and situation effects $\{\beta_j\}$ satisfy constraint such as $\beta_3 = 0$. 
R fitting, assuming exchangeable working correlation structure:

```r
> abortion <- read.table("abortion.dat",header=TRUE)
> abortion
     gender response question case
 1      1         1         1    1
 2      1         1         2    1
 3      1         1         3    1
   ...
5548    0         0         1 1850
5549    0         0         2 1850
5550    0         0         3 1850

> z1 <- ifelse(abortion$question==1,1,0)
> z2 <- ifelse(abortion$question==2,1,0)
> library(gee)
> fit.gee <- gee(response ~ gender + z1 + z2, id=case, family=binomial,
+    corstr="exchangeable", data=abortion)
> summary(fit.gee)

                  Estimate   Naive S.E.   Naive z  Robust S.E.  Robust z
  (Intercept) -0.125325730  0.06782579 -1.84775925 0.06758 212 -1.85442135
   gender  0.003437873  0.08790630  0.03910838 0.08784072 0.03913758
     z1    0.149347107  0.02814374  5.30658404 0.02973865 5.021987 29
     z2    0.052017986  0.02815145  1.84779075 0.02704703 1.923241 79

Working Correlation
       [,1]       [,2]       [,3]
[1,] 1.0000000 0.8173308 0.8173308
[2,] 0.8173308 1.0000000 0.8173308
[3,] 0.8173308 0.8173308 1.0000000
```
R fitting, assuming independence working correlation structure:

```
> abortion <- read.table("abortion.dat",header=TRUE)
> abortion
   gender response question case
1       1        1         1     1
2       1        1         2     1
3       1        1         3     1
...  
5548    0        0         1     1850
5549    0        0         2     1850
5550    0        0         3     1850

> z1 <- ifelse(abortion$question==1,1,0)
> z2 <- ifelse(abortion$question==2,1,0)
> library(gee)
> fit.gee2 <- gee(response ~ gender + z1 + z2, id=case, family=binomial,
+  corstr="independence", data=abortion)
> summary(fit.gee2)

                 Estimate  Naive S.E.  Naive z  Robust S.E.  Robust z
  (Intercept)  -0.125407576  0.05562131 -2.25466795  0.06758012 -1.85562596
   gender    0.003582051  0.05415761  0.06614123  0.08784012  0.04779210
     z1       0.149347113  0.06584875  2.26803253  0.02973865  5.02198753
     z2       0.052017989  0.06586692  0.78974374  0.02704704  1.92324166
```

Because of strong positive correlation between pairs of responses (estimated exchangeable correlation = 0.817), naive SE’s based on independence working correlations are very misleading.
Multinomial response: Touloumis approach (Touloumis, Agresti, Kateri 2013) characterizes pairwise associations by *local odds ratios*

\[
\theta^L_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i,j+1}\pi_{i+1,j}}
\]

For \(r \times c\) contingency table with ordered rows and columns represented by ordered row scores \(\{u_i\}\) and ordered column scores \(\{v_j\}\), *linear-by-linear association model* for expected frequencies \(\mu_{ij} = n\pi_{ij}\) is

\[
\log \mu_{ij} = \lambda + \lambda_i^x + \lambda_j^y + \beta u_i v_j.
\]

With \(\{u_i = i\}, \{v_j = j\}\), each \(\log(\theta^L_{ij}) = \beta\) (*uniform association model*).

In practice, with repeated measurement we usually expect positive assoc., \(\beta > 0\).

In GEE context, \(r = c\) and we take \(u_i = v_i\). With nominal variables, we treat scores as parameters, as no ordering is specified.
Example: Randomized Clinical Trial for Treating Insomnia

Randomized, double-blind clinical trial compared hypnotic drug with placebo in patients with insomnia problems

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Initial</th>
<th>Follow-up</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;20</td>
<td>20–30</td>
<td>30–60</td>
<td>&gt;60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–30</td>
<td>11</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–60</td>
<td>13</td>
<td>23</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;60</td>
<td>9</td>
<td>17</td>
<td>13</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–30</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–60</td>
<td>6</td>
<td>9</td>
<td>18</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;60</td>
<td>4</td>
<td>11</td>
<td>14</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ Y_t = \text{time to fall asleep} \]
\[ x = \text{treatment} (0 = \text{placebo}, 1 = \text{active}) \]
\[ t = \text{occasion} (0 = \text{initial}, 1 = \text{follow-up after 2 weeks}) \]

