ST 544: Applied Categorical Data Analysis

Daowen Zhang

zhang@stat.ncsu.edu
http://www4.stat.ncsu.edu/~dzhang2
## Contents

1. **Introduction** 3
2. **Contingency Tables** 40
3. **Generalized Linear Models (GLMs)** 122
4. **Logistic Regression** 189
5. **Building and Applying Logistic Regression Models** 248
6. **Multicategory Logit Models** 299
8. **Models for Matched Pairs** 366
9. **Modeling Correlated, Clustered, Longitudinal Categorical Data** 435
10. **Random Effects: Generalized Linear Mixed Models (GLMMs)** 480
1 Introduction

I. Categorical Data

Definition

- A categorical variable is a (random) variable that can only take finite or countably many values (categories).

- Type of categorical variables:
  - Gender: F/M or 0/1; Race: White, Black, Others – *Nominal*
  - Patient’s Health Status: Excellent, Good, Fair, Bad – *Ordinal*
  - # of car accidents in next Jan in Wake County – *Interval*
• Application of math operations:

<table>
<thead>
<tr>
<th>Type</th>
<th>Nominal</th>
<th>Ordinal</th>
<th>Interval</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>Gender, Race</td>
<td>Patient’s Health Status</td>
<td># of car accidents</td>
<td>Height</td>
</tr>
<tr>
<td>Math Operation</td>
<td>None</td>
<td>&gt;, &lt;</td>
<td>&gt;, &lt;, ±</td>
<td>Any</td>
</tr>
</tbody>
</table>

• **Response (Dependent) Variable:** $Y$
  
  **Explanatory (Independent, Covariate) Variable:** $X$.

• We focus on the cases where $Y$ is categorical.
II. Common Distributions

II.1 Binomial distribution

- We have a Bernoulli process:
  1. \( n \) independent trials, \( n > 0 \) – fixed integer
  2. Each trial produces 1 of 2 outcomes: \( S \) for success & \( F \) for failure
  3. Success probability at each trial is the same (\( \pi \in (0, 1) \))

- \( Y = \) total \# of successes out of \( n \) trials, \( Y \sim \text{Bin}(n, \pi) \) and has a probability mass function (pmf):

\[
p(y) = P[Y = y] = \frac{n!}{y!(n-y)!} \pi^y (1 - \pi)^{n-y}, \quad y = 0, 1, 2, ..., n.
\]

\( \frac{n!}{y!(n-y)!} \) is usually denoted as \( \binom{n}{y} \), and usually is \( nCr \) in your calculator.

- The above pmf is useful in calculating probabilities associated with a binomial distribution (for a known \( \pi \)).
Table 1.1. Binomial Distribution with $n = 10$ and $\pi = 0.20$, 0.50, and 0.80. The Distribution is Symmetric when $\pi = 0.50$

<table>
<thead>
<tr>
<th>$y$</th>
<th>$P(y)$ when $\pi = 0.20$</th>
<th>$P(y)$ when $\pi = 0.50$</th>
<th>$P(y)$ when $\pi = 0.80$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.107</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0.268</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.302</td>
<td>0.044</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.201</td>
<td>0.117</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.088</td>
<td>0.205</td>
<td>0.005</td>
</tr>
<tr>
<td>5</td>
<td>0.027</td>
<td>0.246</td>
<td>0.027</td>
</tr>
<tr>
<td>6</td>
<td>0.005</td>
<td>0.205</td>
<td>0.088</td>
</tr>
<tr>
<td>7</td>
<td>0.001</td>
<td>0.117</td>
<td>0.201</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td>0.044</td>
<td>0.302</td>
</tr>
<tr>
<td>9</td>
<td>0.000</td>
<td>0.010</td>
<td>0.268</td>
</tr>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.001</td>
<td>0.107</td>
</tr>
</tbody>
</table>
• Examples: Suppose two people (A and B) are to play \( n = 10 \) chess games with no tie. If we assume that the games are independent to each other and \( \pi = P[A \text{ wins } B \text{ in a single game}] = 0.6 \).

1. Find the prob that A wins 4 games.

\[
P[Y = 4] = \binom{10}{4} 0.6^4 (1 - 0.6)^{10-4} = 0.1115
\]

2. Find the prob that A wins at least 4 games.

\[
P[Y \geq 4] = 1 - P[Y \leq 3] = 1 - 0.0548 = 0.9452.
\]

3. Find the prob that B wins more than A.

\[
\]
• Properties of a binomial distribution $Y \sim \text{Bin}(n, \pi)$:

1. $Y = Y_1 + Y_2 + \cdots + Y_n$, where $Y_i = 1/0$ is the number of success in the $i$th trial, $Y_i$ indep of $Y_j$ for $i \neq j$.

2. Mean, variance and standard deviation of $Y$:

   \[
   \begin{align*}
   \mathbb{E}(Y) &= n\pi \\
   \text{var}(Y) &= n\pi(1 - \pi) \\
   \sigma &= \sqrt{\text{var}(Y)} = \sqrt{n\pi(1 - \pi)}
   \end{align*}
   \]

3. $Y$ has smaller variation when $\pi$ is closer to 0 or 1.

• When $n$ is large, $\text{Bin}(n, \pi)$ can be well approximated by a normal dist.  
Requirement: $n\pi \geq 5$ & $n(1 - \pi) \geq 5$.  


Normal Approximation to Bin(12, 0.5)
II.2 Multinomial distribution (for nominal or ordinal categorical variables)

\[ \begin{array}{c|cccc} Y & 1 & 2 & \cdots & c \\ \hline \text{Prob} & \pi_1 & \pi_2 & \cdots & \pi_c \end{array} \]

where \( \pi_i = P[Y = j] > 0, \sum_{j=1}^{c} \pi_j = 1. \)

- Each trial of \( n \) trials results in an outcome in one (and only one) of \( c \) categories, represented by
  \[
  \mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{ic} \end{bmatrix}, \quad i = 1, 2, \ldots, n. \]
  For example, \( \mathbf{Y}_i = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}. \)

Only one of \( \{Y_{ij}\}_{j=1}^{c} \) is 1, others are 0; \( \pi_j = P[Y_{ij} = 1]. \)

- Prob of observing \( \mathbf{Y}_i \): \( \pi_1^{Y_{i1}} \pi_2^{Y_{i2}} \cdots \pi_c^{Y_{ic}} \)
• Often time, we may not have the individual outcome. Instead, we have the following summary:

\[ \tilde{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_c \end{bmatrix}, \]

where \( n_j \) is the \# of trials resulting outcome in the \( j \) category. That is \( n_j = \sum_{i=1}^{n} Y_{ij} \).

• The probability of observing \( \tilde{n} \) is

\[ p(n_1, n_2, \ldots, n_c) = \frac{n!}{n_1!n_2!\cdots n_c!} \pi_1^{n_1} \pi_2^{n_2} \cdots \pi_c^{n_c}. \]

• We often denote \( \tilde{n} \sim \text{multinomial}(n, (\pi_1, \ldots, \pi_c)) \).
• In practice, we want to keep the data in the original form of $Y_i$, or the category the $i$th observation fell, together with other covariate information if such information is available. This is especially the case if each $i$ represents a subject and we would like to use the covariate information to predict which category the individual $i$ most likely falls (regression setting).
• Properties of a multinomial distribution:

1. \( n_j \sim \text{Bin}(n, \pi_j) \Rightarrow \)

\[
E(n_j) = n\pi_j, \quad \text{var}(n_j) = n\pi_j(1 - \pi_j).
\]

2. \( n_i \) and \( n_j \ (i \neq j) \) are negatively associated:

\[
cov(n_i, n_j) = -n\pi_i\pi_j, \ i \neq j.
\]

• \( \tilde{n} \) can be written:

\[
\tilde{n} = \begin{bmatrix}
n_1 \\
n_2 \\
\vdots \\
n_c
\end{bmatrix} = \sum_{i=1}^{n} \tilde{Y}_i.
\]

By CLT, \( \tilde{n} \) approximately has a (multivariate) normal distribution when \( n \) is large.
III. Large-Sample Inference on $\pi$ in a Binomial Distribution

III.1 Likelihood function and maximum likelihood estimation (MLE)

- The parameter $\pi$ in $\text{Bin}(n, \pi)$ is usually unknown and we would like to learn about $\pi$ based on data $y$ from $\text{Bin}(n, \pi)$.

- An intuitive estimate of $\pi$ is the sample proportion

$$p = \frac{y}{n} = \frac{y_1 + y_2 + \ldots + y_n}{n}.$$

1. $p$ is an unbiased estimator (as a random variable):

$$\mathbb{E}(p) = \pi.$$

2. $p$ has a better accuracy when $n$ gets larger:

$$\text{var}(p) = \frac{\pi(1 - \pi)}{n}.$$

3. When $n$ is large, $p$ has an approximate normal distribution (sampling distribution)
Sample proportion \( p \) is the MLE of \( \pi \):

1. Given data \( y \sim \text{Bin}(n, \pi) \), we exchange the roles of \( y \) and \( \pi \) in the pmf and treat it as a function of \( \pi \):

\[
L(\pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}.
\]

This function is called the likelihood function of \( \pi \) for given data \( y \).

2. For example, if \( y = 6 \) out of \( n = 10 \) Bernoulli trials, the likelihood function of \( \pi \) is

\[
L(\pi) = \binom{10}{6} \pi^6 (1 - \pi)^{10-6} = 210 \pi^6 (1 - \pi)^4.
\]

3. Intuitively, the best estimate of \( \pi \) would be the one that maximizes this likelihood or the log-likelihood:

\[
\ell(\pi) = \text{const} + y \log(\pi) + (n - y) \log(1 - \pi).
\]

Note that we use natural log here.

4. It can be shown that the MLE \( \hat{\pi} \) of \( \pi \) is \( p = y/n \).
Figure 1.1. Binomial likelihood functions for \( y = 0 \) successes and for \( y = 6 \) successes in \( n = 10 \) trials.
• In general, the MLE of a parameter has many good statistical properties:

1. When sample size $n$ is large, an MLE is unbiased.
2. When sample size $n$ is large, the variance of an MLE $\to 0$.
3. When sample size $n$ is large, an MLE has an approximate normal distribution.
4. Under some conditions, the MLE is the most efficient estimator.

• We will use ML method most of time in this course.
III.2 Significance test on $\pi$

- Test $H_0: \pi = \pi_0$ v.s. $H_a: \pi \neq \pi_0$ based on data $y \sim \text{Bin}(n, \pi)$.
- The MLE $\hat{\pi} = p = y/n$ has properties:
  \[ E(p) = \pi, \quad \sigma(p) = \sqrt{\pi(1 - \pi)/n} \] (standard error).

- Three classical tests:
  1. Wald test (less reliable):
     \[ Z = \frac{p - \pi_0}{\sqrt{p(1 - p)/n}}, \quad \text{or} \quad Z^2 = \left( \frac{p - \pi_0}{\sqrt{p(1 - p)/n}} \right)^2. \]

     Compare $Z$ to $N(0, 1)$, or compare $Z^2$ to $\chi^2_1$ if $n$ is large.

     That is, if $|Z| \geq z_{\alpha/2}$ or $Z^2 \geq \chi^2_{1, \alpha}$, then we reject $H_0$ at the significance level $\alpha$.

     Large-sample p-value = $2P[Z \geq |z|] = P[\chi^2_1 \geq z^2]$. 

2. Score test (more reliable):

\[ Z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}/n}, \quad \text{or} \quad Z^2 = \left( \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \right)^2. \]

Compare \( Z \) to \( N(0, 1) \), or compare \( Z^2 \) to \( \chi^2_1 \) if \( n \) is large.

That is, if \(|Z| \geq z_{\alpha/2}\) or \(Z^2 \geq \chi^2_{1,\alpha}\), then we reject \( H_0 \) at the significance level \( \alpha \).

Large-sample p-value = \( 2P[Z \geq |z|] = P[\chi^2_1 \geq z^2] \).
3. Likelihood ratio test (LRT):

\[ \ell_0 = y \log \pi_0 + (n - y) \log (1 - \pi_0) \]
\[ \ell_1 = y \log p + (n - y) \log (1 - p) \]
\[ G^2 = 2(\ell_1 - \ell_0) \]
\[ = 2 \left[ y(\log p - \log \pi_0) + (n - y)\{\log (1 - p) - \log (1 - \pi_0)\} \right] \]
\[ = 2 \left[ y \log \frac{p}{\pi_0} + (n - y) \log \frac{1-p}{1-\pi_0} \right] \]
\[ = 2 \left[ y \log \frac{np}{n\pi_0} + (n - y) \log \frac{n(1-p)}{n(1-\pi_0)} \right] \]
\[ = 2 \left[ y \log \frac{y}{n\pi_0} + (n - y) \log \frac{n-y}{n-n\pi_0} \right] \]
\[ = 2 \sum_{\text{obs.}} \log \frac{\text{obs.}}{\text{exp.}} \text{ cells} \]
Compare $G^2$ to $\chi^2_1$.

That is, if $G^2 \geq \chi^2_{1,\alpha}$, then we reject $H_0$ at the significance level $\alpha$.

Large-sample p-value = $P[\chi^2_1 \geq G^2]$. 
• Example: In 2002 GSS, 400 out of 893 responded yes to “...for a pregnant woman to obtain a legal abortion if ...”

• Test $H_0 : \pi = 0.5$ v.s. $H_a : \pi \neq 0.5$ at significance level 0.05.

• $p = y/n = 400/893 = 0.448$.

  1. Wald test:

$$z = \frac{p - \pi_0}{\sqrt{p(1-p)/n}} = \frac{0.448 - 0.5}{\sqrt{0.448 \times (1 - 0.448)/893}} = -3.12.$$  

  Since $z < -1.96$, reject $H_0$ at 0.05 significance level.

  Large sample p-value $= 2P[Z \geq | -3.12|] = 0.0018$. 
2. Score test:

\[ z = \frac{p - \pi_0}{\sqrt{\pi_0 (1 - \pi_0) / n}} = \frac{0.448 - 0.5}{\sqrt{0.5 \times (1 - 0.5) / 893}} = -3.11. \]

Since \( z < -1.96 \), reject \( H_0 \) at 0.05 significance level.

Large sample p-value = \( 2P[Z \geq | -3.11|] = 0.0019. \)
3. LRT:

\[ G^2 = 2 \sum_{2 \text{ cells}} \text{obs.} \log \frac{\text{obs.}}{\text{exp.}} \]

\[ = 2 \times \log \left\{ \frac{400}{893 \times 0.5} \right\} \]

\[ + (893 - 400) \times \log \left\{ \frac{(893 - 400)}{893 - 893 \times 0.5} \right\} \]

\[ = 9.7 > 1.96^2 = 3.84, \]

\[ \Rightarrow \text{Reject } H_0 \text{ at 0.05 significance level.} \]

Large sample p-value = \( P[\chi^2_1 \geq 9.7] = 0.0018. \)

- **Note:** These three tests can be extended to test other parameters.
III.C Large-Sample Confidence Interval (CI) for $\pi$

- Wald CI of $\pi$: For given confidence level $1 - \alpha$, solve the following inequality for $\pi_0$

$$\left| \frac{p - \pi_0}{\sqrt{p(1 - p)/n}} \right| \leq z_{\alpha/2}$$

$$\Rightarrow [p - z_{\alpha/2} \sqrt{p(1 - p)/n}, p + z_{\alpha/2} \sqrt{p(1 - p)/n}].$$

**Note:** $\sqrt{p(1 - p)/n}$ is called the *estimated standard error (SE)* of $p$.

The Wald CI has the form: $Est. \pm z_{\alpha/2} SE$.

For the 2002 GSS example, a 95% Wald CI for $\pi$ is:

$$[0.448 - 1.96 \sqrt{0.448(1 - 0.448)/893},$$

$$0.448 + 1.96 \sqrt{0.448(1 - 0.448)/893}]$$

$$= [0.415, 0.481]$$

---

Slide 25
Note. The Wald CI is not very reliable for small $n$ and $p \approx 0$ or 1. Remedy for 95% CI: add 2 successes and 2 failures to the data and re-construct the 95% Wald CI.

For example, $y = 2, n = 10$, 95% Wald CI:

$$[0.2 - 1.96 \times \sqrt{0.2 \times 0.8/10}, 0.2 + 1.96 \times \sqrt{0.2 \times 0.8/10}] = [-0.048, 0.448].$$

With the remedy, $y^* = 4$, $n^* = 14$, $p^* = 4/14 = 0.286$, 95% Wald CI is

$$[0.286 - 1.96 \times \sqrt{0.286 \times 0.714/14}, 0.286 + 1.96 \times \sqrt{0.286 \times 0.714/14} = [0.049, 0.523].$$
• Score CI of \( \pi \): For given confidence level \( 1 - \alpha \), solve the following inequality for \( \pi_0 \)

\[
\left| \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} \right| \leq z_{\alpha/2}
\]

For the 2002 GSS example, a 95% score CI solves

\[
\left| \frac{0.448 - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/893}} \right| \leq 1.96
\]

\[ \Rightarrow [0.416, 0.481]. \]

**Note:** Here the sample size \( n \) is very large, the Wald CI and the score CI are very close.
Absolute values of the score statistic as a function of $\pi_0$
- Likelihood ratio CI: For given confidence level $1 - \alpha$, solve for $\pi_0$:

$$2 \left[ y \log \left\{ \frac{y}{n\pi_0} \right\} + (n - y) \log \left\{ \frac{(n - y)}{n - n\pi_0} \right\} \right] \leq z_{\alpha/2}^2.$$ 

- For the 2002 GSS example, a 95% LR CI solves:

$$2 \left[ 400 \log \left\{ \frac{400}{893\pi_0} \right\} + (893 - 400) \log \left\{ \frac{(893 - 400)}{893 - 893\pi_0} \right\} \right] \leq 1.96^2$$

$\Rightarrow [0.415, 0.481].$
LRT statistic as a function of $\pi_0$
Note: We see from the GSS example that, for large sample size $n$, the Wald, score, LR CIs are all very close. However, if $n$ is not large, there will be some discrepancy among them.

For example, if $y = 9$, $n = 10$, then:

1. Wald CI: $[0.714, 1.086] = [0.714, 1]$
2. Score CI: $[0.596, 0.982]$
3. LR CI: $[0.628, 0.994]$
IV. Other Inference Approaches

IV.1 Small-sample inference for $\pi$ in $\text{Bin}(n, \pi)$

1. One-sided test: $H_0 : \pi = \pi_0$ v.s. $H_a : \pi > \pi_0$.

Given data $y \sim \text{Bin}(n, \pi)$, the testing procedure would be: Reject $H_0$ if $y$ is large.

Exact p-value = $P[Y \geq y|H_0]$.

For example, $H_0 : \pi = 0.5$ v.s. $H_a : \pi > 0.5$, and $y = 6, n = 10$. Then exact p-value = $P[Y \geq 6|\pi = 0.5] = 0.377$. 
2. Two-sided test: $H_0 : \pi = \pi_0 \ \text{v.s.} \ \ H_a : \pi \neq \pi_0$.

Given data $y \sim \text{Bin}(n, \pi)$, the testing procedure would be: *Reject $H_0$ if $|y - n\pi_0|$ is large.*

Exact p-value = $P[|Y - n\pi_0| \geq |y - n\pi_0||H_0]$.

For example, $H_0 : \pi = 0.5 \ \text{v.s.} \ \ H_a : \pi \neq 0.5$, and $y = 6, n = 10$. Then

\[
\text{exact p-value} = P[|Y - 10 \times 0.5| \geq |6 - 10 \times 0.5||H_0] \\
= P[|Y - 5| \geq 1|H_0] \\
= P[Y - 5 \geq 1|H_0] + P[Y - 5 \leq -1|H_0] \\
= P[Y \geq 6|H_0] + P[Y \leq 4|H_0] \\
= 0.377 + 0.377 = 0.754.
\]

Using exact p-value can be conservative!
Table 1.2. Null Binomial Distribution and One-Sided
$P$-values for Testing $H_0: \pi = 0.50$ against $H_a: \pi > 0.50$
with $n = 10$

<table>
<thead>
<tr>
<th>$y$</th>
<th>$P(y)$</th>
<th>$P$-value</th>
<th>Mid $P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.001</td>
<td>1.000</td>
<td>0.9995</td>
</tr>
<tr>
<td>1</td>
<td>0.010</td>
<td>0.999</td>
<td>0.994</td>
</tr>
<tr>
<td>2</td>
<td>0.044</td>
<td>0.989</td>
<td>0.967</td>
</tr>
<tr>
<td>3</td>
<td>0.117</td>
<td>0.945</td>
<td>0.887</td>
</tr>
<tr>
<td>4</td>
<td>0.205</td>
<td>0.828</td>
<td>0.726</td>
</tr>
<tr>
<td>5</td>
<td>0.246</td>
<td>0.623</td>
<td>0.500</td>
</tr>
<tr>
<td>6</td>
<td>0.205</td>
<td>0.377</td>
<td>0.274</td>
</tr>
<tr>
<td>7</td>
<td>0.117</td>
<td>0.172</td>
<td>0.113</td>
</tr>
<tr>
<td>8</td>
<td>0.044</td>
<td>0.055</td>
<td>0.033</td>
</tr>
<tr>
<td>9</td>
<td>0.010</td>
<td>0.011</td>
<td>0.006</td>
</tr>
<tr>
<td>10</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0005</td>
</tr>
</tbody>
</table>
• Using exact p-value is conservative!

For example, if we are testing $H_0 : \pi = 0.5$ v.s. $H_a : \pi > 0.5$ and our significance level $\alpha = 0.05$ using data $y$ from $\text{Bin}(n = 10, \pi)$. Then based on Table 1.2, we should reject $H_0$ only if $y = 9$ or $y = 10$. However, the actual type I error probability is $0.011 < \alpha = 0.05$. Conservative!
IV.2 Inference based on the mid p-value

- For testing $H_0 : \pi = 0.5 \text{ v.s. } H_a : \pi > 0.5$ with data $y$ from $\text{Bin}(n, \pi)$, we calculate the

$$\text{mid p-value} = 0.5 P[Y = y|H_0] + [Y = y + 1|H_0] + \cdots [Y = n|H_0].$$

For example, suppose $y = 9, n = 10$, then

$$\text{mid p-value} = 0.5 P[Y = 9|H_0] + [Y = 10|H_0] = 0.006.$$

With the use of mid p-value, we will reject $H_0 : \pi = 0.5$ in favor of $H_a : \pi > 0.5$ if $y = 8, 9, 10$. The actual type I error probability is 0.055, much closer to the significance level $\alpha = 0.05$. 
IV.3 Exact confidence interval for $\pi$ using exact p-value

- For given confidence level $(1 - \alpha)$ and observed $y \sim \text{Bin}(n, \pi)$, solve

$$P_{\pi}[Y \geq y] = \sum_{i=y}^{n} \binom{n}{i} \pi^i (1 - \pi)^{n-i} = \alpha/2$$

to get lower limit $\hat{\pi}_L$; if $y = 0$, then set $\hat{\pi}_L = 0$.

Solve

$$P_{\pi}[Y \leq y] = \sum_{i=0}^{y} \binom{n}{i} \pi^i (1 - \pi)^{n-i} = \alpha/2$$

to get upper limit $\hat{\pi}_U$; if $y = n$, then set $\hat{\pi}_U = 1$.

$\Rightarrow [\hat{\pi}_L, \hat{\pi}_U]$ is an exact $(1 - \alpha)$ for $\pi$.