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

GEE estimates (with independence working equations and robust SE): 
\[
\hat{\beta}_1 = 1.04 \ (SE = 0.17), \quad \text{placebo occasion effect}
\]
\[
\hat{\beta}_2 = 0.03 \ (SE = 0.24), \quad \text{treatment effect initially}
\]
\[
\hat{\beta}_3 = 0.71 \ (SE = 0.24), \quad \text{interaction}
\]

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

Odds ratios \(e^{1.04} = 2.8\), \(e^{1.75} = 5.7\)

Treatment effect \(e^{0.03} = 1.03\) initial odds ratio, 
\(e^{0.03+0.71} = 2.1\) follow-up odds ratio
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>0</td>
<td>0</td>
<td>1</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>-2.26709</td>
<td>0.20274</td>
<td>-11.182</td>
<td>0.21877</td>
<td>-10.36333</td>
</tr>
<tr>
<td>factor(cuts)2</td>
<td>-0.95146</td>
<td>0.17848</td>
<td>-5.331</td>
<td>0.18092</td>
<td>-5.2591017</td>
</tr>
<tr>
<td>factor(cuts)3</td>
<td>0.35174</td>
<td>0.17269</td>
<td>2.037</td>
<td>0.17842</td>
<td>1.9713794</td>
</tr>
<tr>
<td>treat</td>
<td>0.03361</td>
<td>0.23690</td>
<td>0.142</td>
<td>0.23844</td>
<td>0.1409595</td>
</tr>
<tr>
<td>occasion</td>
<td>1.03808</td>
<td>0.23760</td>
<td>4.369</td>
<td>0.18092</td>
<td>6.1943093</td>
</tr>
<tr>
<td>treat:occasion</td>
<td>0.70776</td>
<td>0.33418</td>
<td>2.179</td>
<td>0.24352</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
case treat y1 y2
1 1 1 1 1
...
239 239 0 4 4
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 **   |

Local Odds Ratios Estimates:

```
[,1] 0.000 0.000 0.000 1.721 1.721 1.721
[,2] 0.000 0.000 0.000 1.721 1.721 1.721
[,3] 0.000 0.000 0.000 1.721 1.721 1.721
[,4] 1.721 1.721 1.721 0.000 0.000 0.000
[,5] 1.721 1.721 1.721 0.000 0.000 0.000
[,6] 1.721 1.721 1.721 0.000 0.000 0.000
```

A. Agresti (UF)
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

| Parameter        | Estimate | Error  | Limits     | Z    | Pr > |Z| |
|------------------|----------|--------|------------|------|------|---|
| Intercept1       | -2.2671  | 0.2188 | -2.6959    | 1.8383 -10.36 | < .0001 |
| Intercept2       | -0.9515  | 0.1809 | -1.3061    | 0.5969 -5.26  | < .0001 |
| Intercept3       | 0.3517   | 0.1784 | 0.0020     | 0.7014 1.97   | 0.0487  |
| treat            | 0.0336   | 0.2384 | -0.4337    | 0.5009 0.14   | 0.8879  |
| occasion         | 1.0381   | 0.1676 | 0.7096     | 1.3665 6.19   | < .0001 |
| treat*occasion   | 0.7078   | 0.2435 | 0.2305     | 1.1850 2.91   | 0.0037  |
Alternative (*transitional*) model (*CDA*, Sec. 12.4)  
$$\text{logit}[P(Y_2 \leq j)] = \alpha_j + \beta_1 x + \beta_2 y_1$$

This is an ordinary univariate model that can be fitted easily by ML (e.g., in R with `vglm()` function in VGAM library, and with PROC GENMOD or PROC LOGISTIC in SAS), 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.

This approach has advantages over marginal models when initial marginal distributions differ (*CDA*, Sec. 12.4.5).
Random Effects Models for Clustered Data (CDA, Chap. 13)

For binary response, “random intercept model” for observation \( t \) in cluster (e.g., subject) \( i \) is the \textit{generalized linear mixed model} (GLMM)

\[
\text{logit}[P(Y_{it} = 1)] = u_i + \alpha + \beta^T x_{it}
\]

where \( \{u_i\} \) are \textit{iid} from \( \text{N}(0, \sigma^2) \). Introducing random \( u_i \) in model induces correlation between repeated responses because of \textit{subject heterogeneity}.

- e.g., Large positive \( u_i \) implies high \( P(Y_{it} = 1) \) for each \( t \) (tend to result in \( (Y_{i1}, Y_{i2}) = (1, 1) \) sequences of responses).
- Large negative \( u_i \) implies high \( P(Y_{it} = 0) \) each \( t \) (tend to result in \( (Y_{i1}, Y_{i2}) = (0, 0) \) sequences).
- As \( \sigma \) increases, \( (Y_{i1}, Y_{i2}) \) correlation increases.
- Recall that if \( Y_1 = U + X, Y_2 = U + Z \), with \( U, X, \) and \( Z \) uncorrelated, then \( \text{Corr}(y_1, y_2) = \frac{\text{Var}(U)}{\sqrt{\text{Var}(U) + \text{Var}(X)}[\text{Var}(U) + \text{Var}(Z)]} \).
- \( \sigma = 0 \) corresponds to repeated responses being independent.
Analogous models exist for other links, multinomial responses.

Can have differing numbers of observations per cluster and explanatory variables with values varying by observation.

GLMM approach has disadvantage of adding normal dist. assumption about random effect, but results usually robust to that choice.

GLMM approach extends to multivariate normal for multivariate random effect (e.g., useful for multilevel models).

GLMMs are subject-specific models; predictors effects are described at subject level rather than population-averaged as in marginal models.

Since effect $\beta$ is subject-specific, it differs in size from effect in corresponding marginal model.

As correlation increases among observations in cluster (i.e., var(random effects) increases), subject-specific (conditional) effects in GLMMs tend to increase in magnitude relative to population-averaged (marginal) effects.

$$\beta_M \approx \beta_C (1 - \rho)$$
Logistic random-intercept model, showing the *conditional* subject-specific curves and the *marginal* (population-averaged) curve averaging over these.
Given the random effect, GLMM is a GLM. Repeated responses treated as independent; marginally (averaging over subjects), there is positive within-cluster correlation.

GLMM fitting uses marginal ML: Integrate out $u_i$ to obtain marginal likelihood depending on $\beta$ and “variance components.”

E.g., for ordinary random intercept logit model

$$
\text{logit}[P(Y_{it} = 1 \mid u_i)] = u_i + x_{it}^T \beta
$$

likelihood function with $N(0, \sigma^2)$ pdf $f(u_i; \sigma^2)$ for $u_i$ is

$$
\ell(\beta, \sigma^2) = \prod_i \left( \int_{-\infty}^{\infty} \prod_t \left[ \frac{\exp(u_i + x_{it}^T \beta)}{1 + \exp(u_i + x_{it}^T \beta)} \right]^{y_{it}} \right) \left[ \frac{1}{1 + \exp(u_i + x_{it}^T \beta)} \right]^{1-y_{it}} f(u_i; \sigma^2) du_i.
$$
Integrating out random effects can be computationally complex. Numerical integration with Gauss-Hermite quadrature approximates likelihood by finite sum; approximation improves as increase number of *quadrature points*.

Many-dimensional random effects or complex structure requires Monte Carlo methods to approximate likelihood and ML parameter estimates.

Once obtain likelihood function, software uses Newton-Raphson to maximize wrt \((\beta, \sigma)\), obtain *SE* values.

For ML fitting, ordinary likelihood inferences apply.

In testing \(H_0: \sigma = 0\), problems occur from parameter on boundary of parameter space; e.g., LR statistic is approximate equal mixture of chi-squared with \(df = 1\) and degenerate at 0 (when \(\hat{\sigma} = 0\)). For how to handle various cases, see Molenberghs and Verbeke, *American Statistician* 2007.

SAS: PROC NLMIXED uses Gauss-Hermite quadrature for ML fitting of GLMMs, extending PROC MIXED to handle non-normal response and link functions of GLMs.

R: functions include *glmmML* in glmmML library.
**Example:** Support for legalizing abortion in three situations

<table>
<thead>
<tr>
<th>Gender</th>
<th>(1,1,1)</th>
<th>(1,1,0)</th>
<th>(0,1,1)</th>
<th>(0,1,0)</th>
<th>(1,0,1)</th>
<th>(1,0,0)</th>
<th>(0,0,1)</th>
<th>(0,0,0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>342</td>
<td>26</td>
<td>6</td>
<td>21</td>
<td>11</td>
<td>32</td>
<td>19</td>
<td>356</td>
</tr>
<tr>
<td>Female</td>
<td>440</td>
<td>25</td>
<td>14</td>
<td>18</td>
<td>14</td>
<td>47</td>
<td>22</td>
<td>457</td>
</tr>
</tbody>
</table>

Situations are (1) if the family has a very low income and cannot afford any more children, (2) when the woman is not married and does not want to marry the man, and (3) when the woman wants it for any reason. 1, yes; 0, no.