- For example, $y = 3$, $n = 10$, an exact 95% CI is [0.07, 0.65]. That is,

$$P_{\pi=0.07}[Y \geq 3] = 0.025, \quad P_{\pi=0.65}[Y \leq 3] = 0.025.$$  

This exact CI is conservative, that is, too wide.
$P[Y \geq 3|\pi] \text{ (—) and } P[Y \leq 3|\pi] \text{ (…) as functions of } \pi$
IV.4 Exact confidence interval for $\pi$ using exact mid p-value

- For given confidence level $(1 - \alpha)$ and observed $y \sim \text{Bin}(n, \pi)$, solve
  \[
  \frac{1}{2} P_\pi[Y = y] + P_\pi[Y > y] = \alpha/2
  \]
  to get lower limit $\hat{\pi}_L$; if $y = 0$, then $\hat{\pi}_L = 0$.

  Solve
  \[
  \frac{1}{2} P_\pi[Y = y] + P_\pi[Y < y] = \alpha/2
  \]
  to get upper limit $\hat{\pi}_U$; if $y = n$, then $\hat{\pi}_U = 1$.

  $\Rightarrow [\hat{\pi}_L, \hat{\pi}_U]$ is an exact $(1 - \alpha)$ for $\pi$ using mid p-value

- For example, $y = 3$, $n = 10$, an exact 95% CI is [0.08, 0.62]. That is,
  \[
  \frac{1}{2} P_{\pi=0.08}[Y = 3] + P_{\pi=0.08}[Y > 3] = 0.025
  \]
  \[
  \frac{1}{2} P_{\pi=0.62}[Y = 3] + P_{\pi=0.62}[Y < 3] = 0.025.
  \]
  This exact CI may be anti-conservative, that is, too short.
2 Contingency Tables

I. Probability Structure of a 2-way Contingency Table

I.1 Contingency tables

- $X, Y$: cat. var. $Y$ — usually random (except in a case-control study), response; $X$ — can be random or fixed, usually acts like a covariate. $X$ has $I$ levels, $Y$ has $J$ levels.

- A contingency table for $X, Y$ is an $I \times J$ table filled with data.

- For example,
• For example, from a random sample of \( n = 1127 \) Americans, we have the following contingency table:

Table 2.1. Cross classification of Belief in Afterlife by gender

<table>
<thead>
<tr>
<th>Belief in afterlife</th>
<th>Yes</th>
<th>No/Undecided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>509</td>
<td>116</td>
</tr>
<tr>
<td>Male</td>
<td>398</td>
<td>104</td>
</tr>
</tbody>
</table>

• With a contingency table for \( X, Y \), we would like to understand the association between \( X \) and \( Y \), the underlying probability structure of the table, etc.

• For example, for the afterlife table, we would like to see if one gender is more likely to believe in afterlife, or the overall proportion with belief in afterlife in the population, etc.
I.2 Sampling schemes, types of studies, probability structure

- Sampling schemes - ways to get data (tables):
  1. Multinomial sampling: From the population, we obtain a random sample, then cross classify individuals to table cells.

    ★ An example on belief in afterlife from \( n = 1127 \) Americans

    Table 2.1. Cross classification of Belief in Afterlife by gender

    | Belief in afterlife | Yes | No/Undecided |
    |--------------------|-----|-------------|
    | Female             | 509 | 116         |
    | Male               | 398 | 104         |

    ★ This is an example of Multinomial sampling.
    ★ The study using this sampling method is called a cross-sectional study.
In general, a $2 \times 2$ table from *multinomial sampling* \( Y \)

\[
\begin{array}{c|cc|c}
 & 1 & 2 & \\hline
X & n_{11} & n_{12} & n_{1+} \\
1 & n_{21} & n_{22} & n_{2+} \\
2 & n_{+1} & n_{+2} & n \\
\end{array}
\]

where \((n_{11}, n_{12}, n_{21}, n_{22})\) are random variables that have a
multinomial distribution with sample size \(n\) \((n = n_{11} + n_{12} + n_{21} + n_{22})\) and probabilities \(Y\)

\[
\begin{array}{c|cc|c}
 & 1 & 2 & \\hline
X & \pi_{11} & \pi_{12} & \\hline
1 & \pi_{21} & \pi_{22} & \\hline
2 & & & \\
\end{array}
\]

\((\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})\) define the *probability structure* of the
contingency table.
\* \( \pi_{ij} \)'s can be estimated by \( p_{ij} = \frac{n_{ij}}{n} \).

\* With \textit{multinomial sampling}, we can estimate many relevant quantities:

\[
\hat{P}[Y = 1] = \frac{n_{11} + n_{21}}{n} = \frac{n_{1+}}{n}
\]

\[
\hat{P}[X = 1] = \frac{n_{11} + n_{12}}{n} = \frac{n_{1+}}{n}
\]

\[
\hat{P}[Y = 1|X = 1] = \frac{n_{11}}{n_{11} + n_{12}} = \frac{n_{11}}{n_{1+}}
\]

\[
\hat{P}[X = 1|Y = 1] = \frac{n_{11}}{n_{11} + n_{21}} = \frac{n_{11}}{n_{1+}}
\]

\* For \textit{afterlife} example, we estimated that

\[
\hat{P}[\text{belief in afterlife}] = \frac{509 + 398}{1127} = 80\%
\]

\[
\hat{P}[\text{belief in afterlife}|\text{Female}] = \frac{509}{509 + 116} = 81\%
\]

\[
\hat{P}[\text{belief in afterlife}|\text{Male}] = \frac{398}{398 + 104} = 79\%...
\]
2. Product-multinomial sampling on $X$: For example, in a clinical trial for heart disease, we randomly assign 200 patients to treatment 1 and 100 patients to treatment 2 and may obtain potential data like the following:

<table>
<thead>
<tr>
<th></th>
<th>Better</th>
<th>No Change</th>
<th>Worse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>$n_{13}$</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$n_{23}$</td>
</tr>
</tbody>
</table>

Here we have

$(n_{11}, n_{12}, n_{13}) \perp (n_{21}, n_{22}, n_{23})$

$(n_{11}, n_{12}, n_{13}) \sim \text{multinomial}(200, (\pi_1, \pi_2, \pi_3)), \pi_1 + \pi_2 + \pi_3 = 1$

$(n_{21}, n_{22}, n_{23}) \sim \text{multinomial}(100, (\tau_1, \tau_2, \tau_3)), \tau_1 + \tau_2 + \tau_3 = 1$

$(\pi_1, \pi_2, \pi_3)$ and $(\tau_1, \tau_2, \tau_3)$ define the probability structure of this contingency table.
In general, the data looks like

\[
\begin{array}{c|c|c|c|c}
& 1 & 2 & 3 & \text{Total} \\
\hline
X = 1 & \begin{array}{c} n_{11} \\ n_{21} \end{array} & \begin{array}{c} n_{12} \\ n_{22} \end{array} & \begin{array}{c} n_{13} \\ n_{23} \end{array} & n_{1+} \\
\hline
X = 2 & & & & n_{2+}
\end{array}
\]

where \( n_{1+} \) and \( n_{2+} \), the sample sizes for \( X = 1 \) and \( X = 2 \), are fixed.

\((n_{11}, n_{12}, n_{13}) \perp (n_{21}, n_{22}, n_{23})\)

\((n_{11}, n_{12}, n_{13}) \sim \text{multinom}(n_{1+}, (\pi_1, \pi_2, \pi_3)), \pi_1 + \pi_2 + \pi_3 = 1\)

\((n_{21}, n_{22}, n_{23}) \sim \text{multinom}(n_{2+}, (\tau_1, \tau_2, \tau_3)), \tau_1 + \tau_2 + \tau_3 = 1\)

Since the likelihood of \( \pi \)'s and \( \tau \)'s is the product of the likelihood of \( \pi \)'s and the likelihood of \( \tau \)'s, this sampling scheme is called product-multinomial sampling on \( X \).

Clinical trials, cohort studies (prospective studies) all use this sampling scheme.
When $X$ is also random (so has a distribution in the population), $(\pi_1, \pi_2, \pi_3)'s$ defines the conditional distribution of $Y$ given $X = 1$

$(\tau_1, \tau_2, \tau_3)'s$ defines the conditional distribution of $Y$ given $X = 2$.

With *product-multinomial sampling on $X$*, we can only estimate conditional probabilities of $Y|X = x$. Other probabilities are not *estimable*. For example, we cannot estimate $P[Y = 1]$. 

\[ \text{Slide 47} \]
3. Product multinomial sampling on $Y$:

If $Y$ represents a rare event, then a prospective study is *inefficient*. For example, if we would like to investigate the association between *smoking* and *lung cancer* and conduct a prospective study

<table>
<thead>
<tr>
<th>Lung Cancer</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>No</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
</tbody>
</table>

then $n_{11}$, $n_{21}$ will be small unless $n_{1+}$ and $n_{2+}$ are very large.

This will yield an inefficient study.
We may consider a design such as the following one:

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>No</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
</tbody>
</table>

$n_{+1} = 100$  $n_{+2} = 200$

All cell counts will not be small $\Rightarrow$ efficient.

$n_{11} \perp n_{12}$

$n_{11} \sim \text{Bin}(n_{+1}, \pi_1)$, $\pi_1 = P[\text{smoking}| \text{case}]$.

$n_{12} \sim \text{Bin}(n_{+2}, \pi_2)$, $\pi_2 = P[\text{smoking}| \text{control}]$.

We can still investigate the association between smoking and lung cancer using this design.

This sampling scheme is product-multinomial on $Y$.

The study is often called the case-control study.
* In general,

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>No</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
</tbody>
</table>

where $n_1, n_2$, are all fixed.

$n_{11} \perp n_{12}$

$n_{11} \sim \text{Bin}(n_1, \pi_1), \pi_1 = P[\text{smoking|case}].$

$n_{12} \sim \text{Bin}(n_2, \pi_2), \pi_2 = P[\text{smoking|control}].$
Example of a case-control study on MI (Table 2.4)

Table 2.4. Case-Control Study on MI

<table>
<thead>
<tr>
<th>Ever Smoker</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>172</td>
<td>173</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>346</td>
</tr>
</tbody>
</table>

Myocardial Infarction

262 519

where 262 is the sample size for MI cases, 519 is the sample size for controls.

From this study, we cannot estimate the quantities such as

- $P[MI]$ 
- $P[\text{Ever Smoking}]$ 
- $P[MI|\text{Ever smokers}]$ 
- $P[MI|\text{Never smokers}]$ ...
**Note:** *Multinomial sampling \( \Rightarrow \) product-multinomial sampling.*

For example, if we have data from a multinomial sampling with sample size \( n \):

\[
\begin{array}{cc}
Y & 1 & 2 \\
X & 1 & n_{11} & n_{12} \\
& 2 & n_{21} & n_{22} \\
\end{array}
\quad
\begin{array}{cc}
Y & 1 & 2 \\
X & 1 & \pi_{11} & \pi_{12} \\
& 2 & \pi_{21} & \pi_{22} \\
\end{array}
\]

Then we can view the data from *product-multinomial sampling on X* or *product-multinomial sampling on Y*.

That is:

\[
n_{11}|n_1^+ \sim \text{Bin}(n_1^+, \frac{\pi_{11}}{\pi_{11} + \pi_{12}}) \perp n_{21}|n_1^+ \sim \text{Bin}(n_2^+, \frac{\pi_{21}}{\pi_{21} + \pi_{22}})
\]

Or

\[
n_{11}|n^+_1 \sim \text{Bin}(n^+_1, \frac{\pi_{11}}{\pi_{11} + \pi_{21}}) \perp n_{12}|n^+_1 \sim \text{Bin}(n^+_2, \frac{\pi_{12}}{\pi_{12} + \pi_{22}})
\]
I.3 Sensitivity & Specificity in Diagnostic Tests

- In a diagnostic test, $X = \text{true disease status}, Y = \text{test result}$. Then we can form a $2 \times 2$ table:

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Disease</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Using data from *multinomial sampling* or *product-multinomial sampling* on $X$, we can estimate

Sensitivity = $P[Y = \text{Positive} | X = \text{Disease}]$ (True positive rate)

Specificity = $P[Y = \text{Negative} | X = \text{No disease}]$ (True negative rate)

1-sensitivity = false negative rate, 1-specificity = false positive rate.

These two quantities tell us how accurate a test/device is.

Manufacturer of a test device usually provides these two measures.
However, a customer (or potential patient) may be more interested in the following quantities:

\[ P[X = \text{Disease} | Y = \text{Positive}] \] (PV+)
\[ P[X = \text{No disease} | Y = \text{Negative}] \] (PV-)

An accurate test may not yield high PV+ and/or PV-.

For example, assume a mammogram (for breast cancer) has sensitivity=0.86 and specificity=0.88. If \( P[\text{breast cancer}] = 0.01 \). Then

\[
\text{PV+} = P[X = \text{BR} | Y = +] = \frac{P[X = \text{BR}, Y = +]}{P[Y = +]} \\
= \frac{P[Y = + | X = \text{BR}] P[X = \text{BR}]}{P[Y = + | X = \text{BR}] P[X = \text{BR}] + P[Y = + | X = \text{No BR}] P[X = \text{No BR}]} \\
= \frac{0.86 \times 0.01}{0.86 \times 0.01 + (1 - 0.88) \times (1 - 0.01)} = 6.8\% \\
\]

Similarly, PV- = 99.8% (without the test, \( P[\text{No BR}] = 0.99 \)).
### I.4 Independence of $X$ and $Y$

- $X$ and $Y$ are random with the underlying probability structure

\[
\begin{array}{c|ccc}
& 1 & 2 & J \\
\hline
X & \pi_{11} & \pi_{12} & . & \pi_{1J} \\
2 & \pi_{21} & \pi_{22} & . & \pi_{2J} \\
. & . & . & . & . \\
I & \pi_{I1} & \pi_{I2} & . & \pi_{IJ} \\
\end{array}
\]

- $X \perp Y$

$\iff P[X = i, Y = j] = P[X = i]P[Y = j]$ for $i = 1, 2, \ldots, I, j = 1, 2, \ldots, J$.  

$\iff \pi_{ij} = \pi_{i+} \pi_{+j}$ for $i = 1, 2, \ldots, I, j = 1, 2, \ldots, J$.  

$\iff \pi_{i+} = \pi_{i1} + \pi_{i2} + \ldots + \pi_{iJ}, \pi_{+j} = \pi_{1j} + \pi_{2j} + \ldots + \pi_{IJ}$  

$\iff P[Y = j|X = i] = P[Y = j|X = k]$ for all $i, j, k$. 

---

Slide 55
• When $X$ and $Y$ are random 2-level cat. variables, the underlying probability structure is

$$
\begin{array}{c|cc}
Y & 1 & 2 \\
\hline
X & \pi_{11} & \pi_{12} \\
2 & \pi_{21} & \pi_{22} \\
\end{array}
$$

• $X \perp Y$

$\iff \pi_{ij} = \pi_{i+} \pi_{j+}$ for $i, j = 1, 2$ ($\pi_{i+} = \pi_{i1} + \pi_{i2}, \pi_{j+} = \pi_{1j} + \pi_{2j}$)

We only need one of them, e.g. $\pi_{11} = \pi_{1+} \pi_{+1}$

$\iff P[Y = 1|X = 1] = P[Y = 1|X = 2]$, i.e.

$$
\pi_1 = \frac{\pi_{11}}{\pi_{1+}} = \frac{\pi_{21}}{\pi_{2+}} = \pi_2
$$
II Comparing Proportions in $2 \times 2$ Tables

II.1 Difference of proportions

- Given data from a \textit{multinomial sampling} or \textit{product-multinomial sampling} on $X$

\[
\begin{array}{ccc}
 & Y & \\
X & 1 & 2 \\
1 & n_{11} & n_{12} & n_{1+} \\
2 & n_{21} & n_{22} & n_{2+} \\
\end{array}
\]

we would like to make inference on $\pi_1 - \pi_2$ where

$\pi_1 = P[Y = 1|X = 1]$ is the success probability for row 1 and

$\pi_2 = P[Y = 1|X = 2]$ is the success probability for row 2.

- $X \perp Y \iff \pi_1 - \pi_2 = 0.$
1. Estimate of $\pi_1 - \pi_2$:

$$p_1 - p_2 = \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}}.$$

2. Estimated $SE$ (standard error):

$$SE(p_1 - p_2) = \sqrt{\frac{p_1(1 - p_1)}{n_{1+}} + \frac{p_2(1 - p_2)}{n_{2+}}}$$

3. Large-sample $(1 - \alpha)$ CI for $\pi_1 - \pi_2$:

$$p_1 - p_2 \pm z_{\alpha/2} SE(p_1 - p_2).$$

If this CI does not contain 0, we can reject $H_0 : X \perp Y$ at significance level $\alpha$. 
Example: Aspirin and heart attack.

In a 5-yr study, 22,000+ physicians were randomized (blinded) to the placebo/aspirin (one tablet every other day) group:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Placebo</th>
<th>Aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Placebo</td>
<td>189</td>
<td>10,845</td>
</tr>
<tr>
<td>Aspirin</td>
<td>104</td>
<td>10,933</td>
</tr>
</tbody>
</table>

1. Difference of MI probabilities between placebo and aspirin groups:

\[ p_1 - p_2 = \frac{189}{11034} - \frac{104}{11037} = 0.0171 - 0.0094 = 0.0077. \]

2. \[ SE = \sqrt{0.0171(1 - 0.0171)/11034 + 0.0094(1 - 0.0094)/11037} = 0.0015. \]

3. Large sample 95% CI of Difference of MI probabilities:

\[ 0.0077 \pm 1.96 \times 0.0015 = [0.0048, 0.0106]. \]

⇒ Physicians in placebo group are more likely to develop MI.
II.2 Relative Risk

- When both \( \pi_1 \) and \( \pi_2 \) are close to zero (rare event), the difference \( \pi_1 - \pi_2 \) may not be very meaningful.

For example,

Case 1: \( \pi_1 = 0.01, \pi_2 = 0.001 \Rightarrow \pi_1 - \pi_2 = 0.009 \)

Case 2: \( \pi_1 = 0.41, \pi_2 = 0.401 \Rightarrow \pi_1 - \pi_2 = 0.009 \)

The above cases have the same difference \( \pi_1 - \pi_2 \). However, the meanings are totally different.

- For rare events, a more relevant measure for difference is the relative risk (RR):

\[
RR = \frac{\pi_1}{\pi_2}.
\]
• Properties of the *relative risk* (RR):

1. $0 < RR < \infty$

2. $\pi_1 > \pi_2 \iff RR > 1$
   
   $\pi_1 = \pi_2 \iff RR = 1$

3. $\pi_1 < \pi_2 \iff RR < 1$

3. $X \perp Y \iff RR = 1$

• Estimate of RR: Given the $2 \times 2$ table from *multinomial sampling* or *product-multinomial sampling* on $X$, RR can be estimated by

$$\hat{RR} = \frac{p_1}{p_2}. $$
• RR also has a nice interpretation. For the Aspirin Study, the RR estimate is

\[
\hat{RR} = \frac{p_1}{p_2} = \frac{0.0171}{0.0094} = 1.82.
\]

⇒ Physicians receiving the placebo are 82% more likely to develop MI (over 5 yrs) than physicians receiving aspirin.

• SE and CI for RR are complicated, Proc Freq calculates CI for RR and other measures:

```plaintext
data table2_3;
  input group $ mi $ count @@;
datalines;
  placebo yes 189 placebo no 10845
  aspirin yes 104 aspirin no 10933
;

title "Analysis of MI data";
proc freq data=table2_3 order=data;
  weight count;
  tables group*mi / norow nocol nopercent or;
run;
```
Output from the above SAS program:

The FREQ Procedure

Table of group by mi

<table>
<thead>
<tr>
<th>group</th>
<th>frequency</th>
<th>no</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>placebo</td>
<td>189</td>
<td>10845</td>
<td>11034</td>
</tr>
<tr>
<td>aspirin</td>
<td>104</td>
<td>10933</td>
<td>11037</td>
</tr>
<tr>
<td>total</td>
<td>293</td>
<td>21778</td>
<td>22071</td>
</tr>
</tbody>
</table>

Statistics for Table of group by mi

Odds Ratio and Relative Risks

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio</td>
<td>1.8321</td>
<td>1.4400 2.3308</td>
</tr>
<tr>
<td>Relative Risk (Column 1)</td>
<td>1.8178</td>
<td>1.4330 2.3059</td>
</tr>
<tr>
<td>Relative Risk (Column 2)</td>
<td>0.9922</td>
<td>0.9892 0.9953</td>
</tr>
</tbody>
</table>

Sample Size = 22071

A 95% CI for RR is [1.43, 2.31]. We are 95% sure that physicians receiving the placebo is at least 43% and at most 131% more likely to develop MI (over 5 yrs) than physicians receiving aspirin.
II.3 Odds Ratio

- Odds of a prob (of an event): $\pi = P(A)$, then

$$\omega = \frac{\pi}{1 - \pi} = \frac{\text{success prob}}{\text{failure prob}}$$

is called the *odds* of $\pi$ (or of the event $A$). $0 < \omega < \infty$.

For example, $\pi = 0.75$, then $\omega = 0.75/(1 - 0.75) = 3$.

For a rare event ($\pi \approx 0$), $\pi \approx \omega$.

- The event prob $\pi$ is related to odds $\omega$ as:

$$\pi = \frac{\omega}{1 + \omega}.$$  

For example, $\omega = 4$, then $\pi = 4/(1 + 4) = 0.8$. 

• For the $2 \times 2$ table

\[
\begin{array}{c|cc}
Y & 1 & 2 \\
\hline
X & 1 & \text{ } \\
2 & \text{ } & \text{ }
\end{array}
\]

the odds ratio between row 1 ($\pi_1 = P[Y = 1|X = 1]$) and row 2 ($\pi_2 = P[Y = 1|X = 2]$) is defined as

\[
\theta = \frac{\text{odds}_1}{\text{odds}_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}.
\]

• Properties of the odds ratio

1. $0 < \theta < \infty$.
2. $\pi_1 > \pi_2 \Leftrightarrow \theta > 1$; $\pi_1 = \pi_2 \Leftrightarrow \theta = 1$; $\pi_1 < \pi_2 \Leftrightarrow \theta < 1$.
3. $X \perp Y \Leftrightarrow \theta = 1$. 
• Given the $2 \times 2$ table from *multinomial sampling or product-multinomial sampling* on $X$:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
</tbody>
</table>

*odds ratio* $\theta$ can be estimated by

$$\hat{\theta} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{1+}}{(1 - n_{11}/n_{1+})} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}},$$

• $\text{var}(\log \hat{\theta})$ can be estimated by

$$\text{var}(\log \hat{\theta}) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}.$$
We can construct a \((1 - \alpha)\) CI for true \(\theta\) as follows:

1. Get \((1 - \alpha)\) CI for \(\log(\theta)\):

\[
\log \hat{\theta} \pm z_{\alpha/2} SE(\log \hat{\theta}).
\]

2. Exponentiate both ends to get the CI for \(\theta\).

For the Aspirin Study,

\[
\hat{\theta} = \frac{189 \times 10933}{10845 \times 104} = 1.8321 (\approx RR)
\]

\[
\text{var}(\log \hat{\theta}) = \frac{1}{189} + \frac{1}{10845} + \frac{1}{104} + \frac{1}{10933} = 0.01509
\]

95\% CI for \(\log \theta\): \(\log(1.8321) \pm 1.96 \sqrt{0.01509} = [0.3647, 0.8462]\).