Random effects model

\[
\logit[P(Y_{ij} = 1 \mid u_i)] = \alpha + u_i + \beta_j + \gamma x_i,
\]

where \( x_i = 1 \) for females and 0 for males and \( \{u_i\} \) are independent \( N(0, \sigma^2) \).
R for random effects modeling of abortion opinion data, using initial values for adaptive Gauss-Hermite quadrature.

```r
> abortion <- read.table("abortion.dat",header=TRUE)
> abortion
  gender response question case
  1 1 1 1 1
  2 1 1 2 1
  3 1 1 3 1
...
  5548 0 0 1 1850
  5549 0 0 2 1850
  5550 0 0 3 1850

> z1 <- ifelse(abortion$question==1,1,0)
> z2 <- ifelse(abortion$question==2,1,0)
> library(glmmML)
> fit.glmm <- glmmML(response ~ gender + z1 + z2,
+    cluster=abortion$case, family=binomial, data=abortion,
+    method = "ghq", n.points=70, start.sigma=9)
> summary(fit.glmm)

            coef  se(coef)    z Pr(>|z|)
(Intercept) -0.61874  0.3777 -1.63840  1.01e-01
gender       0.01259  0.4888  0.02575  9.79e-01
z1           0.83470  0.1601  5.21347  1.85e-07
z2           0.29240  0.1567  1.86622  6.20e-02

Scale parameter in mixing distribution: 8.736 gaussian
Std. Error: 0.5421
LR p-value for H_0: sigma = 0: 0
```

A. Agresti (UF)
CDA
March 8-10, 2016 168 / 191
SAS for random effects modeling of abortion data, using PROC NLMIXED with Gauss-Hermite quadrature

data new;
input sex poor single any count;
datalines;
1 1 1 1 342
1 1 1 0 26
1 1 0 1 11
1 1 0 0 32
1 0 1 1 6
1 0 1 0 21
1 0 0 1 19
1 0 0 0 356
2 1 1 1 440
2 1 1 0 25
2 1 0 1 14
2 1 0 0 47
2 0 1 1 14
2 0 1 0 18
2 0 0 1 22
2 0 0 0 457
;

A. Agresti (UF)
data new;
set new;
sex = sex - 1;
subject = _n_;  
q1 = 1; q2 = 0;
resp = poor;
output;
q1 = 0; q2 = 1;
resp = single;
output;
q1 = 0; q2 = 0;
resp = any;
output;
drop poor single any;

proc nlmixed qpoints = 1000;
  parms alpha=-0.7 beta1=.8 beta2=.3 gamma=0 sigma=8.7;
  eta = alpha + beta1*q1 + beta2*q2 + gamma*sex + u;
  p = exp(eta)/(1 + exp(eta));
  model resp ~ binary(p);
  random u ~ normal(0,sigma*sigma) subject = subject;
  replicate count;
run;

Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>-0.6193</td>
<td>0.3783</td>
<td>1849</td>
<td>-1.64</td>
<td>0.1018</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta1</td>
<td>0.8348</td>
<td>0.1601</td>
<td>1849</td>
<td>5.22</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta2</td>
<td>0.2924</td>
<td>0.1567</td>
<td>1849</td>
<td>1.87</td>
<td>0.0621</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>0.01261</td>
<td>0.4897</td>
<td>1849</td>
<td>0.03</td>
<td>0.9795</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sigma</td>
<td>8.7450</td>
<td>0.5390</td>
<td>1849</td>
<td>16.22</td>
<td>&lt;.0001</td>
<td>0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Abortion data: Comparison of estimates with random effects and marginal models

<table>
<thead>
<tr>
<th>Estimate</th>
<th>GLMM</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.835</td>
<td>0.149</td>
</tr>
<tr>
<td>(0.160)</td>
<td>(0.066)</td>
<td></td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.292</td>
<td>0.052</td>
</tr>
<tr>
<td>(0.157)</td>
<td>(0.066)</td>
<td></td>
</tr>
</tbody>
</table>

Note the item effects are much stronger with the random effects model (GLMM) than with the marginal model (GEE), which is not surprising because of the strong correlation (large random effects variability).
Summary

- Multivariate data requires ways of dealing with within-cluster correlation.

- Matched-pairs data: Marginal homogeneity and symmetry, comparing dependent proportions (e.g., McNemar test)

- Marginal modeling of repeated categorical measurement data with covariates, using logits for dependent proportions and quasi-likelihood (GEE) methodology

- Generalized linear mixed models provide a subject-specific approach that is an alternative to marginal models; e.g., random effect for intercept can induce dependence.

- Marginal model often preferred for overall summary of “between-subject” effects (e.g., compare females and males), while GLMM preferred to describe “within-subject” effects and summarize heterogeneity among subjects (random effects $\sigma$).
Final comments

- A shorter, less technical, textbook introduction to this topic is *An Introduction to Categorical Data Analysis*, by A. Agresti (2nd edition, Wiley 2008).

- Also less technical, with R guidance, is M. Kateri’s *Contingency Table Analysis: Methods and Implementation Using R* (2014).

- I hope this short course has been useful to you. Thanks very much for your time and attention!
Exercise 1: The 2014 General Social Survey asked “Human beings, as we know them today, developed from earlier species of animals. True or False?” Is the response to this question associated with one’s political ideology? Let $y =$ opinion about evolution (1 = true, 0 = false). Let $x =$ political ideology (1 = extremely conservative, 2 = conservative, 3 = slightly conservative, 4 = moderate, 5 = slightly liberal, 6 = liberal, 7 = extremely liberal). The table shows four of the 1064 observations. You can copy the complete data file Evolution.dat for $n = 1064$ at www.stat.ufl.edu/~aa/smss/data.

Table: GSS Data on $y =$ Opinion about Evolution (1 = True, 0 = False) and $x =$ Political Ideology (from 1 = extremely conservative to 7 = extremely liberal).

<table>
<thead>
<tr>
<th>Subject</th>
<th>$x$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>
Assigning the category numbers (1, 2, 3, 4, 5, 6, 7) to political ideology, we get the following results for using R to fit the logistic regression model.

```
> summary(glm(y ~ polviews, family=binomial(link="logit"))

                        Estimate  Std. Error   z value Pr(>|z|)
(Intercept)          -1.75658    0.20500   -8.569  <2e-16
polviews             0.49422    0.05092    9.706  <2e-16
```

a. Explain why the estimated probability of believing in evolution increases as political ideology moves in the more liberal direction, with estimated probability equal 0.50 at \( x = 3.55 \).

b. Show the formula for the estimated probability of believing in evolution. Show that for subjects with ideology \( x = 1 \), the most conservative category, the estimated probability equals 0.221, and that for \( x = 7 \), the most liberal category, the estimated probability equals 0.846.
c. Show that at $x = 4$ (moderate ideology), a line drawn tangent to the curve has slope approximately equal to 0.12. So, a 1-category increase in political ideology (i.e., from ‘moderate’ to ‘slightly liberal’) has approximately to a 0.12 increase in the estimated probability of belief in evolution.

d. Show that when political ideology increases by 1 category in the liberal direction, the estimated odds of belief in evolution multiply by 1.64; that is, they increase by 64%.

e. Copy the data from the website and fit the logistic regression model, verifying the results in the table.
Exercise 2: Analyze data from a study to compare two devices used to secure airway in patients undergoing surgery: Data shown on p. 244 of CDA. Response variable: \( Y = \) Sore throat (1 = yes, 0 = no) Explanatory variables: \( D = \) duration of surgery (minutes), \( T = \) type of airway (1 = tracheal tube, 0 = laryngeal mask)

<table>
<thead>
<tr>
<th>Duration</th>
<th>Type</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The entire data set can be copied from data file (8) at www.stat.ufl.edu/~aa/cda/data.html

- Fit a main effects model, and interpret effects.
- Show how to conduct inference about the type of airway effect.
- Add an interaction term, and interpret effects for this model.
- Analyze whether the model permitting interaction gives a significantly better fit than the main effects model.
Exercise 3: In several places beginning in Sec. 4.3.2, the text *Categorical Data Analysis* analyzes data from a study of female horseshoe crabs on an island in the Gulf of Mexico. During spawning season, the females migrate to a shore to breed, with a male attached to her posterior spine, and she burrows into the sand and lays clusters of eggs. During spawning, other male crabs may group around the pair and may also fertilize the eggs. These male crabs that cluster around the female crab are called *satellites*. Here, we analyze whether a female crab had any satellites (1 = yes, 0 = no) using as predictors that crab’s width of shell (in centimeters) and color (categories medium light, medium, medium dark, dark). The entire data set can be seen at www.stat.ufl.edu/~aa/cda/data.html