95\% CI for \(\theta\): \([e^{0.3647}, e^{0.8462}] = [1.44, 2.33]\).
• **Note 1:** If we have *multinomial sampling*:

\[
\begin{array}{cc|cc}
   & Y & \multicolumn{2}{c}{Y} \\
   & 1 & 2 & 1 & 2 \\
X & 1 & n_{11} & n_{12} & \pi_{11} & \pi_{12} \\
   & 2 & n_{21} & n_{22} & \pi_{21} & \pi_{22} \\
\end{array}
\]

the *odds ratio* \( \theta \) can be also defined as

\[
\theta = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}.
\]

MLE of \( \pi_{ij} \)'s are \( \hat{\pi}_{ij} = \frac{n_{ij}}{n} \Rightarrow \text{the same estimate of} \ \theta : \)

\[
\hat{\theta} = \frac{\hat{\pi}_{11}\hat{\pi}_{22}}{\hat{\pi}_{12}\hat{\pi}_{21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.
\]

• **Note 2:** If some of \( n_{ij} \)'s are small, add 0.5 to each cell then re-calculate \( \hat{\theta} \) and \( \text{var}(\log\hat{\theta}) \), *e.g.*

\[
\hat{\theta} = \frac{(n_{11} + 0.5)(n_{22} + 0.5)}{(n_{12} + 0.5)(n_{21} + 0.5)}
\]
The relationship between $\theta$ and RR:

\[
\theta = \frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)} = \frac{\pi_1}{\pi_2} \times \frac{(1 - \pi_2)}{(1 - \pi_1)} = RR \times \frac{(1 - \pi_2)}{(1 - \pi_1)}
\]

1. $RR = 1 \iff \theta = 1 \iff X \perp Y$.
2. $\pi_1 > \pi_2 \iff \theta > RR > 1$.
3. $\pi_1 < \pi_2 \iff \theta < RR < 1$.
4. When $\pi_1 \approx 0$ & $\pi_2 \approx 0$ (rare events), $\theta \approx RR$. 

\[\begin{array}{ccccccc}
0 & \theta & RR & 1 & RR & \theta \\
\end{array}\]
The odds ratio for case-control studies:

For the MI study (page 32)

Table 2.4. Case-Control Study on MI

<table>
<thead>
<tr>
<th>Ever Smoker</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>172</td>
<td>173</td>
</tr>
<tr>
<td>No</td>
<td>90</td>
<td>346</td>
</tr>
</tbody>
</table>

we know that we cannot estimate \( \pi_1 = P[MI|Eversmokers] \) and \( \pi_2 = P[MI|Neversmokers] \), and hence cannot estimate

\[
RR = \frac{\pi_1}{\pi_2}.
\]

However, we still want to assess the association between smoking and MI.
From the design, we can estimate

\[ \tau_1 = P[\text{Ever smoking} | \text{MI Case}] : \hat{\tau}_1 = 172/262 = 0.6565 \]

\[ \tau_2 = P[\text{Ever smoking} | \text{MI Control}] : \hat{\tau}_2 = 172/262 = 0.3333 \]

and the odds ratio between \( \tau_1 \) and \( \tau_2 \)

\[ \theta^* = \frac{\tau_1 / (1 - \tau_1)}{\tau_2 / (1 - \tau_2)} : \hat{\theta}^* = \frac{\hat{\tau}_1 / (1 - \hat{\tau}_1)}{\hat{\tau}_2 / (1 - \hat{\tau}_2)} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = 3.82. \]

It can be shown that

\[ \theta^* = \frac{\pi_1 / (1 - \pi_1)}{\pi_2 / (1 - \pi_2)} = \theta \]

So we can use a case-control study to make inference on \( \theta \)!

The formula for \( \text{var} \left( \log \hat{\theta} \right) \) is the same:

\[ \text{var} \left( \log \hat{\theta} \right) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}. \]
Therefore, for the Aspirin case-control study, the odds ratio of developing MI between ever smokers and never smokers is estimated as

\[ \hat{\theta} = 3.82. \]

\[
\text{var}(\log \hat{\theta}) = \frac{1}{172} + \frac{1}{173} + \frac{1}{90} + \frac{1}{346} = 0.0256.
\]

95% CI for \( \log \theta \):

\[
\log(3.82) \pm 1.96 \times \sqrt{0.0256} = [1.02665, 1.65385]
\]

95% CI for \( \theta \):

\[
[e^{1.02665}, e^{1.65385}] = [2.79, 5.227].
\]

Since MI is a rare event, \( RR \approx \theta \), so

\[ \widehat{RR} \approx 3.82 \approx 4. \]

That is, ever smokers is about 3 times more likely to develop MI than never smokers.
III $\chi^2$ Test for Independence between $X$ and $Y$ (nominal)

Suppose $X$ and $Y$ are random and have the prob structure:

\[
\begin{array}{c|ccc|c}
  & 1 & 2 & \cdots & J \\
\hline
X & \pi_{11} & \pi_{12} & \cdots & \pi_{1J} \\
2 & \pi_{21} & \pi_{22} & \cdots & \pi_{2J} \\
. & . & . & \cdots & . \\
I & \pi_{I1} & \pi_{I2} & \cdots & \pi_{IJ} \\
\end{array}
\]

Given data $\{n_{ij}\}$’s from a multinomial sampling, we would like to test $H_0: \pi_{ij} = \pi_{ij}(\theta)$, for $i = 1, \ldots, I$, and $j = 1, \ldots, J$, where $\theta$ is a parameter vector with $\text{dim}(\theta) = k$.

If $\text{dim}(\theta) = 0$, then $\pi_{ij}$’s are totally known under $H_0$. 
III.1 General Pearson $\chi^2$ test and LRT

- MLE $\hat{\theta}$ of $\theta$ under $H_0$; $\hat{\mu}_{ij} = n\pi_{ij}(\hat{\theta})$, where $n = n_{++}$.

- If $H_0$ is true and $n$ is large such as $\hat{\mu}_{ij}$’s are reasonably large ($\hat{\mu}_{ij} \geq 5$), then the Pearson stat

$$\chi^2 = \sum_{\text{all cells}} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \overset{H_0}{\sim} \chi^2_{df}$$

where $df = IJ - 1 - \dim(\theta)$.

Reject $H_0$ at level $\alpha$ if $\chi^2 \geq \chi^2_{df, \alpha}$.

- LRT

$$G^2 = 2 \sum_{\text{all cells}} n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right) \overset{H_0}{\sim} \chi^2_{df}.$$ 

- Calculation of $df$:

$df = \# \text{ of unknown parameters under } H_1 \cup H_0 - \# \text{ of unknown parameters under } H_0.$
Some $\chi^2$ distributions

![Graph showing chi-squared distributions with different degrees of freedom (df = 1, df = 5, df = 10, df = 20).](image)

Figure 2.2. Examples of chi-squared distributions.
III.2 Test of independence

- $X \perp Y \iff H_0 : \pi_{ij} = \pi_i \pi_j, i = 1, ..., I, j = 1, ..., J$

- The MLE of $\pi_{i+}$'s and $\pi_{+j}$'s are

  \[ \hat{\pi}_{i+} = \frac{n_{i+}}{n}, \hat{\pi}_{+j} = \frac{n_{+j}}{n} \]

- $\hat{\mu}_{ij}$ is equal to

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

- Pearson $\chi^2$ and LRT:

  \[ \chi^2 = \sum_{\text{all cells}} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}, G^2 = 2 \sum_{\text{all cells}} n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right) \sim H_0 \chi^2_{df} \]

  \[ df =IJ - 1 - (I - 1 + J - 1) = (I - 1)(J - 1). \]

  Reject $H_0 : X \perp Y$ if $\chi^2$ or $G^2 \geq \chi^2_{df, \alpha}$. 
• **Note**: With data \( \{n_{ij}\} \)'s from a *multinomial sampling* or *product-multinomial sampling* on \( X \), we can test \( H_0 : X \perp Y \) by testing

\[
H_0 : P[Y = j|X = i] = P[Y = j|X = k] \quad \text{for all } i, j, k
\]

(cond. dist. of \( Y \) given \( X \) is the same across all levels of \( X \))

It can be shown that the Pearson \( \chi^2 \) and LRT test stats are the same with the *same* null dist \( \chi^2_{(I-1)(J-1)} \).
• Example: Gender gap in party identification

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X – Gender</strong></td>
<td>Female</td>
<td>762</td>
<td>327</td>
<td>468</td>
</tr>
<tr>
<td>Male</td>
<td>484</td>
<td>239</td>
<td>477</td>
<td></td>
</tr>
<tr>
<td>**</td>
<td>1246</td>
<td>566</td>
<td>945</td>
<td></td>
</tr>
</tbody>
</table>

Then $\hat{\mu}_{11} = 1557 \times 1246/2757 = 703.7$, $\hat{\mu}_{12} = 1557 \times 566/2757 = 319.6$, etc.

$\chi^2 = \frac{(762 - 703.7)^2}{703.7} + \frac{(327 - 319.6)^2}{319.6} + ... = 30.1$

$G^2 = 2(762 \log(762/703.7) + 327 \log(327/319.6) + ...) = 30.0$

$\chi^2_{2,0.05} = 5.99$

Both Pearson test and LRT reject $H_0: X \perp Y$ at level 0.05.

**Note:** $\chi^2 \approx G^2$ even if $H_0$ is likely not true.
• **SAS program for the example:**

```sas
data table2_5;
  input gender $ party $ count @@;
datalines;
  female dem 762 female ind 327 female rep 468
  male   dem 484 male   ind 239 male   rep 477
;

title "Analysis of Party Identification data";
proc freq data=table2_5 order=data;
  weight count;
  tables gender*party / norow nocol nopercent chisq expected measures cmh;
run;
```

• **Output from the above program:**

```
Analysis of Party Identification data

The FREQ Procedure

Table of gender by party

<table>
<thead>
<tr>
<th>gender</th>
<th>party</th>
<th>Frequency</th>
<th>Expected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>dem</td>
<td>ind</td>
<td>rep</td>
</tr>
<tr>
<td>female</td>
<td></td>
<td>762</td>
<td>327</td>
<td>468</td>
</tr>
<tr>
<td></td>
<td></td>
<td>703.67</td>
<td>319.65</td>
<td>533.68</td>
</tr>
<tr>
<td>male</td>
<td></td>
<td>484</td>
<td>239</td>
<td>477</td>
</tr>
<tr>
<td></td>
<td></td>
<td>542.33</td>
<td>246.35</td>
<td>411.32</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1246</td>
<td>566</td>
<td>945</td>
</tr>
</tbody>
</table>
```

Slide 79
### Statistics for Table of gender by party

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>2</td>
<td>30.0701</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>2</td>
<td>30.0167</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>28.9797</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Phi Coefficient</td>
<td></td>
<td>0.1044</td>
<td></td>
</tr>
<tr>
<td>Contingency Coefficient</td>
<td></td>
<td>0.1039</td>
<td></td>
</tr>
<tr>
<td>Cramer's V</td>
<td></td>
<td>0.1044</td>
<td></td>
</tr>
</tbody>
</table>

Sample Size = 2757

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>ASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>0.1710</td>
<td>0.0315</td>
</tr>
<tr>
<td>Kendall's Tau-b</td>
<td>0.0964</td>
<td>0.0180</td>
</tr>
<tr>
<td>Stuart’s Tau-c</td>
<td>0.1078</td>
<td>0.0202</td>
</tr>
<tr>
<td>Somers’ D C</td>
<td>R</td>
<td>0.1097</td>
</tr>
<tr>
<td>Somers’ D R</td>
<td>C</td>
<td>0.0848</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>0.1025</td>
<td>0.0190</td>
</tr>
<tr>
<td>Spearman Correlation</td>
<td>0.1016</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

### Summary Statistics for gender by party

#### 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>28.9797</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>1</td>
<td>28.9797</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>2</td>
<td>30.0592</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
III.3 Cell residuals for a contingency table

- Under $H_0 : X \perp Y$,
  \[ \hat{\mu}_{ij} = \frac{n_i + n_j}{n}. \]

- Then we calculate standardized Pearson residuals:
  \[ e_{ij}^{st} = \frac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1 - p_i^+)(1 - p_+^j)}}. \]

- Under $H_0 : X \perp Y$, $E(e_{ij}^{st}) \approx 0$, $\text{var}(e_{ij}^{st}) \approx 1$, and $e_{ij}^{st}$ behaves like a $N(0, 1)$ variable.

- We can use $e_{ij}^{st}$ to check the departure from $H_0 : X \perp Y$.

- For the Party Identification example, $p_{1^+} = 1557/2757 = 0.565$, $p_{+1} = 1246/2757 = 0.452$

  \[ \Rightarrow e_{11}^{st} = \frac{762 - 703.7}{\sqrt{703.7(1 - 0.565)(1 - 0.452)}} = 4.50 \]
• We can use Proc Genmod of SAS to get the *standardized Pearson residuals*:

```
Proc Genmod order=data;
  class gender party;
  model count = gender party / dist=poisson link=log residuals;
run;
```

• Part of the output:

<table>
<thead>
<tr>
<th>Observation</th>
<th>Raw Residual</th>
<th>Pearson Residual</th>
<th>Deviance Residual</th>
<th>Std Deviance Residual</th>
<th>Std Pearson Residual</th>
<th>Likelihood Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58.328618</td>
<td>2.1988558</td>
<td>2.1694814</td>
<td>4.4419109</td>
<td>4.5020535</td>
<td>4.4877799</td>
</tr>
<tr>
<td>2</td>
<td>7.3547334</td>
<td>0.4113702</td>
<td>0.4098076</td>
<td>0.6967948</td>
<td>0.6994517</td>
<td>0.6985339</td>
</tr>
<tr>
<td>3</td>
<td>-65.683351</td>
<td>-2.84324</td>
<td>-2.904774</td>
<td>-5.430995</td>
<td>-5.315946</td>
<td>-5.34911</td>
</tr>
<tr>
<td>4</td>
<td>-58.328618</td>
<td>-2.504669</td>
<td>-2.551707</td>
<td>-4.586602</td>
<td>-4.502054</td>
<td>-4.528391</td>
</tr>
<tr>
<td>5</td>
<td>-7.354733</td>
<td>-0.468583</td>
<td>-0.470944</td>
<td>-0.702976</td>
<td>-0.699452</td>
<td>-0.701036</td>
</tr>
<tr>
<td>6</td>
<td>65.683351</td>
<td>3.2386734</td>
<td>3.157751</td>
<td>5.1831197</td>
<td>5.3159455</td>
<td>5.2670354</td>
</tr>
</tbody>
</table>

The observation order is for row 1, then row 2, etc.
• Put the standardized Pearson residuals in the original table:

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>Independent</th>
<th>Republican</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X – Gender</strong></td>
<td><strong>Female</strong></td>
<td>4.5</td>
<td>0.7</td>
<td>-5.3</td>
</tr>
<tr>
<td></td>
<td><strong>Male</strong></td>
<td>-4.5</td>
<td>-0.7</td>
<td>5.3</td>
</tr>
</tbody>
</table>

We see from the table that the independence model does not fit the data well.

There are significantly more democrat females (less males) than predicted by the independence model, there are significantly less republican females (more males) than predicted by the model.
IV Testing Independence for Ordinal Data

IV.1 $X, Y$ are both ordinal random cat. variables; Mantel-Haenszel $M^2$ (CMH1)

- Assign scores $u_1 < u_2 < \cdots < u_I$ to $X$ and $v_1 < v_2 < \cdots < v_J$ to $Y$

<table>
<thead>
<tr>
<th></th>
<th>$1(v_1)$</th>
<th>$j(v_j)$</th>
<th>$J(v_J)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1(u_1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i(u_i)$</td>
<td></td>
<td>$\pi_{ij}$</td>
<td></td>
</tr>
<tr>
<td>$I(u_I)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Want to test $H_0 : X \perp Y$ given data such as
\[
\begin{array}{lll}
X & & \\
\begin{array}{c}
u_1 \\
u_2 \\
u_3 \\
\end{array} & \\
\begin{array}{ccc}
2 & 1 & 3 \\
1 & 2 & 1 \\
1 & 1 & 2 \\
\end{array}
\end{array}
\Rightarrow
\begin{array}{ll}
\text{Patient} & X & Y \\
1 & u_1 & v_1 \\
2 & u_1 & v_1 \\
3 & u_1 & v_2 \\
4 & u_1 & v_3 \\
5 & u_1 & v_3 \\
6 & u_1 & v_3 \\
7 & u_2 & v_1 \\
8 & u_2 & v_2 \\
9 & u_2 & v_2 \\
10 & u_2 & v_3 \\
11 & u_3 & v_1 \\
12 & u_3 & v_2 \\
13 & u_3 & v_3 \\
14 & u_3 & v_3 \\
\end{array}
\]
• Pearson correlation coefficient describes linear relationship between $X$ and $Y$ and can be used to test $H_0 : X \perp Y$:

$$r = \frac{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{n} \sum_{i=1}^{I} n_{i+} u_i = \sum_{i=1}^{I} p_{i+} u_i = \bar{u}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{n} \sum_{j=1}^{J} n_{+j} v_j = \sum_{j=1}^{J} p_{+j} v_j = \bar{v}$$
\[ r = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} (u_i - \bar{u})(v_j - \bar{v})}{\sqrt{\sum_{i=1}^{I} p_{i+}(u_i - \bar{u})^2 \sum_{j=1}^{J} p_{+j}(v_j - \bar{v})^2}} \]

- It can be shown that under \( H_0 : X \perp Y \)

\[ \sqrt{n-1}r \overset{a}{\sim} N(0, 1) \]

\[ M^2 = (n-1)r^2 \overset{a}{\sim} \chi_1^2 \]

This is the Mantel-Haenszel test for \( H_0 : X \perp Y \) (cmh1 in SAS).

- **Note:** We don’t have to expand the data to calculate \( r \). Proc Freq calculates \( r \) and \( M^2 \).
• How to choose scores \( \{u_i\} \)'s for \( X \) and \( \{v_j\} \)'s for \( Y \):

1. Any increasing/decreasing seq is ok for \( \{u_i\} \)'s and \( \{v_j\} \)'s. They have to be chosen before analyzing data.
2. Mid-rank. For example,

\[
\begin{array}{ccccc}
 & 1 & 2 & 3 & u_i \\
1 & 2 & 1 & 3 & 6 & 3.5 \\
2 & 1 & 2 & 1 & 4 & 8.5 \\
3 & 1 & 1 & 2 & 4 & 12.5 \\
4 & 4 & 4 & 6 & \\
\end{array}
\]

\[
\begin{array}{ccc}
v_j \\
2.5 & 6.5 & 11.5 \\
\end{array}
\]

```
Proc Freq order=data
   tables x*y/CMH1 Scores=rank;
run;
```

3. The default is “1, 2, \cdots, I” for \( X \) and “1, 2, \cdots, J” for \( Y \) in SAS.
• **Note 1**: $M^2$ only detects “linear trend” between $X$ and $Y$, Pearson $\chi^2$ and LRT $G^2$ detects any deviation from indep.

• **Note 2**: Proc corr of SAS uses (as the default)

\[
t = \frac{(n - 2)^{1/2}}{\sqrt{1 - r^2}} \left( \frac{r^2}{1 - r^2} \right)^{1/2}
\]

to test $H_0 : \rho = 0$ by comparing $t$ to $t_{n-2}$. $M^2$ and $t^2$ are asym. equiv. under $H_0$.

• From slide 80, $M^2 = 28.98$ using 1,2 for gender and 1,2,3 for party identification. Reject $H_0 : X \perp Y$.

• **Note 3**: $M^2$ is for a 2-sided test. We can use $\sqrt{n - 1}r$ for a one-sided test.

From slide 80, $\sqrt{n - 1}r = \sqrt{28.98} = 5.4 \Rightarrow$ reject $H_0 : X \perp Y$ in favor of $H_1 : \rho > 0$ (even if $r = 0.1$).
• Example: Mother’s alcohol consumption and infant malformation (Table 2.7 on p. 42)

<table>
<thead>
<tr>
<th>Alcohol Consumption</th>
<th>Malformation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Present ($Y = 1$)</td>
<td>Absent ($Y = 0$)</td>
</tr>
<tr>
<td>0</td>
<td>48</td>
<td>17,066</td>
</tr>
<tr>
<td>&lt; 1</td>
<td>38</td>
<td>14,464</td>
</tr>
<tr>
<td>1 – 2</td>
<td>5</td>
<td>788</td>
</tr>
<tr>
<td>3 – 5</td>
<td>1</td>
<td>126</td>
</tr>
<tr>
<td>≥ 6</td>
<td>1</td>
<td>37</td>
</tr>
</tbody>
</table>

$\chi^2 = 12.1$ (p-value = 0.016), $G^2 = 6.2$ (p-value = 0.185) ⇒ mixed results.

Assigned scores for alcohol consumption: 0, 0.5, 1.5, 4, 7 and 0/1 for absent/present ⇒ $r = 0.0142$, $M^2 = 6.6$, p-value = $P[\chi^2_1 \geq M^2] = 0.01$.

$\chi^2$, $G^2$, $M^2$ may not be valid ⇒ Exact test (later).
• SAS program:

```sas
data table2_7;
  input alcohol malform count @@;
datalines;
  0  1  48  0  0  17066
  0.5  1  38  0.5  0  14464
  1.5  1  15  1.5  0  788
  4  1  1  4  0  126
  7  1  1  7  0  37
; title "Analysis of infant malformation data";
proc freq data=table2_7;
  weight count;
  tables alcohol*malform / measures chisq cmh;
run;
```

• Part of the output:

```
Statistics for Table of alcohol by malform

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>4</td>
<td>12.0821</td>
<td>0.0168</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>4</td>
<td>6.2020</td>
<td>0.1846</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>6.5699</td>
<td>0.0104</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th></th>
<th>Value</th>
<th>ASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td></td>
<td>0.0142</td>
<td>0.0106</td>
</tr>
<tr>
<td>Spearman Correlation</td>
<td></td>
<td>0.0033</td>
<td>0.0059</td>
</tr>
</tbody>
</table>
```

Slide 91
### IV.2 Trend test for $I \times 2$ and $2 \times J$ tables

- For an $I \times 2$ table where $X$ is an $I$-level ordinal variable and $Y$ is a 2-level variable (such as the infant malformation table) from a *multinomial sampling* or *product-multinomial sampling* on $X$:

$$
\begin{array}{c|cc|c}
Y & 1 & 0 \\
\hline
u_1 & n_{11} & n_{12} & n_{1+} \\
\hline
u_2 & n_{21} & n_{22} & n_{2+} \\
\vdots & \vdots & \vdots & \vdots \\
u_I & n_{I1} & n_{I2} & n_{I+} \\
\end{array}
$$

we can assign scores to $X$ and any scores (usually 0/1) to $Y \Rightarrow M^2$. 