The next page shows some output from model fitting, in which $c_1$, $c_2$, and $c_3$ are indicators for the first three colors:
Software Output (based on SAS) for Model with Width and Color Predictors of Whether Horseshoe Crab Has Satellites

<table>
<thead>
<tr>
<th>Criteria For Assessing Goodness Of Fit</th>
<th>DF</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>168</td>
<td>187.4570</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>168</td>
<td>168.6590</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>-93.7285</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Likelihood-Ratio Confidence Limits</th>
<th>Chi-Square</th>
<th>Pr&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-12.7151</td>
<td>2.7618</td>
<td>-18.4564 -7.5788</td>
<td>21.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c1</td>
<td>1.3299</td>
<td>0.8525</td>
<td>-0.2738 3.1354</td>
<td>2.43</td>
<td>0.1188</td>
</tr>
<tr>
<td>c2</td>
<td>1.4023</td>
<td>0.5484</td>
<td>0.3527 2.5260</td>
<td>6.54</td>
<td>0.0106</td>
</tr>
<tr>
<td>c3</td>
<td>1.1061</td>
<td>0.5921</td>
<td>-0.0279 2.3138</td>
<td>3.49</td>
<td>0.0617</td>
</tr>
<tr>
<td>width</td>
<td>0.4680</td>
<td>0.1055</td>
<td>0.2713 0.6870</td>
<td>19.66</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
a. Write down the prediction equation, and show how to find the estimated probability of a satellite for a crab of mean width (26.3 cm) that is (i) dark, (ii) medium light.

b. Interpret the width effect.

c. Show how to conduct inference (test and confidence interval) about the width effect. Interpret.

Exercise 4: For the horseshoe crab data mentioned in the previous exercise, download the data from the text website and using all the explanatory variables given there, select a model for the probability that a female horseshoe crab has at least one satellite. (You need to set up an indicator function for the response by letting $y = 1$ if the count of satellites is positive and $y = 0$ if there are no satellites.) Interpret the fit of the model.
Exercise 5: For the high school survey data analyzed in the short course section on loglinear models, fit the model denoted in the notes by \((AC, CM)\). Present the results of (a) the deviance goodness-of-fit test, (b) interpreting the estimated \(CM\) association, (c) interpreting results of a test for the \(CM\) association, (d) interpreting results of a 95% confidence interval for the \(CM\) conditional association.

Exercise 6: In the previous exercise, if you regard \(M\) as a response variable, what logit model is equivalent to this loglinear model? Fit that model, and show how results are equivalent.
Exercise 7: Consider the mental impairment data at www.stat.ufl.edu/~aa/ordinal/data.html in the first file shown. Here, SES is a binary indicator of socioeconomic status (1 = high, 0 = low), and life events is a composite measure of the number and severity of important life events that occurred to the subject within the past three years (such as birth of a child, a new job, a divorce, a death in the family).

a. Using your choice of software, fit the cumulative logit model with main effects, and interpret.

b. Conduct a likelihood-ratio or Wald test about the life events effect, and interpret.

c. Construct a 95% confidence interval for a cumulative odds ratio to interpret the life events effect.

d. Now fit the more general model that allows interaction between life events and SES in their effects on mental impairment. Interpret the nature of the interaction. Test whether the interaction effect is needed in the model.

Exercise 8: Fit the random effects model to the abortion opinion data, data file (22) at www.stat.ufl.edu/~aa/cda/data.html, using a variety of initial values and number of quadrature points. Summarize what you learn.
Exercise 1

a. Since $\hat{\beta} = 0.494 > 0$, the estimated probability of believing in evolution increases as political ideology moves in the more liberal direction (i.e., higher $x$ scores). The estimated probability equals 0.50 at $x = -\hat{\alpha}/\hat{\beta} = 1.757/0.494 = 3.55$.