---

Slide 92
The Mantel-Haenszel $M^2$ can be derived in a different way (taken from Section 3.2.1)

Consider

$$\pi_i = P[Y = 1|X = u_i].$$

Assume a linear trend model for $\pi_i$:

$$\pi_i = \alpha + \beta u_i$$

Then $H_0 : X \perp Y \implies H_0^* : \beta = 0$

An unbiased estimate of $\pi_i$:

$$\hat{\pi}_i = \frac{n_i 1}{n_i +} = p_i \leftarrow \text{sample proportion at } X = u_i$$

The trend model implies the following linear model for $p_i$:

$$p_i = \alpha + \beta u_i + \varepsilon_i,$$
\[ \text{var}(\varepsilon_i) = \pi_i(1 - \pi_i)/n_{i+}, \text{ which equals } \alpha(1 - \alpha)/n_{i+} \text{ under } H_0^*: \beta = 0 \]

\[ \implies \text{WLS (weighted LS, weighted by sample size } n_{i+}) \text{ estimate of } \beta \]

\[ \hat{\beta} = \frac{\sum_{i=1}^{I} n_{i+}(u_i - \bar{u})(p_i - p)}{\sum_{i=1}^{I} n_{i+}(u_i - \bar{u})^2}, \]

where

\[ \bar{u} = \frac{1}{n} \sum_{i=1}^{I} n_{i+}u_i \leftarrow \text{sample mean of } \{X_i\} \]

\[ p = \frac{n+1}{n} \leftarrow \text{pooled sample response rate} \]

\[ \text{var}(\hat{\beta}) \text{ under } H_0 \text{ can be estimated by} \]

\[ \hat{\text{var}}_{H_0}(\hat{\beta}) = \frac{p(1 - p)}{\sum_{i=1}^{I} n_{i+}(u_i - \bar{u})^2}. \]
For testing $H_0^* : \beta = 0$, let’s use Wald test

$$Z = \frac{\hat{\beta}}{\sqrt{\text{var}_{H_0}(\hat{\beta})}}$$

Under $H_0 : X \perp Y$, $Z \sim N(0, 1)$ or $Z^2 \sim \chi_1^2$.

- $Z^2$ or $Z$ is the Cochran-Armitage Trend test.
  
  It can be shown that $Z^2 = nr^2$. Remember $M^2 = (n - 1)r^2$
  
  $$\Rightarrow Z^2 = \frac{n}{n - 1} M^2 \approx M^2$$

- SAS program:

```
  title "Trend test of infant malformation data";
  proc freq data=table2_7 order=data;
    weight count;
    tables alcohol*malform / trend;
  run;
```
• Part of the output:

Statistics for Table of alcohol by malform

Cochran-Armitage Trend Test

<table>
<thead>
<tr>
<th>Statistic (Z)</th>
<th>2.5632</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-sided Pr &gt; Z</td>
<td>0.0052</td>
</tr>
<tr>
<td>Two-sided Pr &gt;</td>
<td>Z</td>
</tr>
</tbody>
</table>

Sample Size = 32574

• We see that $Z = 2.5632$. Both one-sided and 2-sided p-values are significant. Since $Z > 0$, we conclude that $\beta > 0$.

We can confirm the relationship:

$$Z^2 = \frac{n}{n-1} M^2.$$
For a $2 \times J$ table where $X$ is nominal or ordinal variable, $Y$ is an ordinal variable with data $\{n_{ij}\}$'s from a multinomial sampling or product-multinomial sampling on $X$

$$
\begin{array}{cccc}
& v_1 & v_2 & \cdots & v_J \\
X & 1 & n_{11} & n_{12} & \cdots & n_{1J} \\
& n_{21} & n_{22} & \cdots & n_{2J} \\
\end{array}
$$

We have a situation similar to the two sample $t$-test for comparing the means of $Y$ scores between $X = 1$ and $X = 2$. It can be shown that $t^2 \approx M^2$ ($M^2$ will be independent of the score choice for $X$).

If we use mid-ranks as the scores for $Y$, $M^2$ is the same as Mann-Whitney test.
IV.3 Tests for nominal-ordinal tables

- $X$ – nominal, $Y$ – ordinal with data from multinomial sampling or product-multinomial sampling on $X$ such as:

\[
\begin{array}{c|ccc|c}
\text{Y} & v_1 & v_2 & v_3 & \\
\hline
1 & n_{11} & n_{12} & n_{13} & n_{1+} \\
2 & n_{21} & n_{22} & n_{23} & n_{2+} \\
3 & n_{31} & n_{32} & n_{33} & n_{3+} \\
\end{array}
\]

- $H_0 : X \perp Y$
  \[\downarrow\]
  The cond. dists. of $Y$ are the same across levels of $X$
  \[\downarrow\]
  The mean scores of $Y$ at $X = i$ are the same across levels of $X$

- This is an ANOVA problem.
• We can use the ANOVA $F$-test to test $X \perp Y$:

$$F = \frac{SST/(I-1)}{SSE/(n-I)} \overset{H_0}{\sim} F_{I-1,n-I}$$

• Equivalently (for large $n$), we can use

$$\chi^2 = \frac{SST}{SSE^*/(n-1)} \overset{H_0}{\sim} \chi^2_{I-1}$$

where $SSE^*$ is the modified sum of squares of errors.

The test $\chi^2$ is called cmh2 by SAS:

```plaintext
proc freq;
  weight count;
  tables x*y / cmh2;
run;
```
V. Exact Inference for Sparse Tables

V.1 Fisher’s exact test for $2 \times 2$ tables

- $X, Y$ – 2 level cat. variables with structure

\[
\begin{array}{c|cc}
   & 1 & 2 \\
\hline
X 1 & \pi_{11} & \pi_{12} \\
2   & \pi_{21} & \pi_{22} \\
\end{array}
\]

- Want to test $H_0: X \perp Y$ given data, WLOG, assuming from a multinomial sampling:

\[
\begin{array}{c|cc}
   & 1 & 2 \\
\hline
X 1 & n_{11} & n_{12} \\
2   & n_{21} & n_{22} \\
\end{array}
\]
• When \( \{n_{ij}\} \)'s are large, we can use the Pearson \( \chi^2 \) or LRT \( G^2 \) to test \( H_0 : X \perp Y \).

• However, when some cell counts \( \{n_{ij}\} \)'s are small, the exact dist. of \( \chi^2 \) or LRT \( G^2 \) under \( H_0 \) may be far from \( \chi^2_1 \), \( \rightarrow \) use of asym. dist. may give wrong conclusions.

• Fisher’s tea example: Fisher’s colleague, Muriel Bristol claimed she could tell whether or not tea (or milk) was added to the cup first.

<table>
<thead>
<tr>
<th>Muriel’s Guess</th>
<th>Milk</th>
<th>Tea</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Milk</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Tea</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
• By the design of Fisher’s tea example, Pearson $\chi^2$ or $G^2$ can at most take 5 different values (there are only 5 possible different tables). Therefore, the $\chi^2_1$ approximate dist. of $\chi^2$ or $G^2$ is very poor!

• Even if we assumed multinomial sampling, there would only be $\binom{8+3}{3} = 165$ tables. Moreover, $n_{ij}$’s are small. The $\chi^2_1$ approximation of Pearson $\chi^2$ or $G^2$ will still be very poor.

• Let us develop an exact test for testing $H_0 : X \perp Y$ in these kind of sparse $2 \times 2$ tables.

• Let us assume multinomial sampling and would like to test $H_0 : \theta = 1(X \perp Y) \text{ v.s. one-sided alternative } H_a : \theta > 1$. 
• With *multinomial sampling*, \((n_{11}, n_{12}, n_{21}, n_{22})\) are random variables (only the sum \(n = n_{++}\) is fixed).

• Under \(H_0: \theta = 1(X \perp Y)\), \(\pi_{ij} = \pi_i+\pi_j\), there are two unknown \(\pi_{1+}, \pi_{+1}\) parameters. So the distribution of data \((n_{11}, n_{12}, n_{21}, n_{22})\) is unknown even under \(H_0\).

• It can be shown that under \(H_0: \theta = 1(X \perp Y)\), the conditional distribution of \(n_{11}|n_{1+}, n_{+1}\) is totally known:

\[
P[n_{11} = t_0] = \frac{{n_{1+}} \cdot {n_{2+}}}{n \choose n_{+1}} \cdot \frac{{n_{+1}} - t_0}{n_{+1} - t_0}.
\]

where \(t_0\) is the observed value of \(n_{11}\). This is a *hyper-geometric* distribution.
V.2 P-values of Fisher’s exact tests:

\[
\begin{array}{ccc}
\text{Y} & & \\
1 & 2 & \\
X & 1 & n_{11} & n_{12} & n_{1+} \\
& 2 & n_{21} & n_{22} & n_{1+} \\
& n_{+1} & n_{+2} & n \\
\end{array}
\]

• Simple algebra shows

\[
\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{n_{11}(n_{+2} - n_{1+} + n_{11})}{(n_{1+} - n_{11})(n_{+1} - n_{11})} \Rightarrow n_{11}
\]

\[\Rightarrow\text{ larger } \hat{\theta} \Leftrightarrow \text{ larger } n_{11}\]

\[\Rightarrow\text{ We should reject } H_0 \text{ in favor of } H_1 \text{ when } n_{11} \text{ is large.}\]

\[\Rightarrow\text{ P-value } = P[n_{11} \geq t_0 | n_{1+}, n_{+1}, H_0] \text{ – one-sided Fisher’s exact test.}\]
• For Fisher’s tea example, one-sided p-value is:

\[
P\text{-value} = P[n_{11} \geq 3| n_{1+}, n_{+1}, H_0] \\
= P[n_{11} = 3| n_{1+}, n_{+1}, H_0] + P[n_{11} = 4| n_{1+}, n_{+1}, H_0] \\
= \frac{\binom{4}{3}\binom{4}{1}}{\binom{8}{4}} + \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}} = 0.229 + 0.014 = 0.243
\]

Mid P-value \(= \frac{0.229}{2} + 0.014 = 0.129.\)

**Note**: In this example, \(n_{1+}, n_{+1}\) are naturally fixed.
• Two-sided Fisher’s exact test: $H_0 : \theta = 1(X \perp Y)$ v.s. two-sided alternative $H_a : \theta \neq 1$.

<table>
<thead>
<tr>
<th>Table</th>
<th>$n_{11} = 0$</th>
<th>$n_{11} = 1$</th>
<th>$n_{11} = 2$</th>
<th>$n_{11} = 3$</th>
<th>$n_{11} = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob</td>
<td>0.014</td>
<td>0.229</td>
<td>0.514</td>
<td>0.229</td>
<td>0.014</td>
</tr>
</tbody>
</table>

• P-value of two-sided Fisher’s exact test:

\[
P\text{-value} = \sum P(n_{11})I\{P(n_{11}) \leq P(t_0)\}
\]
\[
= \text{sum of table probs that are } \leq \text{ observed table prob.}
\]

\[
p\text{-value} = P[n_{11} = 0] + P[n_{11} = 1] + P[n_{11} = 3] + P[n_{11} = 4]
\]
\[
= 0.014 + 0.229 + 0.229 + 0.014 = 0.486.
\]
- SAS program & output for Fisher’s exact test:

```sas
data table2_8;
  input pour $ guess $ count @@;
datalines;
milk milk 3   milk tea 1
  tea milk 1   tea tea 3
;

title "Analysis of Fisher’s tea data";
proc freq data=table2_8;
  weight count;
  tables pour*guess / norow nocol nopercent chisq;
  exact fisher or;
run;
```

The FREQ Procedure

Table of pour by guess

<table>
<thead>
<tr>
<th>pour</th>
<th>guess</th>
<th>Frequency</th>
<th>milk</th>
<th>tea</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>milk</td>
<td>milk 3</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>tea</td>
<td>tea 1</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>4</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Statistics for Table of pour by guess

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>2.0000</td>
<td>0.1573</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>1</td>
<td>2.0930</td>
<td>0.1480</td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell (1,1) Frequency (F)</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left-sided Pr &lt;= F</td>
<td>0.9857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right-sided Pr &gt;= F</td>
<td>0.2429</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table Probability (P)</td>
<td>0.2286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two-sided Pr &lt;= P</td>
<td>0.4857</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>9.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymptotic Conf Limits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.3666</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>220.9270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exact Conf Limits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.2117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>626.2435</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample Size = 8

**Note:** We can also obtain an exact CI for the true $\theta$. 
V.3 Fisher’s exact tests can be conservative

- For the Fisher’s tea example, the exact null distribution of
\[ n_{11} | n_{1+}, n_{+1} : \]

<table>
<thead>
<tr>
<th>Table</th>
<th>( n_{11} = 0 )</th>
<th>( n_{11} = 1 )</th>
<th>( n_{11} = 2 )</th>
<th>( n_{11}=3 )</th>
<th>( n_{11} = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob</td>
<td>0.014</td>
<td>0.229</td>
<td>0.514</td>
<td>0.229</td>
<td>0.014</td>
</tr>
</tbody>
</table>

- If we would like to construct a one-sided test at significance level 0.05 (target type I error prob), then we would only reject \( H_0 : \theta = 1 \) in favor of \( H_a : \theta > 1 \) when \( n_{11} = 4 \). Therefore, the actual type I error prob is

\[
P[n_{11} = 4 | H_0, n_{1+}, n_{+1}] = 0.014 < 0.05.
\]

So the test is very conservative!
VI Association in Three-Way Tables

- $X$, $Y$ – 2 categorical variables

The $X$, $Y$ (marginal) association may not reflect a *Causal relation*. Need to adjust a 3rd variable $Z$, *confounding variable* (related to both $X$, $Y$)

For example,

\[
\begin{align*}
X &= \text{second hand smoking} \\
Y &= \text{lung cancer} \\
Z &= \text{age, may be related to } X \text{ and } Y
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>Lung Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Second Hand Smoking</td>
<td>$\pi_{11}$</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>$\pi_{21}$</td>
</tr>
</tbody>
</table>

Slide 110
VI.1 Partial tables, conditional and marginal associations

- With 3 categorical variables $X, Y$ and $Z$, at each level of $Z$, there is an $XY$ table. Together, they form partial tables.

- Each partial table provides information on conditional associations between $X$ and $Y$ given $Z = k$.

- When collapsing partial tables over $Z$, we get a 2-way $XY$ (marginal) table. This table provides information of marginal association between $X$ and $Y$.

- We need to be aware that the conditional associations and marginal association may be different!
• Death penalty example (Table 2.10). Data from Florida, 1976-1987.

\[ X = \text{defendant’s’ race (W, B)}, Y = \text{death penalty (Yes, No).} \]

\[
\begin{array}{c|cc}
Y – \text{Death Penalty} & \text{Yes} & \text{No} \\
\hline
X – \text{Race} & W & 53 & 430 \\
& B & 15 & 176 \\
\end{array}
\]

Death penalty rate for \( W = \hat{\pi}_1 = \frac{53}{53+430} = 0.11 \)

Death penalty rate for \( B = \hat{\pi}_2 = \frac{15}{15+176} = 0.079 \)

\[ \hat{\psi} = 1.39, \quad \hat{\theta} = \frac{53 \times 176}{430 \times 15} = 1.45 \]

\[ \Rightarrow \text{White defendants are (40%) more likely to receive a death penalty than black defendants.} \]

• Maybe the race of victims (Z) affects the XY association?
When $Z =$ White, $XY$ table is

<table>
<thead>
<tr>
<th></th>
<th>Y – Death Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>$X$ – Race</td>
<td>W</td>
</tr>
<tr>
<td>Yes</td>
<td>53</td>
</tr>
<tr>
<td>No</td>
<td>414</td>
</tr>
</tbody>
</table>

When $Z =$ Black, $XY$ table is

<table>
<thead>
<tr>
<th></th>
<th>Y – Death Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>$X$ – Race</td>
<td>W</td>
</tr>
<tr>
<td>Yes</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>16</td>
</tr>
</tbody>
</table>

- We see that the conditional associations and the marginal association between $X$ and $Y$ have different directions! This phenomenon is called *Simpson’s paradox*. 

  Slide 113
• Reasons causing *Simpson’s paradox*:

  - $Z$ is related to both $X$ and $Y$.
  1. More white victims than black victims.
  2. Given $Z = \text{white}$, defendants ($X$) are about 90% likely to be white.
  3. Given $Z = \text{black}$, defendants ($X$) are only about 10% likely to be white.
  4. More white defendants received death penalty ($X, Y$ are related).
VI.2 Conditional and marginal odds ratios

- When we have $2 \times 2 \times K$ tables for $X, Y$ and $Z$, At $Z = k$, observed table for $XY$ is

\[
\begin{array}{c|cc}
Y \\
\hline
1 & n_{11k} & n_{12k} \\
2 & n_{21k} & n_{22k} \\
\end{array}
\]

Then we have $K$ conditional odds ratios that estimate the conditional associations between $X$ and $Y$ at $Z = k$

\[
\hat{\theta}_{XY}(k) = \frac{n_{11k}n_{22k}}{n_{12k}n_{21k}}.
\]
The marginal $XY$ table is

$$
\begin{array}{c|cc}
\text{Y} & 1 & 2 \\
\hline
1 & n_{11+} & n_{12+} \\
2 & n_{21+} & n_{22+} \\
\end{array}
$$

The marginal odds-ratio estimates the marginal association between $X$ and $Y$:

$$
\hat{\theta}_{XY} = \frac{n_{11+}n_{22+}}{n_{12+}n_{21+}}.
$$
- For the death penalty example,

\[
\hat{\theta}_{XY} = 1.45 \\
\hat{\theta}_{XY(1)} = \frac{53 \times 37}{11 \times 414} = 0.43 \\
\hat{\theta}_{XY(2)} = \frac{0 \times 139}{4 \times 16} = 0 \\
\hat{\theta}_{XY(2)}^{\text{mod}} = \frac{0.5 \times 139.5}{4.5 \times 16.5} = 0.94
\]
VI.3 Conditional and marginal independence

- If $X$ and $Y$ are independent at any level of $Z$, then $X$ and $Y$ are called *conditionally independent* given $Z$.

  If $X, Y$ are 2-level variables, then $X$ and $Y$ *conditionally independent* \iff $\theta_{XY(k)} = 1$, $k = 1, 2, ..., K$.

- $X, Y$ marginally independent if $X, Y$ are independent.

  If $X, Y$ are 2-level variables, then $X$ and $Y$ *marginally independent* \iff $\theta_{XY} = 1$. 
• Example: Conditional independence $\iff$ marginal independence.

\[
\begin{array}{cc}
  Y \\
  S & F \\
  X & A & 18 & 12 \\
  & B & 12 & 8 \\
\end{array}
\quad\hat{\theta}_{XY(1)} = 1 \quad A = B
\]

\[
\begin{array}{cc}
  Y \\
  S & F \\
  X & A & 2 & 8 \\
  & B & 8 & 32 \\
\end{array}
\quad\hat{\theta}_{XY(2)} = 1 \quad A = B
\]

Marginally,

\[
\begin{array}{cc}
  Y \\
  S & F \\
  X & A & 20 & 20 \\
  & B & 20 & 40 \\
\end{array}
\quad\hat{\theta}_{XY} = 2 \quad \Rightarrow A > B
\]
- Example: Marginal independence $\not\Rightarrow$ conditional independence

\[
\begin{array}{c|cc}
X & S & F \\
\hline
A & 4 & 1 \\
B & 9 & 6 \\
\end{array}
\quad \hat{\theta}_{XY}(1) = \frac{8}{3}
\quad \begin{array}{c|cc}
X & S & F \\
\hline
A & 6 & 9 \\
B & 1 & 4 \\
\end{array}
\quad \hat{\theta}_{XY}(2) = \frac{8}{3}
\]

Marginally,

\[
\begin{array}{c|cc}
X & S & F \\
\hline
A & 10 & 10 \\
B & 10 & 10 \\
\end{array}
\quad \hat{\theta}_{XY} = 1 \quad \Rightarrow A = B
\]
VI.4 Homogeneous association

- Assume $X, Y$ are 2-level variables.

*Homogeneous association* (in terms of $\theta$) – no interaction

\[ \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)} \]

When $\theta_{XY(k)}$ are not all the same, $Z$ is called an effect modifier (there is interaction).

- **Note:** Under *homogeneous association*, we cannot claim

\[ \theta_{XY} = \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)}. \]

See previous examples.
3 Generalized Linear Models (GLMs)

0 Introduction

- In a simple linear regression model for continuous $Y$:

$$Y = \alpha + \beta x + \varepsilon,$$

usually $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$.

$Y$ = response

$x$ = (numeric) covariate, indep or explanatory variable

$\beta = E(Y|x + 1) - E(Y|x)$

$2\beta = E(Y|x + 2) - E(Y|x)$, etc.

$\beta$ catches the linear relationship between $X$ and $Y$.

When $\beta = 0$, there is no linear relationship between $X$ and $Y$. 
• Given data \((x_i, y_i), i = 1, 2, \cdots, n\), we can estimate \(\alpha, \beta\), and hence \(E(Y|x)\). A common method to estimate \(\alpha, \beta\) is least squares (LS) by minimizing the following sum of squares (SS)

\[
\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2.
\]

• Minimizing \(\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \Rightarrow \)

\[
\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2},
\]

\[
\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}
\]

where \(\bar{x}\) is the sample mean of \(\{x_i\}'s, \bar{y}\) is the sample mean of \(\{y_i\}'s.

• \(\hat{\alpha}, \hat{\beta}\) have good statistical properties.

• Normality is Not required for the LS estimation.
A Description of the Linear Model

The graph shows a linear relationship between the variables x and y. The data points are plotted on a Cartesian plane with x values ranging from 0.2 to 0.8 and y values ranging from 1 to 4. A line of best fit is also plotted through the data points, indicating a positive linear relationship.

Slide 124
• Under $\varepsilon \overset{iid}{\sim} N(0, \sigma^2)$ (so $Y$ is also normal), the above model can be re-written as

$$Y|x \overset{ind}{\sim} N(\alpha + \beta x, \sigma^2),$$

or equivalently

$$Y|x \overset{ind}{\sim} N(\mu(x), \sigma^2), \quad \mu(x) = \alpha + \beta x$$

• MLE of $(\alpha, \beta) = \text{LSE of } (\alpha, \beta)$.

• Simple linear regression model can be extended to more than 1 covariate:

$$Y|x \overset{ind}{\sim} N(\mu(x), \sigma^2)$$

$$\mu(x) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$  

$\beta_k$: average change in $Y$ with one unit increase in $x_k$ while holding other covariates fixed (if $x_k$’s are unrelated variables)

• The above model can be easily extended to non-normal data $Y$. 

---

*Slide 125*
Three Components of a GLM

- Data: \((x_i, y_i), i = 1, 2, \ldots, n\)
  
  \[ y_i = \text{response} \]
  \[ x_i = (x_{1i}, x_{2i}, \ldots, x_{pi}) \text{ covariate, indep or explanatory variable} \]

- A GLM has 3 components: random component, systematic component and the link function.

**I.1 Random component**

- Response \(Y\) is the random component of a GLM. We need to specify a distribution for \(Y\), such as normal, Bernoulli/Binomial or Poisson. For the normal GLM, we specify the normal distribution for \(Y\).
I.2 Systematic component

- For covariates $x_1, x_2, \cdots, x_p$, form linear combination:

$$\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$ 

This linear combination is called the *systematic component* of a GLM. In a regression setting, the covariate values are viewed as fixed, hence the name of *systematic component*.

**Note:** we allow interactions such as $x_3 = x_1 x_2$, power functions such as $x_2 = x_1^2$ and other transformation for the covariates (*e.g.*, $x_2 = e^{x_1}$). In this case, we have to be careful in interpreting $\beta_k$’s.
I.3 Link function

- Denote $\mu = \mathbb{E}(Y|x)$.

- With a smooth and monotone function $g(\mu)$, we relate $\mu$ and the *systematic component* via the formula:

$$g(\mu) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$  

This function $g(\mu)$ is called the *link* function of a GLM.

- **Note:** Since both $\mu$ and the *systematic component* are both fixed quantities, there is *NO* error term in the above formula!

- Obviously, a normal GLM assumes

$$g(\mu) = \mu.$$  

This link function is called the *identity* link.
• **Note**: In modelling the relationship between continuous response $Y$ and a covariate $x$, often time we would try to apply a transformation function $g(\cdot)$ to $Y$ so that $g(Y)$ may have a distribution closer to *normal* (even though normality is not necessary) and then fit

$$g(Y) = \alpha + \beta x + \epsilon.$$

This is a *transformation model*.

A GLM with link function $g(\mu)$ ($\mu = \text{E}(Y|x)$)

$$g(\mu) = \alpha + \beta x$$

is **NOT** the same as the above transformation model, and we don’t apply the link function to the response $Y$!

Will see more later ...
### Fitting and inference of a GLM

- Since we specify the distribution of $Y$, given data we use Maximum Likelihood (instead of Least squares) approach for estimation and inference on effect parameters $\beta_1, \cdots, \beta_p$.

- There is a unified algorithm for estimation and inference.

- Using Proc Genmod of SAS, we get the estimate, SE and p-value for testing $H_0 : \beta_k = 0$, etc.

```plaintext
proc genmod data=; * if y=1/0, then we need "descending" here;
  model y = x / dist= link=;
run;
```

The default distribution is normal with identity link. Common distributions are:

<table>
<thead>
<tr>
<th>Dist=</th>
<th>Distribution</th>
<th>Default Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial</td>
<td>Bin</td>
<td>B</td>
</tr>
<tr>
<td>Gamma</td>
<td>Gam</td>
<td>G</td>
</tr>
<tr>
<td>NegBin</td>
<td>NB</td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>Nor</td>
<td>N</td>
</tr>
<tr>
<td>Poisson</td>
<td>Poi</td>
<td>P</td>
</tr>
</tbody>
</table>
If $y$ is binary (1/0) with 1 being the success (that is, we would like to model $P[Y = 1]$), we should use descending option in Proc Genmod.

For binomial response $y$ (of course, we should have $n$ - # of Bernoulli trials to get $y$), we have to use:

```sas
proc genmod data=;
  model y/n = x / dist=bin link=;
run;
```

Note: $y$ and $n$ are two variables in the data set. We don’t define a new variable $p = y/n$ and use “model $p = x$”. The / in $y/n$ is just a symbol.

- Data is organized in the same way as for Proc Reg of SAS.
II GLMs for Binary Response $Y$

- When the response $Y$ is binary (1/0, 1=success, 0=failure):

\[ \mu = \mathbb{E}(Y) = 1 \times P[Y = 1] + 0 \times P[Y = 0] = P[Y = 1] = \pi \]

is the success probability.

- A GLM for binary $Y$ with link function $g(\cdot)$ relates $\pi$ to the systematic component in the following:

\[ g(\pi) = \alpha + \beta x. \]

- Different choice of the link function $g(\pi)$ leads to a different binary GLM.
II.1 Linear probability model

- If we choose the link function $g(\cdot)$ to be the identity link $g(\pi) = \pi$, then we have a linear probability model:

  $$\pi = \alpha + \beta x.$$ 

- Linear probability model is reasonable only if $\alpha + \beta x$ yields values in $(0,1)$ for valid values of $x$.

- $\beta$ has a nice interpretation:

  $$\beta = \pi(x + 1) - \pi(x)$$

  risk difference when $x$ increases by one unit.

- When the linear probability fits the data well, we can also use LS to make inference on $\beta$. The LS & ML estimation and inference will be similar.

  Testing $H_0 : \beta = 0$ under this model is basically the same as the Cochran-Armitage trend test.
• Inference for the risk difference in a $2 \times 2$ table can be achieved using the linear probability model:

\[
\begin{array}{c|cc|c}
X & 1 & 0 & \\
\hline
1 & y_1 & n_1 - y_1 & n_1 \\
0 & y_2 & n_2 - y_2 & n_2 \\
\end{array}
\]

Let $\pi_1 = P[Y = 1|x = 1]$, $\pi_0 = P[Y = 1|x = 0]$, and we would like to make inference in $\phi = \pi_1 - \pi_0$, the risk difference between row 1 ($X = 1$) and row 2 ($X = 0$).

We can fit the following linear probability model to the above table

\[
\pi = \alpha + \beta x.
\]

Then $\beta$ is the same as $\phi$. 
• SAS program for making inference on risk difference for a $2 \times 2$ table:

```
data main;
  input x y n;
  1 * *;
  0 * *;
;
proc genmod;
  model y/n = x / dist=bin link=identity;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
- Snoring and Heart Disease Example (Table 3.1 on p. 69)

<table>
<thead>
<tr>
<th>Snoring</th>
<th>Heart Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
</tr>
<tr>
<td>0 Never</td>
<td>0</td>
</tr>
<tr>
<td>Snoring</td>
<td>2 Occasionally</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>4</td>
</tr>
<tr>
<td>5 Every night</td>
<td>5</td>
</tr>
</tbody>
</table>

- After assigning scores \(x_i: 0, 2, 4, 5\) to *snoring*, we can calculate the sample proportions \(p_i\) for each snoring level and plot \(p_i\) against \(x_i\) to see if linear probability model is reasonable.
SAS program and Part of its output:

```sas
data table3_1;
  input snoring score y y0;
  n = y+y0;
  p = y/n;
  logitp = log(p/(1-p));
data lines;
  0 0 24 1355
  1 2 35 603
  2 4 21 192
  3 5 30 224;

title "Snoring and heart disease data using class variable with identity link";
proc genmod;
  class snoring;
  model y/n = snoring / dist=bin link=identity noint;
  estimate "level 1 - level 0" snoring -1 1 0 0;
  estimate "level 2 - level 1" snoring 0 -1 1 0;
  estimate "level 3 - level 2" snoring 0 0 -1 1;
run;

title "Sample proportion vs score";
proc plot;
  plot p*score;
run;

title "Sample logit vs score";
proc plot;
  plot logitp*score;
run;
```
### The GENMOD Procedure

#### Contrast Estimate Results

<table>
<thead>
<tr>
<th>Label</th>
<th>Mean Estimate</th>
<th>Mean Confidence Limits</th>
<th>L’Beta Estimate</th>
<th>Standard Error</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>level 1 - level 0</td>
<td>0.0375</td>
<td>0.0185</td>
<td>0.0564</td>
<td>0.0375</td>
<td>0.0097</td>
</tr>
<tr>
<td>level 2 - level 1</td>
<td>0.0437</td>
<td>-0.0000</td>
<td>0.0875</td>
<td>0.0437</td>
<td>0.0223</td>
</tr>
<tr>
<td>level 3 - level 2</td>
<td>0.0195</td>
<td>-0.0369</td>
<td>0.0759</td>
<td>0.0195</td>
<td>0.0288</td>
</tr>
</tbody>
</table>

Sample proportion vs score

Plot of p*score. Legend: A = 1 obs, B = 2 obs, etc.
• The plots indicates linear probability model with the chosen scores for snoring may fit the data well (good choice of snoring scores).

• Consider linear probability model:

$$\pi = \alpha + \beta x,$$

where $x$ is the snoring score.

• SAS program:

```sas
title "Snoring and heart disease data using score with identity link";
proc genmod;
  model y/n = score / dist=bin link=identity;
run;
```
SAS output:

```
**************************************************************************
Snoring and heart disease data using score with identity link

The GENMOD Procedure

Model Information

Data Set                  WORK.TABLE3_1
Distribution              Binomial
Link Function             Identity
Response Variable (Events) y
Response Variable (Trials) n

Number of Observations Read 4
Number of Observations Used 4
Number of Events           110
Number of Trials           2484

Response Profile

<table>
<thead>
<tr>
<th>Value</th>
<th>Binary Outcome</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Event</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>Nonevent</td>
<td>2374</td>
</tr>
</tbody>
</table>
```
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>2</td>
<td>0.0692</td>
<td>0.0346</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>2</td>
<td>0.0692</td>
<td>0.0346</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>2</td>
<td>0.0688</td>
<td>0.0344</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>2</td>
<td>0.0688</td>
<td>0.0344</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>-417.4960</td>
<td>0.0344</td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-10.1609</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td></td>
<td>24.3217</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td></td>
<td>36.3217</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td></td>
<td>23.0943</td>
<td></td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.0172</td>
<td>0.0034</td>
<td>0.0105 - 0.0240</td>
<td>25.18</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.0198</td>
<td>0.0028</td>
<td>0.0143 - 0.0253</td>
<td>49.97</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

- The fitted model is

\[ \hat{\pi} = 0.017 + 0.0198x, x = 0, 2, 4, 5 \]
From the fitted model, we can calculate the estimated heart disease probability for each level of snoring:

<table>
<thead>
<tr>
<th>Snoring ($x$)</th>
<th>Heart Disease</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes ($y_i$)</td>
<td>No</td>
</tr>
<tr>
<td>0 Never</td>
<td>24</td>
<td>1355</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>35</td>
<td>605</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>21</td>
<td>192</td>
</tr>
<tr>
<td>5 Every night</td>
<td>30</td>
<td>224</td>
</tr>
</tbody>
</table>

Since the fitted values $\hat{\pi} \approx p_i$, the linear probability model fits the data well.

The model has a nice interpretation: For non-snorers, the heart disease prob is 0.017 (the intercept).
For occasional snorers, the HD prob increases 0.04 (more than double), etc.
• **Note**: We can recover the original binary data (1/0 – called \(hd\) in the new data set) with 1 for heart disease, and use the following program to get exactly the same results:

```plaintext
title "Snoring and binary heart disease in proc genmod";
proc genmod descending;
model hd = score / dist=bin link=identity;
run;
```

---

### Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.0172</td>
<td>0.0034</td>
<td>0.0105 0.0240</td>
<td>25.18</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.0198</td>
<td>0.0028</td>
<td>0.0143 0.0253</td>
<td>49.97</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

Without the option `descending`, Proc Genmod models

\[
P[Y = 0] = 1 - \pi:
\]

\[
1 - \pi = 1 - \alpha - \beta x.
\]

Therefore, if we don’t use the option descending, the intercept estimate will be equal to \(1 - 0.0172 = 0.9828\), and the estimate for the coefficient of snoring score \(x\) will be -0.0198.
• We can also fit a linear regression model to the binary data and will get similar results.

```sas
title "Snoring and binary heart disease with LS approach";
proc reg;
   model hd = score;
run;
```

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| |
|----------|----|-------------------|----------------|---------|------|-----------------|
| Intercept| 1  | 0.01687           | 0.00516        | 3.27    | 0.0011 |
| score    | 1  | 0.02004           | 0.00232        | 8.65    | <.0001 |

**Note:** Since `proc reg` models $E(Y) = \pi$, the above results should be similar to the linear prob model with the option `descending` (if binary response data is used).
II.2 Logistic regression model

- For binary response $Y$, if we take the link function $g(\pi)$ in the GLM as

$$g(\pi) = \text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right),$$

then we have a logistic regression model:

$$\text{logit}(\pi) = \alpha + \beta x.$$ 

Here the function $g(\pi) = \text{logit}(\pi) = \log\{\pi/(1 - \pi)\} = \log(\text{odds})$ is called the logit function of $\pi$. Note that with this link, any $x$ and $\alpha, \beta$ will yield a valid $\pi$:

$$\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}.$$ 

- With a fitted logistic regression, the estimated prob at $x$ is given by

$$\hat{\pi}(x) = \frac{e^{\hat{\alpha} + \hat{\beta} x}}{1 + e^{\hat{\alpha} + \hat{\beta} x}}.$$
Figure 3.2. Logistic regression functions.
• Interpretation of $\beta$:

$\pi$ at $x$ : 

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

$\pi$ at $x + 1$ :

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

$$\log \frac{\pi(x + 1)}{1 - \pi(x + 1)} - \log \frac{\pi(x)}{1 - \pi(x)} = \beta$$

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

$$e^\beta = \frac{\pi(x + 1)/\{1 - \pi(x + 1)\}}{\pi(x)/\{1 - \pi(x)\}}$$

odds-ratio with one unit increase in $x$

$\Rightarrow$  

$$2\beta = \log \left\{ \frac{\pi(x + 2)/\{1 - \pi(x + 2)\}}{\pi(x)/\{1 - \pi(x)\}} \right\}$$

log odds-ratio with two unit increase in $x$, etc.
Inference for the odds-ratio in a $2 \times 2$ table can be achieved using the logistic regression model:

\[
\begin{array}{c|c|c|c}
\text{Y} & 1 & 0 \\
\hline
\text{X} & 1 & y_1 & n_1 - y_1 & n_1 \\
0 & y_2 & n_2 - y_2 & n_2 \\
\end{array}
\]

Let $\pi_1 = P[Y = 1|x = 1]$, $\pi_0 = P[Y = 1|x = 0]$, and we would like to make inference on $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$, the odds-ratio between row 1 and row 2.

We can fit the following logistic regression model:

\[
\text{logit}(\pi) = \alpha + \beta x.
\]

Since $x$ can only take 0 and 1, $e^\beta = \theta$ is the odds-ratio of interest. Testing $H_0 : \beta = 0 \Leftrightarrow H_0 : X \perp Y$. 

---

Slide 148
• SAS program for making inference on *odds ratio* for a $2 \times 2$ table:

```sas
data main;
  input x y n;
  1 * * 0 * *
;
proc genmod;
  model y/n = x / dist=bin link=logit;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood Parameter Estimates

Parameter  DF  Estimate  Standard  Wald 95% Confidence  Wald
           Error             Limits  Chi-Square
Intercept 1  *  *  *  *  *  *  *
X          1  *  *  *  *  *  *
Scale      0  1.0000  0.0000  1.0000  1.0000  *  *
```
Logistic regression model for Snoring and Heart Disease Example.

If there is a nearly straight line in the plot of sample logit against $x$ indicates a good fit of the logistic regression:

$$
\text{sample logit} = \log \frac{p_i}{1 - p_i}.
$$

Sample logit vs score

Plot of logitp*score. Legend: A = 1 obs, B = 2 obs, etc.
title "Snoring and heart disease data using score with logit link";
proc genmod;
  model y/n = score / dist=bin link=logit;
run;

**************************************************************************
Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.8662</td>
<td>0.1662</td>
<td>-4.1920 -3.5405</td>
<td>541.06</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.3973</td>
<td>0.0500</td>
<td>0.2993 0.4954</td>
<td>63.12</td>
</tr>
</tbody>
</table>

• Comparison of estimated probs:

<table>
<thead>
<tr>
<th>Snoring($x$)</th>
<th>Heart Disease</th>
<th>Linear</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes ($y_i$)</td>
<td>No</td>
<td>$n_i$</td>
</tr>
<tr>
<td>0 Never</td>
<td>24</td>
<td>1355</td>
<td>1379</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>35</td>
<td>605</td>
<td>638</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>21</td>
<td>192</td>
<td>213</td>
</tr>
<tr>
<td>5 Every night</td>
<td>30</td>
<td>224</td>
<td>254</td>
</tr>
</tbody>
</table>

⇒ Linear prob model is better than the logistic model.
• We can also use the original binary response hd and use the following SAS program with descending option and will get the same results.

```sas
title "Snoring and heart disease data using score with logit link";
proc genmod descending;
  model hd = score / dist=bin link=logit;
run;
```

**************************************************************************

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.8662</td>
<td>0.1662</td>
<td>-4.1920 -3.5405</td>
<td>541.06</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.3973</td>
<td>0.0500</td>
<td>0.2993 0.4954</td>
<td>63.12</td>
</tr>
</tbody>
</table>

• Note: if we don’t use the option descending, then we are modeling $P[Y = 0] = 1 - \pi = \tau$. If the original logistic model for $\pi$ is true, then we also have a logistic model for $\tau$:

$$
\log \left( \frac{\tau}{1 - \tau} \right) = \log \left( \frac{1 - \pi}{\pi} \right) = -\log \left( \frac{\pi}{1 - \pi} \right) = -\alpha - \beta x.
$$

Therefore, all estimates will be the mirror image of those from the previous logistic model.
II.3 Log linear probability model

- For binary response $Y$, if we take the link function $g(\pi)$ in the GLM as the log function, then we have a log-linear probability model:

$$\log(\pi) = \alpha + \beta x.$$ 

- Given $x$ and $\alpha, \beta$, solving for $\pi$ we have:

$$\pi = e^{\alpha + \beta x}.$$ 

Of course, the model is only reasonable if the model produces valid $\pi$’s in (0,1) for $x$ in the valid range.
- Interpretation of $\beta$:

$$\log \pi(x) = \alpha + \beta x$$

$$\log \pi(x + 1) = \alpha + \beta(x + 1)$$

$$\log \pi(x + 1) - \log \pi(x) = \beta$$

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

$$e^\beta = \frac{\pi(x + 1)}{\pi(x)}$$

RR with one unit increase in $x$

$$\Rightarrow e^{2\beta} = \frac{\pi(x + 2)}{\pi(x)}$$

RR with two unit increase in $x$
Inference for the RR in a $2 \times 2$ table can be achieved using the log-linear probability model:

$$
\begin{array}{c|cc|c}
Y & 1 & 0 \\
\hline
X & 1 & y_1 & n_1 \\
& 0 & y_2 & n_2 \\
\hline
\end{array}
$$

Let $\pi_1 = P[Y = 1|x = 1]$, $\pi_0 = P[Y = 1|x = 0]$, and we would like to make inference on $RR = \frac{\pi_1}{\pi_0}$, the relative risk between row 1 and row 2. We can fit the following log-linear probability model:

$$
\log(\pi) = \alpha + \beta x.
$$

Since $x$ can only take 0 and 1, $e^\beta$ is the RR of interest.

Testing $H_0 : \beta = 0 \Leftrightarrow H_0 : X \perp Y$. 
• SAS program for making inference on *relative risk* for a $2 \times 2$ table:

```sas
data main;
  input x y n;
  1 * *
  0 * *
;
proc genmod;
  model y/n = x / dist=bin link=log;
run;
```

• Output would look like:

```
Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
```
II.4 Probit regression model

- For binary response $Y$, if we take the link function in the GLM as $g(\pi) = \Phi^{-1}(\pi)$, the inverse of the cumulative distribution function (cdf) of N(0,1), then we have a probit regression model

$$\Phi^{-1}(\pi) = \alpha + \beta x.$$ 

- For any $x$ and $\alpha, \beta$, the model yields valid $\pi$:

$$\pi = \Phi(\alpha + \beta x).$$

- A probit model is very similar to a logistic regression. That is, if

$$\Phi^{-1}\{\pi(x)\} = \alpha + \beta x$$

is true, then

$$\text{logit}\{\pi(x)\} \approx \alpha^* + \beta^* x$$

with $\alpha^* = 1.7\alpha$ and $\beta^* = 1.7\beta$. However, the fitted probs from these 2 models will be similar.
For the Snoring/Heart Disease example, the fitted results:

```plaintext
title "Snoring and heart disease data using score with probit link";
proc genmod;
  model y/n = score / dist=bin link=probit;
run;
```

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.0606</td>
<td>0.0704</td>
<td>-2.1986 -1.9225</td>
<td>855.49</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.1878</td>
<td>0.0236</td>
<td>0.1415 0.2341</td>
<td>63.14</td>
</tr>
</tbody>
</table>

\( \hat{\pi}(x) = \Phi(-2.0606 + 0.1878x) \).

For example, when \( x = 2 \) (occasional snorers), \( \hat{\pi}(x) \) is:

\[
\hat{\pi}(2) = \Phi(-2.0606 + 0.1878 \times 2) = \Phi(-1.685) = P[Z \leq -1.685] = 0.046.
\]

**Note:** \( 1.7 \times (-2.0606) = -3.5, 1.7 \times 0.1878 = 0.32 \), very close to the estimates from the logistic model.
We can also use the original binary response $hd$ and use the following SAS program with descending option and will get the same results.

```
title "Snoring and heart disease data using score with logit link";
proc genmod descending;
    model hd = score / dist=bin link=probit;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.0606</td>
<td>0.0704</td>
<td>-2.1986 to -1.9225</td>
<td>855.49</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>0.1878</td>
<td>0.0236</td>
<td>0.1415 to 0.2341</td>
<td>63.14</td>
</tr>
</tbody>
</table>

**Note**: if we don’t use the descending option, then we are modeling $P[Y = 0] = 1 - \pi = \tau$. If the original probit model for $\pi$ is true, then we also have a probit model for $\tau$:

$$
\Phi^{-1}(\tau) = \Phi^{-1}(1 - \pi) = -\Phi^{-1}(\pi) = -\alpha - \beta x.
$$

Therefore, all estimates will be the mirror image of those from the previous probit model.
- Comparison of estimated probs from 3 models:

<table>
<thead>
<tr>
<th>Snoring (x)</th>
<th>Heart Disease</th>
<th>Yes (y_i)</th>
<th>No</th>
<th>(n_i)</th>
<th>(p_i)</th>
<th>Linear Fit</th>
<th>Logit Fit</th>
<th>Probit Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Never</td>
<td></td>
<td>24</td>
<td>1355</td>
<td>1379</td>
<td>0.017</td>
<td>0.017</td>
<td>0.021</td>
<td>0.020</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td></td>
<td>35</td>
<td>605</td>
<td>638</td>
<td>0.055</td>
<td>0.057</td>
<td>0.044</td>
<td>0.046</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td></td>
<td>21</td>
<td>192</td>
<td>213</td>
<td>0.099</td>
<td>0.096</td>
<td>0.099</td>
<td>0.095</td>
</tr>
<tr>
<td>5 Every night</td>
<td></td>
<td>30</td>
<td>224</td>
<td>254</td>
<td>0.118</td>
<td>0.116</td>
<td>0.132</td>
<td>0.131</td>
</tr>
</tbody>
</table>

⇒

1. Logistic model and probit model give very close predicted \(\pi\)'s.
2. Linear prob model is better than the logistic model.
Sample proportions and fitted π’s from 3 models

Figure 3.1. Fit of models for snoring and heart disease data.
III GLMs for Count Data

- In many applications, the response $Y$ is *count* data:
  1. Monthly # of car accidents on a particular highway.
  2. Yearly # of new cases of certain disease in counties over US, etc.

- For count data $Y$, a common distributional assumption is $Y \sim \text{Poisson} (\mu)$:
  \[ E(Y) = \text{var}(Y) = \mu. \]

- A GLM for count data $Y$ usually uses log as the link function:
  \[ \log(\mu) = \alpha + \beta x. \]

  \[ \Rightarrow \mu(x) = e^{\alpha + \beta x}. \]

  Of course, other link functions, such as identity link, are also possible.