b. From the estimates in the table, a person with political ideology $x$ has estimated probability of believing in evolution

$$\hat{P}(y = 1) = \frac{e^{-1.757+0.494x}}{1 + e^{-1.757+0.494x}}.$$  

For subjects with ideology $x = 1$, the estimated probability equals

$$\hat{P}(y = 1) = \frac{e^{-1.757+0.494(1)}}{1 + e^{-1.757+0.494(1)}} = \frac{e^{-1.262}}{1 + e^{-1.262}} = \frac{0.283}{1.283} = 0.221.$$  

C. A line drawn tangent to the logistic curve at $x = 4$ has slope approximately equal to $\hat{\beta}/4 = 0.494/4 = 0.12$.  

d. The antilog of $\hat{\beta}$ is $e^{\hat{\beta}} = e^{0.494} = 1.64$. When political ideology increases by 1 category in the liberal direction, the estimated odds of belief in evolution multiply by 1.64; that is, they increase by 64%. When $x = 5$, for example, the estimated odds of belief in evolution are 1.64 times what they are when $x = 4$. When $x = 4$,

$$\text{Estimated odds} = \frac{\hat{P}(y = 1)}{1 - \hat{P}(y = 1)} = e^{-1.757+0.494(4)} = 1.246,$$

whereas when $x = 5$,

$$\text{Estimated odds} = \frac{\hat{P}(y = 1)}{1 - \hat{P}(y = 1)} = e^{-1.757+0.494(5)} = 2.043,$$

which is 1.64 times the value of 1.246 at $x = 4$. In other words, $e^{\hat{\beta}} = e^{0.494} = 1.64 = 2.043/1.246$ is an estimated odds ratio, equaling the estimated odds at $x = 5$ divided by the estimated odds at $x = 4$. 
Exercise 2

a, c

> fit1 <- glm(Y ~ D + T, family=binomial)
> summary(fit1)
Coefficients:

Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.41734 1.09457 -1.295 0.19536
D 0.06868 0.02641 2.600 0.00931 **
T -1.65895 0.92285 -1.798 0.07224 .

---
Null deviance: 46.180 on 34 degrees of freedom
Residual deviance: 30.138 on 32 degrees of freedom
AIC: 36.138

> glm(Y ~ D, family=binomial)
Coefficients:

(Intercept) D

-2.21358 0.07038

Residual Deviance: 33.65 AIC: 37.65

> fit3 <- glm(Y ~ D + T + D*T, family=binomial)
> summary(fit3)
Coefficients:

Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.04979 1.46940 0.034 0.9730
D 0.02848 0.03429 0.831 0.4062
T -4.47224 2.46707 -1.813 0.0699 .
D:T 0.07460 0.05777 1.291 0.1966

---
Residual deviance: 28.321 on 31 degrees of freedom
AIC: 36.321

A. Agresti (UF)  CDA  March 8-10, 2016
b. Likelihood-ratio statistic = difference of deviances = 
33.65 − 30.14 = 3.51, \( df = 1 \), \( P \)-value = 0.06. Or, can construct CI for odds ratio relating type of airway to response, given duration, such as the Wald CI \( \exp[-1.65895 \pm 1.96(0.92285)] \).

c. Note duration effect (change in logit for each minute increase in duration) is 0.028 for laryngeal mask and \( (0.028 + 0.075) = 0.103 \) for tracheal tube.

d. Likelihood-ratio statistic = difference of deviances = 
30.14 − 28.32 = 1.82, \( df = 1 \), \( P \)-value = 0.18.
Exercise 3:

a. \( \text{logit}[\hat{P}(Y = 1)] = -12.715 + 1.33c_1 + 1.40c_2 + 1.11c_3 + 0.468x. \)

For dark crab of mean width, \( \hat{P}(Y = 1) = e^{-12.715+0.468(26.3)}/[1 + e^{-12.715+0.468(26.3)}] = 0.40. \)

b. For a given color, the estimated odds of a satellite multiply by \( e^{0.468} = 1.60 \) for each 1-cm increase in width.

c. Wald chi-squared = 19.66, \( df = 1, P\)-value < 0.0001, so extremely strong evidence of a positive effect of width, adjusting for color. The confidence interval for the multiplicative effect of a 1-cm increase in width on the odds of a satellite (for a given color), based on inverting the likelihood ratio test, is \( (e^{0.2713}, e^{0.6870}) = (1.31, 1.99)) \).

Exercise 4: To find discussion of model fitting for these data, look at Sec. 6.1 of CDA. Good models include those with color and width predictors, where color can be treated either as qualitative, or ordinal with equally-spaced scores, or merely in terms of whether a crab is dark. Weight can be used in place of width, but they are not both needed, because of the strong positive correlation between them (0.887).
Exercise 5:
As seen below, deviance $G^2 = 92.018$ with $df = 2$ suggests a poor fit. The estimated conditional $CM$ odds ratio is $e^{3.224} = 25.1$, with Wald CI $e^{3.224 \pm 1.96(0.161)} = (18.3, 34.5)$. The change in deviance for testing the $CM$ association is $843.83 - 92.018 = 751.81$, with $df = 1$, extremely strong evidence against conditional independence of $C$ and $M$.

\[
\begin{array}{cccc}
\text{a} & \text{c} & \text{m} & \text{count} \\
1 & yes & yes & yes & 911 \\
2 & yes & yes & no & 538 \\
3 & yes & no & yes & 44 \\
4 & yes & no & no & 456 \\
5 & no & yes & yes & 3 \\
6 & no & yes & no & 43 \\
7 & no & no & yes & 2 \\
8 & no & no & no & 279 \\
\text{fit} & \text{<- glm(count~a+c+m+a*c+c*m, family=poisson)} \\
\text{fit2} & \text{<- glm(count~a+c+m+a*c, family=poisson)} \\
\end{array}
\]
Exercise 6:
It is equivalent to the logit model for four binomials (rather than 8 Poissons) in which case there is only a \( C \) main effect (and not an \( A \) effect) in predicting the \( M \) response. In the following output, the first logit model is equivalent to loglinear model \((AC, AM, CM)\) and the second is equivalent to loglinear \((AC, CM)\).

---

```r
> data
  A  C Myes Mno
1 1 1  911 538
2 1 0   44 456
3 0 1   3  43
4 0 0    2  279
> n <- Myes + Mno
> fit <- glm(Myes/n ~ A + C, family=(binomial), weights=n)
> summary(fit)
Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)              -5.3090   0.4752  -11.172  < 2e-16 ***
A                        2.9860   0.4647   6.426  1.31e-10 ***
C                        2.8479   0.1638  17.382  < 2e-16 ***
---
Residual deviance: 0.37399 on 1 degrees of freedom

> fit2 <- glm(Myes/n ~ C, family=(binomial), weights=n)
> summary(fit2)
Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)              -2.7710   0.1522  -18.233  < 2e-16 ***
C                        3.2242   0.1609   20.034  < 2e-16 ***
---
Residual deviance: 92.018 on 2 degrees of freedom
```
Exercise 7
The Wald test about life events has chi-squared statistic $= (-0.3189/0.1194)^2 = 7.13$, $df = 1$, and $P$-value $= 0.008$. The 95% Wald confidence interval for the cumulative odds ratio is $e^{-0.3189\pm1.96(0.1194)}$, or $(0.58, 0.92)$.

---

```r
> mental <- read.table("mental.dat", header=TRUE)
> mental
   impair ses life
 1      1   1   1
 2      1   1   9
 3      1   4   0
 4      5   0   9
 5      3   0   9
 6      5   0   9
 7      3   0   9
 8      5   0   9
 9      3   0   9
10     5   0   9
11     3   0   9
12     5   0   9
13     3   0   9
14     5   0   9
15     3   0   9
16     5   0   9
17     3   0   9
18     5   0   9
19     3   0   9
20     5   0   9
21     3   0   9
22     5   0   9
23     3   0   9
24     5   0   9
25     3   0   9
26     5   0   9
27     3   0   9
28     5   0   9
29     3   0   9
30     5   0   9
31     3   0   9
32     5   0   9
33     3   0   9
34     5   0   9
35     3   0   9
36     5   0   9
37     3   0   9
38     5   0   9
39     3   0   9
40     5   0   9

> install.packages("VGAM")
> 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
```
The model permitting interaction shows that the effect of life events is greater for those at low SES, but not significantly so.

```r
> fit2 <- vglm(impair~ses+life+ses*life,family=cumulative(parallel=TRUE),data=mental)
> summary(fit2)
Coefficients:
               Estimate Std. Error  z value
(Intercept):1  0.098131   0.81107  0.12099
(Intercept):2  1.592521   0.83729  1.90199
(Intercept):3  2.606616   0.90980  2.86504
ses            0.370876   1.13027  0.32813
life           -0.420448   0.19034 -2.20893
ses:life       -0.420448   0.23613 -1.76777
Residual deviance: 98.50444 on 114 degrees of freedom
```