- Interpretation of $\beta$:
  \[ e^\beta = \frac{\mu(x + 1)}{\mu(x)}, e^\beta - 1 = \text{percentage increase in } \mu \text{ with 1 unit increase in } x \]
III.1 Example: Female horseshoe crabs and their satellites (Table 3.2, page 76-77)
• Data (a subset):

```r
data crab;
input color spine width satell weight;
   weight=weight/1000; color=color-1;
datalines;
  3 3 28.3 8 3050
  4 3 22.5 0 1550
  2 1 26.0 9 2300
  4 3 24.8 0 2100
  4 3 26.0 4 2600
  3 3 23.8 0 2100
  2 1 26.5 0 2350
  4 2 24.7 0 1900
.
.
.
```

\[ y_i = \# \text{ of satellites (male crabs) for female crab } i \]

\[ x_i = \text{ carapace width of female crab } i \]

• Model the relationship between \( \mu_i = E(Y_i|x_i) \) and \( x_i \) using the log-linear model

\[ \log(\mu_i) = \alpha + \beta x_i \]

assuming \( Y_i \sim \text{Poisson}(\mu_i) \).
SAS Program and output:

```sas
title "Analysis of crab data using Poisson distribution";
title2 "(without overdispersion) with log link";
proc genmod data=crab;
  model satell = width / dist=poi link=log;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.3048</td>
<td>0.5422</td>
<td>-4.3675 -2.2420</td>
<td>37.14</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1640</td>
<td>0.0200</td>
<td>0.1249 0.2032</td>
<td>67.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{\mu}(x) = e^{-3.3048+0.1640x}. \]

\[ \hat{\beta} = 0.1640 \text{ with } \widehat{SE}(\hat{\beta}_1) = 0.02, \text{ p-value } < 0.0001. \]

However, the inference may not be valid since the count data \( Y \) often has an over-dispersion issue:

\[ \text{var}(Y) > \text{E}(Y). \]
III.2 Over-dispersion in count data

- Empirical check of over-dispersion:

<table>
<thead>
<tr>
<th>Carapace width ($x$)</th>
<th>Num. of Obs.</th>
<th>$\bar{y}$</th>
<th>$S^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\leq 23.25$</td>
<td>14</td>
<td>1</td>
<td>2.77</td>
</tr>
<tr>
<td>$23.25 - 24.25$</td>
<td>14</td>
<td>1.43</td>
<td>8.88</td>
</tr>
<tr>
<td>$24.25 - 25.25$</td>
<td>28</td>
<td>2.39</td>
<td>6.54</td>
</tr>
<tr>
<td>$25.25 - 26.25$</td>
<td>39</td>
<td>2.69</td>
<td>11.38</td>
</tr>
<tr>
<td>$26.25 - 27.25$</td>
<td>22</td>
<td>2.86</td>
<td>6.88</td>
</tr>
<tr>
<td>$27.25 - 28.25$</td>
<td>24</td>
<td>3.87</td>
<td>8.81</td>
</tr>
<tr>
<td>$28.25 - 29.25$</td>
<td>18</td>
<td>3.94</td>
<td>16.88</td>
</tr>
<tr>
<td>$&gt; 29.25$</td>
<td>14</td>
<td>5.14</td>
<td>8.29</td>
</tr>
</tbody>
</table>

Observation: $S^2 \gg \bar{y} \implies \text{var}(Y_i|x_i) > E(Y_i|x_i)$, over-dispersion!
A common approach to take into account over-dispersion in inference is to assume the following variance-mean relationship for count data $Y$:

$$\text{var}(Y) = \phi \text{E}(Y),$$

$\phi$— over-dispersion parameter.

- Estimation of $\phi$ using the *Pearson* statistic

$$\hat{\phi}_P = \frac{1}{df} \sum \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

This can be specified by `scale=pearson` or `scale=p` in Proc Genmod. A common choice.

- Estimation of $\phi$ using the Deviance statistic:

$$\hat{\phi}_D = \frac{2[\log(L_S) - \log(L_M)]}{df}$$

This can be specified by `scale=deviance` or `scale=d` in Proc Genmod.
• SAS program and output:

```
title "Analysis of crab data using overdispersed Poisson";
title2 "distribution with log link";
proc genmod data=crab;
   model satell = width / dist=poi link=log scale=pearson;
run;
```

******************************************************************************

Analysis Of Maximum Likelihood 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>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-3.3048</td>
<td>0.9673</td>
<td>-5.2006 -1.4089</td>
<td>11.67</td>
<td>0.0006</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1640</td>
<td>0.0356</td>
<td>0.0942 0.2339</td>
<td>21.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.7839</td>
<td>0.0000</td>
<td>1.7839 1.7839</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was estimated by the square root of Pearson Chi-Square/DOF.
• With the option `scale=pearson`, the Pearson estimate 
\[ \sqrt{\hat{\phi}_P} = 1.7839, \] indicating a lot of over-dispersion.

• From the output, we see that we got the same estimates of \( \alpha \) and \( \beta \). However, their standard errors are inflated by \( \sqrt{\hat{\phi}} = 1.7839 \) (larger SE’s).

• Based on the estimated model:

\[
\log(\mu) = -3.3048 + 0.1640x
\]

⇒ With 1cm increase in carapace width, the average # of satellites will increase by \( e^{0.1640} - 1 = 0.18 = 18\% \).
III.3 GLM for count data with other links

- Plot of smoothing of raw data indicates the identity link function:

![Figure 3.5. Smoothenings of horseshoe crab counts.](image)
• Consider the GLM with the identity link:

\[ \mu = \alpha + \beta x. \]

• SAS program and output:

```sas
title "Analysis of crab data using overdispersed Poisson";
title2 "distribution with identity link";
proc genmod data=crab;
    model satell = width / dist=poi link=identity scale=pearson;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-11.5321</td>
<td>2.6902</td>
<td>-16.8048</td>
<td>-6.2593</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.5495</td>
<td>0.1056</td>
<td>0.3425</td>
<td>0.7565</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.7811</td>
<td>0.0000</td>
<td>1.7811</td>
<td>1.7811</td>
</tr>
</tbody>
</table>

⇒

1. A lot of over-dispersion: \( \hat{\phi}_P^{1/2} = 1.7811 \).
2. Significant evidence against \( H_0 : \beta = 0 \).
3. Fitted model: \( \hat{\mu} = -11.5321 + 0.5495x \).
Comparison of GLMs with log and identity links

Figure 3.6. Estimated mean number of satellites for log and identity links.
III.4 Negative binomial for over-dispersed count data

- We can assume a negative-binomial distribution for count response $Y$ to automatically handle \textit{over-dispersion}:

\[
E(Y) = \mu, \quad \text{var}(Y) = \mu + D\mu^2,
\]

where $D > 0$ is a positive parameter.

- **Note:** Suppose we have a Bernoulli process with success probability $\pi$ and we would continue the trial until we obtain $r$ successes. Let $Y = \text{extra \# of trial in order to achieve our goal}$, then the distribution of $Y$ is called a \textit{negative binomial} with \textit{pmf}

\[
f(y) = \binom{y + r - 1}{r - 1} \pi^r (1 - \pi)^y, \quad y = 0, 1, 2, ...
\]

\[
\Rightarrow 
E(Y) = \frac{r(1 - \pi)}{\pi}, \quad \text{var}(Y) = \frac{r(1 - \pi)}{\pi^2} = \mu + \frac{1}{r}\mu^2
\]

In this case $D = 1/r$. 
• In the general negative binomial distribution, we can allow $r$ to be a non-integer. If $r \to \infty$, we have the Poisson distribution.

• The above distribution can be specified in SAS using dist=negbin.

• SAS program and output for the crab data:

```sas
title "Analysis of crab data using Negative Binomial distribution with log link";
proc genmod data=crab;
   model satell = width / dist=negbin link=log; * other links are possible;
run;
```

**************************************************************************

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-4.0525</td>
<td>1.2642</td>
<td>-6.5303 -1.5747</td>
<td>10.28</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.1921</td>
<td>0.0476</td>
<td>0.0987 0.2854</td>
<td>16.27</td>
</tr>
<tr>
<td>Dispersion</td>
<td>1</td>
<td>1.1055</td>
<td>0.1971</td>
<td>0.7795 1.5679</td>
<td></td>
</tr>
</tbody>
</table>

$\Rightarrow$ 1. $\hat{D} = 1.1$.

2. Fitted model: $\log(\hat{\mu}) = -4.0525 + 0.1921x$. similar fit.

• **Note:** We don’t use the option scale=. There may be some computational issue with neg. bin. dist.
III.5 GLMs for rate data

- When the response $Y$ represents the number of events occurred over a time window with length $T$ or over a population with size $T$, etc, it may be more meaningful to model the rate data $R = Y/T$.

- Let $\mu = E(Y)$. Then the expected rate $r = E(R)$ is

$$r = \frac{\mu}{T}.$$ 

- If we assume a log-linear model for the rate $r$:

$$\log(r) = \alpha + \beta x,$$

then the model for $\mu$ is

$$\log(\mu) = \log(T) + \alpha + \beta x.$$ 

The term $\log(T)$ is called an offset and can be specified using offset=logt if we define the variable logt = log(T).
- Example: British train accidents over time (Table 3.4, page 83):

<table>
<thead>
<tr>
<th>Year</th>
<th>Train-km</th>
<th>Train Collisions</th>
<th>Train-road Collisions</th>
<th>Year</th>
<th>Train-km</th>
<th>Train Collisions</th>
<th>Train-road Collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>518</td>
<td>0</td>
<td>3</td>
<td>1988</td>
<td>443</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2002</td>
<td>516</td>
<td>1</td>
<td>3</td>
<td>1987</td>
<td>397</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2001</td>
<td>508</td>
<td>0</td>
<td>4</td>
<td>1986</td>
<td>414</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>2000</td>
<td>503</td>
<td>1</td>
<td>3</td>
<td>1985</td>
<td>418</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1999</td>
<td>505</td>
<td>1</td>
<td>2</td>
<td>1984</td>
<td>389</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>1998</td>
<td>487</td>
<td>0</td>
<td>4</td>
<td>1983</td>
<td>401</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1997</td>
<td>463</td>
<td>1</td>
<td>1</td>
<td>1982</td>
<td>372</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1996</td>
<td>437</td>
<td>2</td>
<td>2</td>
<td>1981</td>
<td>417</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1995</td>
<td>423</td>
<td>1</td>
<td>2</td>
<td>1980</td>
<td>430</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1994</td>
<td>415</td>
<td>2</td>
<td>4</td>
<td>1979</td>
<td>426</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1993</td>
<td>425</td>
<td>0</td>
<td>4</td>
<td>1978</td>
<td>430</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1992</td>
<td>430</td>
<td>1</td>
<td>4</td>
<td>1977</td>
<td>425</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1991</td>
<td>439</td>
<td>2</td>
<td>6</td>
<td>1976</td>
<td>426</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>1990</td>
<td>431</td>
<td>1</td>
<td>2</td>
<td>1975</td>
<td>436</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>1989</td>
<td>436</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: British Department of Transport.
\begin{itemize}
  \item $y =$ yearly \# of train accidents with road vehicles from 1975-2003.
  \item $T =$ \# of train-KM's.
  \item $x =$ \# of years since 1975.
  \item Consider log-rate GLM:
    \begin{equation*}
      \log(\mu) = \log(T) + \alpha + \beta x.
    \end{equation*}
\end{itemize}

```
title "Analysis of British train accident data";
proc genmod data=train;
  model y = x / dist=poi link=log offset=logt scale=pearson;
run;
```

\begin{verbatim}
Analysis Of Maximum Likelihood Parameter Estimates

Parameter DF Estimate Standard Error Wald 95% Confidence Limits Wald Chi-Square Pr > ChiSq
Intercept 1  -4.2114   0.1987   449.41 4.6008   -3.8221  <.0001
year 1   -0.0329   0.0134    5.99   -0.0593   -0.0066 0.0144
Scale 0    1.2501   0.0000   5.99    1.2501    1.2501
```

$\Rightarrow \log(\text{rate}) = -4.21 - 0.0329x$. Accidents decline overtime.
Note: If we assume a different model for the expected rate $r$, we will have a different model for $\mu = E(Y)$. The thing that matters is to find a model for $\mu = E(Y)$.

For example, if we assume

$$\frac{1}{r} = \alpha + \beta x, \quad \Rightarrow \frac{T}{\mu} = \alpha + \beta x$$

$$\Rightarrow$$

$$\frac{1}{\mu} = \alpha(1/T) + \beta(x/T).$$

So the link function is $g(\mu) = \mu^{-1}$. If we define $t1$ for $1/T$ and $x1$ for $x/T$ in our data set, then we can use the following program to fit the above model:

```r
proc genmod data=mydata;
    model y = t1 x1 / noint dist=poi link=power(-1) scale=pearson;
run;
```
IV  Inference for GLM and Model Checking

IV.1  Inference for $\beta$ in a GLM

- After we fit a GLM, we can make inference on $\beta$ such as:
  - Wald test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$:
    $$Z = \frac{\hat{\beta}}{SE(\hat{\beta})}$$
    Compare $Z$ to $N(0,1)$ to get p-value (Note: $SE(\hat{\beta})$ has to be the correct SE, e.g. needs to account for over-dispersion).
  - LRT test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$ with NO over-dispersion:
    $$G^2 = 2(\log L_1 - \log L_0),$$
    where $L_0$ is the maximum likelihood of model under $H_0$, $L_1$ is the maximum likelihood of model under $H_0 \cup H_a$.
    Compare $G^2$ to $\chi^2_1$.
    In order to construct the LRT, we need to fit two models, one
under $H_0$, one under $H_0 \cup H_a$.

* LRT test for $H_0 : \beta = 0$ v.s. $H_a : \beta \neq 0$ with over-dispersion:

\[ G^2 = \frac{2(\log L_1 - \log L_0)}{\hat{\phi}} , \]

where $\hat{\phi}$ is the estimate $\phi$ under $H_0 \cup H_a$. Compare $G^2$ to $\chi^2_1$.

For the crab data:

```
proc genmod data=crab;
  model satell = / dist=poi link=log;
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>172</td>
<td>632.7917</td>
<td>3.6790</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>172</td>
<td>632.7917</td>
<td>3.6790</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>172</td>
<td>584.0436</td>
<td>3.3956</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>172</td>
<td>584.0436</td>
<td>3.3956</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>35.9898</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-494.0447</td>
<td></td>
</tr>
</tbody>
</table>
proc genmod data=crab;
  model satell = width / dist=poi link=log;
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>171</td>
<td>567.8786</td>
<td>3.3209</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>171</td>
<td>567.8786</td>
<td>3.3209</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>171</td>
<td>544.1570</td>
<td>3.1822</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>171</td>
<td>544.1570</td>
<td>3.1822</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>68.4463</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-461.5881</td>
<td></td>
</tr>
</tbody>
</table>

\[ G^2 = \frac{2(68.4463 - 35.9898)}{3.1822} = 20.2, \text{ compared to } \chi^2_1. \]
Construct a \((1 - \alpha)\) CI for \(\beta\):

\[
[\hat{\beta} - z_{\alpha/2}SE(\hat{\beta}), \hat{\beta} + z_{\alpha/2}SE(\hat{\beta})] = [\hat{\beta}_L, \hat{\beta}_U]
\]

\[\Rightarrow\] We can get a CI for functions of \(\beta\).

For example, in a logistic regression, \(e^\beta\) is the odds-ratio (\(\theta\)) of success with one unit increase of \(x\). Then a \((1 - \alpha)\) CI for \(e^\beta = \theta\):

\[
[e^{\hat{\beta}_L}, e^{\hat{\beta}_U}].
\]
IV.2 Model checking

- In some situations, we can check to see if a GLM
  \[ g(\mu) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \]
  fits the data well.

- **Conditions**: No over-dispersion (e.g. binary/binomial data), # of unique values of \( x \) is fixed, \( n_i \to \infty \).

- Snoring/Heart disease example:

<table>
<thead>
<tr>
<th>Snoring</th>
<th>Heart Disease</th>
<th>Yes ((y_i))</th>
<th>No</th>
<th>(n_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Never</td>
<td></td>
<td>24</td>
<td>1355</td>
<td>1379</td>
</tr>
<tr>
<td>Snoring</td>
<td></td>
<td>35</td>
<td>605</td>
<td>638</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td></td>
<td>21</td>
<td>192</td>
<td>213</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td></td>
<td>30</td>
<td>224</td>
<td>254</td>
</tr>
<tr>
<td>5 Every night</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Slide 183**
If we consider the data as $y_i | n_i \sim Bin(n_i, \pi_i)$, $i = 1, 2, 3, 4 = I$, we have $I = 4$ data points.

Consider a model such as the logistic regression:

$$\text{logit}\{\pi(x)\} = \alpha + \beta x.$$ 

$\Rightarrow$ ML $L_M$.

A **Saturated model** has a separate $\pi_i$ for each value of $x$ (perfect fit).

$\Rightarrow$ ML $L_S$.

**Deviance** is the LRT comparing current model to the saturated model:

$$\text{Dev} = 2[\log(L_S) - \log(L_M)].$$

If the current model is good, then $\text{Dev} \sim \chi^2_{I-(p+1)}$. A smaller Dev indicates a better fit.
SAS proc genmod automatically presents the Deviance for a model:

```
proc genmod;
  model y/n = score / dist=bin link=logit;
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>2</td>
<td>2.8089</td>
<td>1.4045</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>2</td>
<td>2.8089</td>
<td>1.4045</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>2</td>
<td>2.8743</td>
<td>1.4372</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>2</td>
<td>2.8743</td>
<td>1.4372</td>
</tr>
</tbody>
</table>

Linear probability model is better than the logistic model using deviance!
Note: We can also use the following *Pearson* $\chi^2$ statistic for model checking in this situation:

\[
\chi^2 = \sum \frac{(y_i - \hat{E}(y_i)_{\text{model}})^2}{\hat{\text{var}}(y_i)_{\text{model}}}
\]

where $\hat{E}(y_i)_{\text{model}}$ is the est. mean of $y_i$ under current model, $\hat{\text{var}}(y_i)_{\text{model}}$ is the est. variance of $y_i$ under current model.

If the model fits the data well, $\chi^2 \sim \chi^2_{I-(p+1)}$. A small $\chi^2$ indicates a better fit.

If we use the Pearson $\chi^2$, we get the same conclusion: Linear probability model is better than the logistic model!

Note: If $Y$ is binary, we should use option aggregate= in the model statement:

```latex
proc genmod descending;
  model hd = score / dist=bin link=logit aggregate=score;
run;
```
IV.3 Residuals

- We can obtain Deviance residuals or Pearson $\chi^2$ residuals after fitting a GLM.

- Deviance residuals:

$$\text{Dev} = 2[\log(L_S) - \log(L_M)] = \sum d_i,$$

$$r_{Di} = d_i^{1/2} \text{sign}(y_i - \mu_i)$$ is the deviance residual.

- Standardized Deviance residuals is the standardized version of $r_{Di}$.
  Standardized deviance residuals can be used to identify outliers.

- Pearson residuals:

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\text{var}(y_i)}}.$$

$$\mathbb{E}(e_i) \approx 0, \text{ var}(e_i) < 1.$$
• Standardized Pearson residual:

\[ r_i = \frac{y_i - \hat{\mu}_i}{SE}. \]

\( E(r_i) \approx 0, \ var(r_i) \approx 1, \ r_i \) behaves like a N(0,1) variable. Standardized Pearson residuals can be used to identify outliers.

• Use residuals in the model Statement of Proc Genmod to obtain these residuals.
4 Logistic Regression

I Logistic Model and Its Interpretation

I.1 The logistic regression model

- For binary response $Y$ with $\pi(x) = P[Y = 1|x]$, a logistic regression model for $\pi(x)$ is

$$\text{logit}\{\pi(x)\} = \log\left\{\frac{\pi(x)}{1 - \pi(x)}\right\} = \alpha + \beta x.$$ 

$$\Rightarrow \frac{\pi(x)}{1 - \pi(x)} = e^{\alpha + \beta x}$$

$$\pi(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}.$$
I.2 Odds-ratio interpretation

- Interpretation of $\alpha$, $\beta$:

$$\alpha = \log \left\{ \frac{\pi(0)}{1 - \pi(0)} \right\} : \log \text{odds of success at } x = 0$$

$$\pi(0) = \frac{e^{\alpha}}{1 + e^{\alpha}}.$$  

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

log odds-ratio of success with 1 unit increase of $x$

$$e^\beta = \frac{\pi(x+1) \{1 - \pi(x+1)\}}{\pi(x) \{1 - \pi(x)\}}$$

odds-ratio of success with 1 unit increase of $x$
I.3 Empirical check of the logistic model

- Suppose at $x_i$ there are $n_i$ obs and $y_i$ successes, and $n_i$ is reasonably large. Since $p_i = y_i/n_i$ is a good estimate of $\pi_i$, so if

$$\logit(\pi_i) = \alpha + \beta x_i$$

is a good model, the plot of $p_i$ v.s. $x_i$ will look like a logistic curve. However, not easy to tell visually.

- Better to plot $\logit(p_i)$ v.s. $x_i$. If the logistic model is good, then this plot should roughly show a linear line.

- $p_i$ may be 0 or 1, in which case $\logit(p_i)$ is undefined. Add 0.5 to success and failure and recalculate sample proportion $\tilde{p}_i$. Or equivalently calculate the odds

$$\text{odds}_i = \frac{y_i + 0.5}{n_i - y_i + 0.5}$$

and plot $\log(\text{odds}_i)$ v.s. $x_i$. A roughly linear line indicates the model is reasonable. Better to group data.
I.4 Example: Horseshoe crab data

- For crab data, define binary response $Y_i$ for female crab $i$ as

$$Y_i = \begin{cases} 
1 & \text{if crab } i \text{ has at least one satellite} \\
0 & \text{otherwise}
\end{cases}$$

- Define $\pi(x_i) = P[Y_i = 1|x_i]$, where $x_i$ is the carapace width of female crab $i$.

- First would like to check if

$$\logit \pi(x_i) = \alpha + \beta x_i$$

is reasonable.
• SAS program and output:

```sas
data crab;
input color spine width satell weight;
  weight=weight/1000; color=color-1;
  y=(satell>0);
datalines;
  3 3 28.3 8 3050
  4 3 22.5 0 1550
  2 1 26.0 9 2300
  ....;

title "Define mid width for every [w+0.25, w+1.25]";
data crab; set crab;
  if width <=23.25 then
    mid_width = 22.75;
  else if width <= 29.25 then
    mid_width = ceil(width-0.25) - 0.25;
  else
    mid_width = 29.75;
run;

proc sort data=crab;
  by mid_width;
run;

proc summary data=crab noprint;
  var y;
  by mid_width;
  output out=crab2 sum=y;
run;
```

data crab2; set crab2;
  ni = _FREQ_
  logitpi = log((y + 0.5)/(ni - y + 0.5));
run;

title "Empirical logit vs. mid width";
proc plot;
  plot logitpi*mid_width;
run;

***************************************************************
Empirical logit vs. mid width 1
Plot of logitpi*mid_width. Legend: A = 1 obs, B = 2 obs, etc.

- The above plot indicates that the logistic model may be reasonable.
• We can use Proc GenMod or Proc Logistic to fit

\[ \text{logit} \pi(x_i) = \alpha + \beta x_i. \]

Here we use Proc Logistic:

```sas
title "Logistic fit to the probability of having satellites";
proc logistic data=crab descending;
   model y=width;
run;
```

• **Note**: Here we need to use “descending” option since the response variable \( Y_i \) is 1/0 and we want to model \( P[Y_i = 1|x_i] \). Otherwise, SAS models \( P[Y_i = 0|x_i] \).

• SAS output:

```
LOGISTIC Procedure
Model Information

Data Set             WORK.CRAB
Response Variable    y
Number of Response Levels 2
Model                binary logit
Optimization Technique   Fisher’s scoring
```

Slide 195
Number of Observations Read 173
Number of Observations Used 173

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>Total y Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>2</td>
<td>62</td>
</tr>
</tbody>
</table>

Probability modeled is y=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

<table>
<thead>
<tr>
<th>Model</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>227.759</td>
<td>198.453</td>
</tr>
<tr>
<td>SC</td>
<td>230.912</td>
<td>204.759</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>225.759</td>
<td>194.453</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>31.3059</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>27.8752</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>23.8872</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-12.3508</td>
<td>2.6287</td>
<td>22.0749</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.4972</td>
<td>0.1017</td>
<td>23.8872</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>width</td>
<td>1.644</td>
<td>1.347 2.007</td>
</tr>
</tbody>
</table>

- The estimated model for $\pi(x)$:

$$\text{logit}\hat{\pi}(x) = -12.351 + 0.497x.$$  

- $e^{0.497} = 1.64 = \text{the odds-ratio of having satellites associated with one cm increase in carapace width.}$  
  $\Rightarrow 64\% \text{ increase in odds of having satellites with one cm increase in carapace width.}$
Figure 4.3. Observed and fitted proportions of satellites, by width of female crab.
I.5 Approximate linear interpretation of the logistic model

- From the above fitted model, it is observed that \( \hat{\pi}(x) \) is approximately linear from \( x = 23 \sim 27 \). At \( x_0 = 25 \), \( \hat{\pi}(x_0) \approx 0.5 \).

- Simple algebra shows the slope of \( \pi(x) \) at \( x \) is

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

which can be approximately interpreted as the change in success probability \( \pi(x) \) when \( x \) increases by one unit from \( x \) to \( x + 1 \).

At \( x_0 = -\alpha/\beta \), \( \alpha + \beta x_0 = 0, \Rightarrow \pi(x_0) = 0.5 \)
\[
\Rightarrow \pi'(x_0) = \frac{\beta}{4}
\]

\( \Rightarrow \) Success prob increases (if \( \beta > 0 \)) by \( \beta/4 \) additively when \( x \) increases by one unit from \( x_0 \) to \( x_0 + 1 \) (or \( x \) to \( x + 1 \) for \( x \) around \( x_0 \)).

So success prob increases (if \( \beta > 0 \)) from 0.5 to 0.75 (0.5+1/4) additively when \( x \) increases from \( x_0 = -\alpha/\beta \) to \( x_0 + 1/\beta \).
• For crab data,

\[ \hat{\pi}'(x_0) = \frac{\hat{\beta}}{4} = 0.1243. \]

⇒ With 1 cm increase in carapace width in [23,27], the prob of having satellite increases additively by 12.43%.

• We can also fit a linear probability model (using LS) to the binary data \( y_i \) and got the fit:

\[ \hat{\pi}(x) = -1.766 + 0.092 x. \]

The slope estimate in this model is comparable to \( \frac{\hat{\beta}}{4} = 0.1243 \) from the logistic model.
I.6 Logistic model for retrospective studies (e.g., case-control studies)

<table>
<thead>
<tr>
<th>Covariate</th>
<th>$Y = 1$</th>
<th>$Y = 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$x_I$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_1$</td>
<td></td>
<td>$n_0$</td>
</tr>
</tbody>
</table>

- With a multinomial sample (random sample), or a product-binomial sample on $X$, we can model $\pi(x) = P[Y = 1|x]$.
- Assume the logistic model

$$\logit\{\pi(x)\} = \alpha + \beta x$$

is true in the population, we then can make inference on $\alpha$ and $\beta$ using
the data.
• However, for rare events (either in terms of $Y = 1$ or $Y = 0$), it is not efficient to conduct a multinomial sampling or a product-binomial sampling on $X$. A solution is to conduct case-control studies.
• Question: Suppose we have data from a case-control study, can we still make inference on $\alpha$, $\beta$ (especially on $\beta$)?
• In a case-control study, we (randomly) sample $n_1$ cases and $n_0$ controls (we may over-sample or under-sample cases). Then their exposure history ($x$) is identified.
• Let $\pi^*(x) = P[Y = 1|x, \text{design}]$, then it can be shown that $\pi^*(x)$ also has a logistic model with the same slope $\beta$:

$$\logit\{\pi^*(x)\} = \alpha^* + \beta x,$$

where $\alpha^*$ depends on $\alpha$ and sampling prob's for cases and controls. We can ignore the design and fit the logistic model!

Logistic model is the ONLY GLM that has this invariance property!
I.7 Normal model for $X \Rightarrow$ logistic model for $Y$

- Suppose both $X$ and $Y$ are random variables. $Y = 1/0$, and

$$X|Y = 1 \sim \text{N} (\mu_1, \sigma^2), \quad X|Y = 0 \sim \text{N} (\mu_0, \sigma^2).$$

Then given data $(x_i, y_i)$ $(i = 1, 2, ..., n)$ from a multinomial sampling, we can conduct a two-sample t-test to test $H_0: \mu_1 = \mu_0$.

- It can be shown that $\pi(x) = P[Y = 1|X = x]$ satisfies logistic model:

$$\logit \pi(x) = \alpha + \beta x$$

where $\beta = (\mu_1 - \mu_0)/\sigma$.

- The two-sample t-test for $H_0: \mu_1 = \mu_0 \leftrightarrow H_0: \beta = 0$ from a logistic model!

- If $X|Y = 1$ and $X|Y = 0$ have different variances, then we need an extra term $\beta_2 x^2$ in the logistic model.
II Inference for Logistic Models

II.1 Hypothesis testing

- Model

\[ \text{logit}\{\pi(x)\} = \alpha + \beta x \]

We are interested in testing \( H_0 : \beta = 0 \) (\( x \) has no effect on \( Y \)) v.s \( H_a : \beta \neq 0 \)

1. Wald Test: Compare \( Z = \frac{\hat{\beta}}{\text{SE}(\beta)} \) to \( \text{N}(0,1) \), or \( Z^2 \) to \( \chi^2_1 \).

2. LRT Test:
   - Fit the full model \( \text{logit}\{\pi(x)\} = \alpha + \beta x \Rightarrow \ell_1 \)
   - Fit the null model \( \text{logit}\{\pi(x)\} = \alpha \Rightarrow \ell_0 \)
   - Compare \( G^2 = 2(\ell_1 - \ell_0) \) to \( \chi^2_1 \).

3. Score Test: based on \( U = \frac{\partial \ell}{\partial \beta} \bigg|_{H_0} \).

Proc Logistic of SAS reports all of them.
II.2 Confidence intervals of $\beta$

- Two CI's for $\beta$
  1. Wald CI for $\beta$: $\hat{\beta} \pm z_{\alpha/2}\hat{SE}(\hat{\beta})$.
  2. LR (likelihood ratio) CI for $\beta$: invert the LRT test, i.e., collect all $\beta_0$ such that

$$G^2(Y, x; \beta_0) \leq \chi^2_{1, \alpha}$$

where $G^2(Y, x; \beta_0)$ is the LRT stat for testing $H_0: \beta = \beta_0$.

Software:
Proc Logistic; * may need "descending" here;
model y = x / aggregate=(x) scale=none CLparm=PL Wald Both CLodds=PL Wald Both;
  *or model y/n = x / aggregate=(x) scale=none CLparm=PL Wald Both CLodds=PL Wald Both;
Run;

or
Proc Genmod; * may need "descending" here;
model y = x / dist=bin LRCI;
  * or model y/n = x / dist=bin LRCI;
Run;

aggregate scale=none is for goodness-of-fit $\chi^2$ and Deviance.
II.3 Confidence interval of $\pi(x_0)$

- True success prob $\pi(x_0)$ at $x_0$:

$$\pi(x_0) = \frac{e^{\eta(x_0)}}{1 + e^{\eta(x_0)}},$$

where $\eta(x_0) = \alpha + \beta x_0$, with estimate

$$\widehat{\eta}(x_0) = \widehat{\alpha} + \widehat{\beta} x_0,$$

$$\text{var}(\widehat{\eta}(x_0)) = \text{var}(\widehat{\alpha}) + 2x_0 \text{cov}(\widehat{\alpha}, \widehat{\beta}) + x_0^2 \text{var}(\widehat{\beta})$$

$\implies$ (1 - $\alpha$) CI for $\eta(x_0)$: $\widehat{\eta}(x_0) \pm z_{\alpha/2}\{\text{var}(\widehat{\eta}(x_0))\}^{1/2} = [\widehat{\eta}_1, \widehat{\eta}_2]$

$\implies$ (1 - $\alpha$) CI for $\pi(x_0)$:

$$\left[ \frac{e^{\widehat{\eta}_1}}{1 + e^{\widehat{\eta}_1}}, \frac{e^{\widehat{\eta}_2}}{1 + e^{\widehat{\eta}_2}} \right]$$

- Note: Need to use option covout in Proc Logistic, or option covb in model statement of Proc GenMod to get $\widehat{\text{cov}}(\widehat{\alpha}, \widehat{\beta})$. 
Note: If we define $x^* = x - x_0$ and fit

$$\text{logit} \pi^*(x^*) = \alpha^* + \beta x^*$$

Then $\pi^*(0) = \pi(x_0)$ and

$$\pi^*(0) = \frac{e^{\alpha^*}}{1 + e^{\alpha^*}}$$

$(1 - \alpha)$ CI for $\alpha^*$ is $\hat{\alpha}^* \pm \frac{z_{\alpha/2}}{\hat{\text{SE}}(\hat{\alpha}^*)} = [\hat{\alpha}^*_1, \hat{\alpha}^*_2]$. 

$$\implies (1 - \alpha) \text{ CI for } \pi(x_0) = \pi^*(0) \text{ will be}$$

$$\left[ \frac{e^{\hat{\alpha}^*_1}}{1 + e^{\hat{\alpha}^*_1}}, \frac{e^{\hat{\alpha}^*_2}}{1 + e^{\hat{\alpha}^*_2}} \right].$$
For crab data, the satellite probability at $x_0 = 26.5$ is

$$\hat{\pi}(x_0) = \frac{e^{-12.351+0.497(26.5)}}{1 + e^{-12.351+0.497(26.5)}} = 0.695.$$  

$$\hat{\eta}(x_0) = \hat{\alpha} + \hat{\beta} x_0 = -12.351 + 0.497(26.5) = 0.825$$

$$\text{var}\{\hat{\eta}(x_0)\} = \text{var}(\hat{\alpha}) + 2x_0\text{cov}(\hat{\alpha}, \hat{\beta}) + x_0^2\text{var}(\hat{\beta})$$

$$= 6.9102 + 2(26.5)(-0.2668) + (26.5)^2(0.0103) = 0.038.$$  

The 95% CI for $\eta(x_0)$ is

$$\hat{\eta}(x_0) \pm z_{0.025}\sqrt{\text{var}\{\hat{\eta}(x_0)\}}^{1/2} = 0.825 \pm 1.96\sqrt{0.038} = [0.44, 1.21].$$

The 95% CI for $\pi(x_0)$ is

$$\left[ \frac{e^{0.44}}{1 + e^{0.44}}, \frac{e^{1.21}}{1 + e^{1.21}} \right] = [0.61, 0.77].$$
• **Note:** The CI for $\pi(26.5)$ can also be obtained from **Proc Logistic**:

```plaintext
proc logistic data=crab descending;
  model y=width;
  output out=out predicted=pihat lower=lower upper=upper / alpha=0.05;
run;
```

---

### mid_ Obs color spine width satell weight y width _LEVEL_ pihat lower upper

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>2</td>
<td>3</td>
<td>26.3</td>
<td>1</td>
<td>2.400</td>
<td>1</td>
<td>26.75</td>
<td>1</td>
<td>0.67400</td>
<td>0.59147</td>
<td>0.74700</td>
</tr>
<tr>
<td>98</td>
<td>1</td>
<td>1</td>
<td>26.5</td>
<td>0</td>
<td>2.350</td>
<td>0</td>
<td>26.75</td>
<td>1</td>
<td>0.69546</td>
<td>0.61205</td>
<td>0.76775</td>
</tr>
</tbody>
</table>

• If the value $x_0$ is not in the data set, we can insert one data point with $x_0$ only (others are missing). For example, $x_0 = 22.8$ is not in the data set, then we insert one data point before we run the above program:

```plaintext
data x0;
  input width y;
  cards;
  22.8 . ;
run;

data crab; set crab x0;
run;
```

---

### mid_ Obs color spine width satell weight y width _LEVEL_ pihat lower upper

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>22.5</td>
<td>4</td>
<td>1.475</td>
<td>1</td>
<td>22.75</td>
<td>1</td>
<td>0.23810</td>
<td>0.12999</td>
<td>0.39528</td>
</tr>
<tr>
<td>6</td>
<td>.</td>
<td>.</td>
<td>22.8</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>22.75</td>
<td>1</td>
<td>0.26621</td>
<td>0.15454</td>
<td>0.41861</td>
</tr>
</tbody>
</table>

---

*Slide 209*
II.4 Use model to gain efficiency

- Using a model such as the logistic model can provide a more efficient probability estimate (smaller standard error estimate or shorter confidence interval with the same confidence level).

For example, if we assume the logistic regression model is correct, then the 95% CI for $\pi(26.5)$ is [0.61, 0.77].

In the data set, at $x = 26.5$, there are 6 female crabs with 4 having satellites. So another estimate of $\pi(26.5)$ is $p = 4/6 = 0.667$. A large sample 95% CI without using the logistic model is:

$$
\frac{4}{6} \pm 1.96 \sqrt{\frac{0.667(1 - 0.667)}{6}} = [0.290, 1.044] = [0.29, 1].
$$

The exact 95% CI for $\pi(26.5)$ $4/6$ is [0.22, 0.96]. Both the large sample and exact CIs are much wider than the one based on the model.
III Logistic Model with Categorical Predictors

III.1 Logistic model with indicator variables for $2 \times 2 \times 2$ tables

- Example: AIDS and AZT use (table 4.4, p. 112)

$$Y = \begin{cases} 
1 & \text{AIDS Sym.} \\
0 & \text{No AIDS Sym.}
\end{cases}, \quad X = \begin{cases} 
1 & \text{immediate AZT use} \\
0 & \text{Wait until immunity is weak}
\end{cases}, \quad Z = \begin{cases} 
1 & \text{White} \\
0 & \text{Back}
\end{cases}$$

\[
\begin{array}{c|cc|c}
 & Y = 1 & Y = 0 & \text{Total} \\
\hline
X = 1 & 14 & 93 & 109 \\
X = 0 & 32 & 81 & 113 \\
Z = 1 \quad \text{Total} & 46 & 174 & 220 \\
\hline
Z = 0 \quad \text{Total} & 63 & 182 & 245 \\
\end{array}
\]
• Define prob of having AIDS Symptom

\[ \pi(x, z) = P[Y = 1|x, z], \quad x, z = 0, 1 \]

and consider the following “main-effect” only model

\[ \text{logit}\{\pi(x, z)\} = \alpha + \beta_1 x + \beta_2 z. \]

• Model implies:

\[
\begin{align*}
\text{logit}\pi(x = 1, z) &= \alpha + \beta_1 + \beta_2 z \\
\text{logit}\pi(x = 0, z) &= \alpha + 0 + \beta_2 z \\
\Rightarrow \beta_1 &= \text{logit}\pi(x = 1, z) - \text{logit}\pi(x = 0, z) \\
\Rightarrow e^{\beta_1} &= \frac{\pi(x = 1, z)/\{1 - \pi(x = 1, z)\}}{\pi(x = 0, z)/\{1 - \pi(x = 0, z)\}}
\end{align*}
\]

The odds-ratio between \( X \) and \( Y \) at \( Z = 0 \) (black) is the same as that at \( Z = 1 \) (white) \((= e^{\beta_1})\) \(\Rightarrow\) common odds-ratio!
⇒ The partial associations between $X$ and $Y$ are the same at $Z = 0$ (black) and $Z = 1$ (white) and are equal to $e^{\beta_1}$.
⇒ *homogeneous* $XY$ association across levels of $Z$.

- Model also implies:

$$\begin{align*}
\logit \pi(x, z = 1) &= \alpha + \beta_1 x + \beta_2 \\
\logit \pi(x, z = 0) &= \alpha + \beta_1 x + 0 \\
\Rightarrow \beta_2 &= \logit \pi(x, z = 1) - \logit \pi(x, z = 0) \\
\Rightarrow e^{\beta_2} &= \frac{\pi(x, z = 1)/\{1 - \pi(x, z = 1)\}}{\pi(x, z = 0)/\{1 - \pi(x, z = 0)\}}
\end{align*}$$

⇒ The partial associations between $Z$ and $Y$ are the same at $X = 0$ and $X = 1$ and are equal to $e^{\beta_2}$.
⇒ *homogeneous* $ZY$ association across levels of $X$.

Of course, we are more interested in whether immediate AZT use works. That is, we are more interested in the partial association $e^{\beta_1}$.
• If $\beta_1 = 0 \Rightarrow X, Y$ are conditionally indep given $Z$

If $\beta_2 = 0 \Rightarrow Z, Y$ are conditionally indep given $X$

• Given data in the form of contingency tables

\[
\begin{array}{c|cc|c|cc|c}
 & Y = 1 & Y = 0 & & Y = 1 & Y = 0 & \\
\hline
X & 1 & & & 1 & & \\
0 & & & & & & \\
\hline
Z & & & & & & \\
1 & & & & & & \\
0 & & & & & & \\
\end{array}
\]

we can fit the above *homogeneous* model and test the above conditional independence hypotheses (particularly $X \perp Y | Z$) under the assumed model using the Wald, LRT and score test.
SAS program and partial output:

```sas
data table5_6;
  input azt race sym nosym;
  n = sym+nosym;
  datalines;
  1 1 14 93
  0 1 32 81
  1 0 11 52
  0 0 12 43
;

proc genmod;
  model sym/n = azt race / dist=bin link=logit type3 lrci;
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>1</td>
<td>1.3835</td>
<td>1.3835</td>
</tr>
<tr>
<td>Scaled 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>
<tr>
<td>Scaled Pearson X2</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>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Likelihood Ratio</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.0736</td>
<td>0.2629</td>
<td>-1.6088 -0.5735</td>
<td>16.67</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>azt</td>
<td>1</td>
<td>-0.7195</td>
<td>0.2790</td>
<td>-1.2773 -0.1799</td>
<td>6.65</td>
<td>0.0099</td>
<td></td>
</tr>
<tr>
<td>race</td>
<td>1</td>
<td>0.0555</td>
<td>0.2886</td>
<td>-0.5023 0.6334</td>
<td>0.04</td>
<td>0.8476</td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
<td></td>
<td></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>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>
• Wald test for $H_0 : \beta_1 = 0 (X \perp Y | Z)$: $\chi^2 = 6.65$, p-value=0.01.
LRT for $H_0 : \beta_1 = 0 (X \perp Y | Z)$: $G^2 = 6.87$, p-value=0.009. Strong evidence!

• Score test SAS program and partial output:

```sas
title "Main effect model & score test for AZT";
proc logistic;
    model sym/n = race azt / selection=forward slentry=1 include=1;
run;
```

Summary of Forward Selection

<table>
<thead>
<tr>
<th>Step</th>
<th>Effect Entered</th>
<th>DF</th>
<th>Number In</th>
<th>Score Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>azt</td>
<td>1</td>
<td>2</td>
<td>6.8023</td>
<td>0.0091</td>
</tr>
</tbody>
</table>

• Score test for $H_0 : \beta_1 = 0 (X \perp Y | Z)$: $\chi^2 = 6.8$, p-value=0.009, closer to LRT.
• From the output, we have:

\[
\hat{\beta}_1 = -0.72 \\
e^{\hat{\beta}_1} = 0.49 \\
\text{SE}(\hat{\beta}_1) = 0.2790 \\
95\% \text{ LRCI for } \beta_1 = [-1.2773, -0.1799] \\
95\% \text{ LRCI for } e^{\beta_1} = [e^{-1.2773}, e^{-0.1799}] = [0.28, 0.84].
\]

⇒ For each race, the odds of having AIDS symptom for patients with immediate AZT treatment is only about half of the odds for patients with delayed AZT treatment.

• **Note 1**: The first program also gives goodness-of-fit Pearson \( \chi^2 = 1.39 \) and deviance=1.38, with \( df = 1 \), p-value=0.24, indicating reasonable fit of the model to the data.
• **Note 2:** We can also consider a model with interaction between AZT use \((x)\) and race \((z)\) in the above logistic model:

\[
\text{logit}\{\pi(x, z)\} = \alpha + \beta_1 x + \beta_2 z + \beta_3 xz.
\]

Model implies:

\[
\text{logit}\pi(x = 1, z) = \alpha + \beta_1 + \beta_2 z + \beta_3 z
\]

\[
\text{logit}\pi(x = 0, z) = \alpha + 0 + \beta_2 z + 0
\]

\[
\Rightarrow \text{logit}\pi(x = 1, z) - \text{logit}\pi(x = 0, z) = \beta_1 + \beta_3 z
\]

\[
\Rightarrow \frac{\pi(x = 1, z)/\{1 - \pi(x = 1, z)\}}{\pi(x = 0, z)/\{1 - \pi(x = 0, z)\}} = e^{\beta_1 + \beta_3 z}
\]

The model allows different treatment effects for different races.

We can test \(H_0 : \beta_3 = 0\) to see if the homogeneous model is adequate.
```r
proc genmod;
  model sym/n = azt race azt*race / dist=bin type3 lrci;
run;
```

Analysis Of Maximum Likelihood Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Likelihood Ratio</th>
<th>Wald</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-1.2763</td>
<td>0.3265</td>
<td>-1.9611 -0.6692</td>
<td>15.28</td>
<td>15.28</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>azt</td>
<td>1</td>
<td>-0.2771</td>
<td>0.4655</td>
<td>-1.2024 0.6394</td>
<td>0.35</td>
<td>0.35</td>
<td>0.5518</td>
</tr>
<tr>
<td>race</td>
<td>1</td>
<td>0.3476</td>
<td>0.3875</td>
<td>-0.3930 1.1367</td>
<td>0.80</td>
<td>0.80</td>
<td>0.3698</td>
</tr>
<tr>
<td>azt*race</td>
<td>1</td>
<td>-0.6878</td>
<td>0.5852</td>
<td>-1.8452 0.4599</td>
<td>1.38</td>
<td>1.38</td>
<td>0.2399</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was held fixed.

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>azt</td>
<td>1</td>
<td>0.35</td>
<td>0.5515</td>
</tr>
<tr>
<td>race</td>
<td>1</td>
<td>0.83</td>
<td>0.3635</td>
</tr>
<tr>
<td>azt*race</td>
<td>1</td>
<td>1.38</td>
<td>0.2395</td>
</tr>
</tbody>
</table>

The Wald and LRT statistics are all equal to 1.38 ($df = 1$), with p-value=0.24.

The LRT statistic 1.38 is the same as the deviance 1.38 from the homogeneous model since the model with interaction is saturated.
III.2 Logistic model for $2 \times 2 \times K$ tables

- An example of multi-center clinical trial evaluating a cream in curing skin infection

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>11</td>
<td>25</td>
<td>16</td>
<td>4</td>
<td>14</td>
<td>5</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>27</td>
<td>22</td>
<td>10</td>
<td>1</td>
<td>8</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$Z = 1$</td>
<td></td>
<td></td>
<td>$Z = 2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
<th>S</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>6</td>
<td>11</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$Z = 5$</td>
<td></td>
<td></td>
<td>$Z = 6$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What we observed: There is a lot of variation in success probabilities among centers.
If we collapse the tables over centers, we got:

\[
\begin{array}{c|cc}
 & Y & \\
S & 55 & 75 \\
F & 47 & 96 \\
\end{array}
\]

\[
\hat{\theta}_{XY} = \frac{96 \times 55}{47 \times 75} \approx 1.5
\]

The above estimate \(\hat{\theta}_{XY}\) may not be very useful since this is not a random sample, so we cannot use the famous formula for calculating the variance of \(\log \hat{\theta}_{XY}\):

\[
\hat{\text{var}}(\log \hat{\theta}_{XY}) \neq \frac{1}{55} + \frac{1}{75} + \frac{1}{47} + \frac{1}{96}
\]

(would be the results if we run model \(y/n=trt\))

\(\Rightarrow\) Should focus on conditional association!
• Let $\pi(x, z) = P[Y = 1|x, z]$, where

\[
Y = 1 \text{ for success, 0 for failure}
\]

\[
x = 1 \text{ for treatment, 0 for control}
\]

\[
z = 1, 2, \ldots, 8 \text{ for centers}
\]

and consider the ANOVA type of (homogeneous) model:

\[
\logit \{ \pi(x, z = k) \} = \alpha + \beta x + \beta^z_k - - - (\star)
\]

• $\Rightarrow$ common odds-ratio model:

\[
\frac{\pi(x = 1, z = k)/\{1 - \pi(x = 1, z = k)\}}{\pi(x = 0, z = k)/\{1 - \pi(x = 0, z = k)\}} = e^{\beta} \text{ trt effect at center } k
\]

\[
\pi(x = 0, z = k)/\{1 - \pi(x = 0, z = k)\} = e^{\alpha+\beta^z_k}
\]

$\beta = 0 \Leftrightarrow X \perp Y|Z$.

Note: Usually, we set $\beta^z_8 = 0$ (reference coding in Proc logistic).
SAS program and output:

```sas
data cream;
input center trt y y0;
n=y+y0;
cards;
1 1 11 25
1 0 10 27
2 1 16 4
2 0 22 10
...
title "Use homogeneous model to test no treatment effect at each center";
proc logistic;
    class center / param=ref;
    model y/n = center trt / selection=f include=1 slentry=1;
run;
```

---

**Summary of Forward Selection**

<table>
<thead>
<tr>
<th>Step</th>
<th>Effect Entered</th>
<th>DF</th>
<th>In</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>trt</td>
<td>1</td>
<td>2</td>
<td>6.5583</td>
<td>0.0104</td>
</tr>
</tbody>
</table>

**Type 3 Analysis of Effects**

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>center</td>
<td>7</td>
<td>58.4897</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>6.4174</td>
<td>0.0113</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>83.8082</td>
<td>8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>76.8096</td>
<td>8</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>58.9946</td>
<td>8</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.8859</td>
<td>0.6755</td>
<td>1.7201</td>
<td>0.1897</td>
</tr>
<tr>
<td>center</td>
<td>1</td>
<td>-2.2079</td>
<td>0.7195</td>
<td>9.4166</td>
<td>0.0022</td>
</tr>
<tr>
<td>center</td>
<td>2</td>
<td>-0.1525</td>
<td>0.7381</td>
<td>0.0427</td>
<td>0.8363</td>
</tr>
<tr>
<td>center</td>
<td>3</td>
<td>-1.0550</td>
<td>0.7457</td>
<td>2.0015</td>
<td>0.1571</td>
</tr>
<tr>
<td>center</td>
<td>4</td>
<td>-3.6264</td>
<td>0.9071</td>
<td>15.9813</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>center</td>
<td>5</td>
<td>-2.7278</td>
<td>0.8184</td>
<td>11.1104</td>
<td>0.0009</td>
</tr>
<tr>
<td>center</td>
<td>6</td>
<td>-4.3548</td>
<td>1.2293</td>
<td>12.5499</td>
<td>0.0004</td>
</tr>
<tr>
<td>center</td>
<td>7</td>
<td>-3.0056</td>
<td>1.0200</td>
<td>8.6836</td>
<td>0.0032</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>0.7769</td>
<td>0.3067</td>
<td>6.4174</td>
<td>0.0113</td>
</tr>
</tbody>
</table>

- From the output:
  \( \hat{\beta} = 0.7769, e^{\hat{\beta}} = 2.17 \).
  \( \hat{SE}(\hat{\beta}) = 0.3067 \implies \) 95% Wald CI of \( \beta \): [0.176, 1.378], 95% Wald CI for \( e^{\beta} \): [1.19, 3.97]

Wald test for \( H_0 : \beta = 0 (X \perp Y \mid Z) : \chi^2 = 6.42 \), p-value=0.01
Score test for \( H_0 : \beta = 0 (X \perp Y \mid Z) : \chi^2 = 6.56 \), p-value=0.01.
**Note:** We can also get LR CI for $\beta$ and LRT for $H_0: \beta = 0$:

```latex
proc genmod;
  class center;
  model y/n = center trt / type3 lrci;
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>7</td>
<td>9.7463</td>
<td>1.3923</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>7</td>
<td>9.7463</td>
<td>1.3923</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>7</td>
<td>8.0256</td>
<td>1.1465</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>7</td>
<td>8.0256</td>
<td>1.1465</td>
</tr>
</tbody>
</table>

***Analysis Of Maximum Likelihood Parameter Estimates***

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Likelihood Ratio</th>
<th>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1</td>
<td>0.7769</td>
<td>0.3067</td>
<td>0.1851</td>
<td>1.3915</td>
<td>6.42</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>center</td>
<td>7</td>
<td>81.21</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>trt</td>
<td>1</td>
<td>6.67</td>
<td>0.0098</td>
</tr>
</tbody>
</table>
LR CI for $\beta$: [0.185, 1.392], LR CI for $e^{\beta}$: $[e^{0.185}, e^{1.392}] = [1.20, 4.02]$. LRT for $H_0 : \beta = 0 (X \perp Y | Z)$: $G^2 = 6.67$, p-value = 0.0098.

The above program also gives the Pearson $\chi^2 = 8.03$ and deviance = 9.75 with $df = 7$ for goodness-of-fit (p-values = 0.33 and 0.20).
### III.2 Cochran-Mantel-Haenszel (CMH) test for $2 \times 2 \times K$ tables

- Another way to test $X \perp Y | Z$ is to use the CMH test. The data at center $k$ can be represented as

$$
\begin{array}{ccc}
  & Y & \\
  S & F & \\
  X & \text{trt} & \\
  \text{control} & n_{11k} & n_{12k} & n_{1+k} \\
  & n_{21k} & n_{22k} & n_{2+k} \\
  & n_{+1k} & n_{+2k} & n_{++k} \\
  Z = k & \\
\end{array}
$$
• Under $H_0 : X \perp Y \mid Z$, $n_{11k} \mid n_{1+k}, n_{+1k} \sim$ hypergeometric distribution:

$$E(n_{11k} \mid H_0, n_{1+k}, n_{+1k}) = \frac{n_{1+k}n_{+1k}}{n_{++k}} = \mu_{11k},$$

$$\text{var}(n_{11k} \mid H_0, n_{1+k}, n_{+1k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^2(n_{++k} - 1)}.$$ 

$\Rightarrow$

$$\chi^2 = \frac{\left[\sum_{k=1}^{K} (n_{11k} - \mu_{11k})\right]^2}{\sum_{k=1}^{K} \text{var}(n_{11k} \mid H_0, n_{1+k}, n_{+1k})} \overset{H_0}{\sim} \chi^2_1.$$ 

This is the Cochran-Mental-Haenszel test for $H_0 : X \perp Y \mid Z$.

• CMH with correction:

$$\chi^2_c = \frac{\left\{\sum_{k=1}^{K} (n_{11k} - \mu_{11k}) - 0.5\right\}^2}{\sum_{k=1}^{K} \text{var}(n_{11k} \mid H_0, n_{1+k}, n_{+1k})} \overset{H_0}{\sim} \chi^2_1.$$ 

• The CMH does not require the homogeneous model.
• For our data, the CMH $\chi^2$ can be calculated as

$$\chi^2 = \frac{\{|(11 - 36 \times 21/73) + (16 - 20 \times 38/52 + \cdots | - 0.5\}^2}{36 \times 37 \times 21 \times 52/(73^2 \times 72) + 20 \times 32 \times 38 \times 14/(52^2 \times 51) + \cdots}$$

$$= 6.38.$$  

Compare $\chi^2 = 6.38$ to $\chi^2_1$ and get p-value= 0.0115.

• **Note**: If we don't reject $H_0 : X \perp Y|Z$ using the CMH test, it may be either $H_0 : X \perp Y|Z$ is true or the conditional association between $X$ and $Y$ have different directions at different levels of $Z$.

• We can use `proc freq` to conduct the above CMH test.
data y1; set cream;
count=y;
drop y0;
y=1;
run;

data y0; set cream;
count=y0;
drop y0;
y=0;
run;

data new; set y1 y0;
run;

title "MH test for conditional independence and MH common OR";
proc freq data=new order=data;
   weight count;
   tables center*trt*y/nopercent norow nocol cmh;
run;

*******************************************************************************
MH test for conditional independence and MH common OR 8
The FREQ Procedure
Summary Statistics for trt by y
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>6.3841</td>
<td>0.0115</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>1</td>
<td>6.3841</td>
<td>0.0115</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>1</td>
<td>6.3841</td>
<td>0.0115</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 Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control</td>
<td>Mantel-Haenszel Logit **</td>
<td>2.1345</td>
<td>1.1776 3.8692</td>
</tr>
<tr>
<td>(Odds Ratio)</td>
<td></td>
<td>1.9497</td>
<td>1.0574 3.5949</td>
</tr>
<tr>
<td>Cohort</td>
<td>Mantel-Haenszel Logit **</td>
<td>1.4245</td>
<td>1.0786 1.8812</td>
</tr>
<tr>
<td>(Col1 Risk)</td>
<td></td>
<td>1.2194</td>
<td>0.9572 1.5536</td>
</tr>
<tr>
<td>Cohort</td>
<td>Mantel-Haenszel Logit</td>
<td>0.8129</td>
<td>0.6914 0.9557</td>
</tr>
<tr>
<td>(Col2 Risk)</td>
<td></td>
<td>0.8730</td>
<td>0.7783 0.9792</td>
</tr>
</tbody>
</table>

** These logit estimators use a correction of 0.5 in every cell of those tables that contain a zero.

Breslow-Day Test for Homogeneity of the Odds Ratios

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9955</td>
<td>7</td>
<td>0.3330</td>
</tr>
</tbody>
</table>

CMH $\chi^2 = 6.3841$, $df = 1$, $p$-value = 0.0115.  
MH Common odds-ratio estimate $\hat{\theta}_{MH} = 2.1345$ with 95% CI [1.1776, 3.8692].  
Breslow-Day Test for common odds-ratio: $\chi^2 = 7.9955$, $df = 7$, $p$-value = 0.3330, similar to the GOF test.
IV Multiple Logistic Regression Models

• $Y$ - binary, multiple $x_1, x_2, \cdots, x_p$, let $\pi(x) = P[Y = 1|x_1, \cdots, x_p]$, a multiple logistic regression model for $\pi(x)$ is

$$\text{logit}\{\pi(x)\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$ 

• If $x_1, x_2, \cdots, x_p$ represent $p$ different covariates, then $\beta_k$ can be interpreted as follows:

$$\text{logit}\{\pi(x_k + 1)\} = \alpha + \beta_1 x_1 + \cdots + \beta_k(x_k + 1) + \cdots + \beta_p x_p$$
$$\text{logit}\{\pi(x_k)\} = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k + \cdots + \beta_p x_p$$
$$\text{logit}\{\pi(x_k + 1)\} - \text{logit}\{\pi(x_k)\} = \beta_k$$
$$\beta_k = \log \left\{ \frac{\pi(x_k + 1)/[1 - \pi(x_k + 1)]}{\pi(x_k)/[1 - \pi(x_k)]} \right\}$$
$$e^{\beta_k} = \frac{\pi(x_k + 1)/[1 - \pi(x_k + 1)]}{\pi(x_k)/[1 - \pi(x_k)]},$$

odds-ratio with 1 unit increase in $x_k$ while other $x$'s are fixed.
• If $x_1, x_2, \cdots, x_p$ do not represent $p$ different covariates, for example, $x_3$ may be defined as $x_1 x_2$. In this case, we have to interpret $\beta_k$'s case by case.

• For example, if $x_1, x_2$ are two unrelated covariates and $x_3 = x_1 x_2$. Then when $x_1$ increases from $x_1$ to $x_1 + 1$ with $x_2$ fixed, then

$$\begin{align*}
\logit\{\pi(x_1 + 1, x_2)\} &= \alpha + \beta_1 (x_1 + 1) + \beta_2 x_2 + \beta_3 (x_1 + 1)x_2 \\
\logit\{\pi(x_1, x_2)\} &= \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 \\
\beta_1 + \beta_3 x_2 &= \logit\{\pi(x_1 + 1, x_2)\} - \logit\{\pi(x_1, x_2)\} \\
e^{\beta_1 + \beta_3 x_2} &= \frac{\pi(x_1 + 1, x_2) / [1 - \pi(x_1 + 1, x_2)]}{\pi(x_1, x_2) / [1 - \pi(x_1, x_2)]}
\end{align*}$$

$\Rightarrow$ The effect of $x_1$ on $\pi(x)$ depends on $x_2$, so $x_2$ is an effect modifier.
IV.1 Logistic model with numeric and categorical covariates.

- Example: Crab data
  - \( x \) – carapace width
  - \( color \) – ordinal variable: medium-light (1), medium (2), medium-dark (3) and dark (4).

- Consider model \( M_1 \) for \( \pi(x, c) = P[Y = 1|x, c_1, c_2, c_3, c_4] \):

\[
M_1 : \text{logit}\{\pi(x, c)\} = \alpha + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 x
\]

- \( c_1 \) dummy for color = medium light
- \( c_2 \) dummy for color = medium
- \( c_3 \) dummy for color = medium dark
- color = dark is used as a reference color

\( \beta_1 \) – log odds-ratio of having a least one satellite between medium-light crabs and dark crabs given that they have the same carapace width.
$\beta_1 - \beta_2$ – comparison between medium-light and medium crabs with the same width.

- SAS program and output:

```sas
proc genmod data=crab descending;
class color;
model y = width color / dist=bin link=logit type3;
run;
```

![Analysis Of Maximum Likelihood Parameter Estimates](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-12.7151</td>
<td>2.7618</td>
<td>-18.1281 -7.3021</td>
<td>21.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.4680</td>
<td>0.1055</td>
<td>0.2611 0.6748</td>
<td>19.66</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>color 1</td>
<td>1</td>
<td>1.3299</td>
<td>0.8525</td>
<td>-0.3410 3.0008</td>
<td>2.43</td>
<td>0.1188</td>
</tr>
<tr>
<td>color 2</td>
<td>1</td>
<td>1.4023</td>
<td>0.5484</td>
<td>0.3274 2.4773</td>
<td>6.54</td>
<td>0.0106</td>
</tr>
<tr>
<td>color 3</td>
<td>1</td>
<td>1.1061</td>
<td>0.5921</td>
<td>-0.0543 2.2666</td>
<td>3.49</td>
<td>0.0617</td>
</tr>
<tr>
<td>color 4</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was held fixed.

![LR Statistics For Type 3 Analysis](image)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>width</td>
<td>1</td>
<td>24.60</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>color</td>
<td>3</td>
<td>7.00</td>
<td>0.0720</td>
</tr>
</tbody>
</table>
The fitted model is

\[ M_1 : \logit\{ \pi(x, c) \} = -12.715 + 1.330 c_1 + 1.402 c_2 + 1.106 c_3 + 0.468 x \]

\[ \hat{\beta}_1 = 1.330, \quad e^{\hat{\beta}_1} = e^{1.330} = 3.78. \] The odds that medium light crabs have satellites is 3.78 times the odds that dark crabs have satellites.

For crabs with the same color, one cm increase in carapace width will increase the odds by \( e^{0.468} - 1 = 0.60 \) (60%).

From the fitted model, we can obtain a fitted model for crabs with a particular color. For example, for medium light crabs with width \( x \), the fitted model is

\[ \logit\{ \pi(x, c = 1) \} = -12.715 + 1.330 + 0.468 x = -11.385 + 0.468 x. \]
Predicted probabilities from model $M_1$

**Figure 4.4.** Logistic regression model using width and color predictors.
• We can test $H_0: \text{no color effects}$ by testing $H_0: \beta_1 = \beta_2 = \beta_3 = 0$. The LRT for $H_0$ is $\chi^2 = 7$ with $df = 3$, p-value=0.0720. Marginally significant.

• Color is an ordinal categorical variable. One way to take this into account is to assign scores to color and treat it as a numerical variable. For example, we may use $c = (1, 2, 3, 4)$ for those 4 color categories and fit

$$M_2 : \text{logit}\{\pi(x, c)\} = \alpha + \beta_1 c + \beta_2 x$$

The fitted model is

$$M_2 : \text{logit}\{\pi(x, c)\} = -10.071 - 0.509c + 0.458x$$
From this fitted model, we obtain:

\[
\frac{\text{odds}(c = 1)}{\text{odds}(c = 4)} = e^{-0.509 \times 1 - (-0.509 \times 4)} = e^{1.527} = 4.6
\]

\[
\frac{\text{odds}(c = 2)}{\text{odds}(c = 4)} = e^{-0.509 \times 2 - (-0.509 \times 4)} = e^{1.018} = 2.768
\]

\[
\frac{\text{odds}(c = 3)}{\text{odds}(c = 4)} = e^{-0.509 \times 3 - (-0.509 \times 4)} = e^{0.509} = 1.664
\]

The LRT comparing \( M_2 \) to \( M_1 \) (\( M_2 \subset M_1 \)):

\[
G^2 = 2 \{-93.7285 - (-94.5606)\} = 1.66, \text{ with } df = 2. \text{ P-value}=0.436
\]

\( \Rightarrow \) Reasonable fit.

However, the estimated effects from these 2 models are very different.
• Fitted model ($M_1$) and Figure 4.4 showed that $c_1$, $c_2$ and $c_3$ have similar effects, indicating that we can group crabs with colors 1, 2, 3 and divide crabs into 2 groups: non-dark (color = 1, 2, 3) and dark (color = 4). Denote $c = 1$ for non-dark crabs and $c = 0$ for dark crabs and consider the model

$$M_3 : \text{logit}\{\pi(x, c)\} = \alpha + \beta_1 c + \beta_2 x$$

The fitted model is

$$M_3 : \text{logit}\{\pi(x, c)\} = -12.980 + 1.301 c + 0.478 x$$

The estimates are very close to those of $M_1$.

The LRT comparing $M_3$ to $M_1$ ($M_3 \subset M_1$):

$$G^2 = 2\{-93.7285 - (-93.9789)\} = 0.501, \text{ with } df = 2.$$ 

P-value=0.778. $\Rightarrow M_3$ has a better fit than $M_2$. 


• We can consider interactions between color and width in the previous models. For example, in $M_3$, we can consider the interaction $c \times x$:

$$M_4 : \text{logit}\{\pi(x, c)\} = \alpha + \beta_1 c + \beta_2 x + \beta_3 c \times x.$$  

The fitted model is

$$M_4 : \text{logit}\{\pi(x, c)\} = -5.854 - 6.958c + 0.200x + 0.322c \times x.$$  

From this, the fitted model for non-dark crabs ($c = 1$):

$$\text{logit}\{\pi(x, c = 1)\} = -5.854 - 6.958 + 0.200x + 0.322x = -12.812 + 0.522x.$$  

The fitted model for dark crabs:

$$\text{logit}\{\pi(x, c = 0)\} = -5.854 + 0.200x.$$  

$$\pi(x, c = 1) > \pi(x, c = 0) \iff -12.812 + 0.522x > -5.854 + 0.200x \iff x > 21.68.$$
title "Logistic model with width and color interaction";
proc genmod data=crab descending;
    model y = c width c*width / dist=bin link=logit type3;
run;
************************************************************************************
Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-5.8538</td>
<td>6.6939</td>
<td>-18.9737 - 7.2660</td>
<td>0.76</td>
<td>0.3818</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>-6.9578</td>
<td>7.3182</td>
<td>-21.3013 - 7.3857</td>
<td>0.90</td>
<td>0.3417</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.2004</td>
<td>0.2617</td>
<td>-0.3124 - 0.7133</td>
<td>0.59</td>
<td>0.4437</td>
</tr>
<tr>
<td>c*width</td>
<td>1</td>
<td>0.3217</td>
<td>0.2857</td>
<td>-0.2381 - 0.8816</td>
<td>1.27</td>
<td>0.2600</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: The scale parameter was held fixed.

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>c</td>
<td>1</td>
<td>0.84</td>
<td>0.3591</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.62</td>
<td>0.4326</td>
</tr>
<tr>
<td>c*width</td>
<td>1</td>
<td>1.17</td>
<td>0.2791</td>
</tr>
</tbody>
</table>

The LRT for the interaction: $G^2 = 1.17 \ (df = 1)$, p-value=0.28, not significant.
V. Summarizing Effects in Logistic Regression Models

- $Y$ - binary, multiple $x_1, x_2, \ldots, x_p$, let $\pi(x) = P[Y = 1|x_1, \ldots, x_p]$, a multiple logistic regression model for $\pi(x)$ is

$$\text{logit}\{\pi(x)\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p.$$

- When $x_1, x_2, \ldots, x_p$ represent $p$ different covariates, then $e^{\beta_k}$ can be interpreted as the odds-ratio of success (disease) with 1 unit increase in $x_k$ while other $x$’s are fixed.

- When $[Y = 1|x]$’s are rare events for some $x$’s, then $e^{\beta_k}$ can be approximately interpreted as the relative risk of disease with 1 unit increase in $x_k$ while other $x$’s are fixed.
• When \([Y = 1|x]’s\) are not rare events \((\pi(x)’s\) are not close to 0), we can apply the linear approximation to \(\pi(x)\):

\[
\frac{\partial \pi(x)}{\partial x_k} = \beta_k \pi(x) \{1 - \pi(x)\}.
\]

⇒ With 1 unit increase in \(x_k\), the success probability will increase additively by approximately \(\beta_k \pi(x) \{1 - \pi(x)\}\).

The approximation will be better around \(x_0\) such that \(\pi(x_0) = 0.5\), where the success prob will increase additively by \(\beta_k/4\).

With multiple \(x’s\), we need to find meaningful \(x_0\). That is, \(x_0\) should represent a meaningful population.
For example, for the crab data with the fitted model:

\[ M_3 : \text{logit}\{\pi(x, c)\} = -12.980 + 1.301c + 0.478x, \]

where \( c = 1 \) for non-dark crabs, \( c = 0 \) for dark crabs, \( x = \) carapace width.

If we set \( x_0 = 24.43, c_0 = 1 \), then \( \pi(x_0, c_0) = 0.5 \). That is, for non-dark crabs, around \( x_0 = 24.43 \), with one cm increase of carapace width, the probability of having satellites increase additively by approximately \( 0.478/4 = 0.12 \).

Alternatively, we can interpret the color effect by fixing \( x \) at its sample mean \( \bar{x} = 26.3 \) cm:

\[
\text{color}=0 : \quad \pi(c = 0, \bar{x}) = \frac{e^{-12.980+0.478\times\bar{x}}}{1 + e^{-12.980+0.478\times\bar{x}}} = 0.40
\]

\[
\text{color}=1 : \quad \pi(c = 1, \bar{x}) = \frac{e^{-12.980+1.301+0.478\times\bar{x}}}{1 + e^{-12.980+1.301+0.478\times\bar{x}}} = 0.71
\]
So when $c$ increases from 0 to 1, the prob increases from 0.4 to 0.71. The difference is 0.31.

This difference $\approx 1.301 \times 0.4 \times (1 - 0.4) = 0.312$.

We may also interpret the width effect by comparing $\pi(c, x)$ at $x_{LQ} = 24.9$ and $x_{UQ} = 27.7$ of $x$ by fixing $c$ at $\bar{c} = 0.873$:

$x_{LQ}$: $\pi(\bar{c}, x_{LQ}) = \frac{e^{-12.980+1.301\times0.873+0.478\times24.9}}{1 + e^{12.980+1.301\times0.873+0.478\times24.9}} = 0.51$

$x_{UQ}$: $\pi(\bar{c}, x_{UQ}) = \frac{e^{-12.980+1.301\times0.873+0.478\times27.7}}{1 + e^{12.980+1.301\times0.873+0.478\times27.7}} = 0.80$

The change rate in prob: $(0.80 - 0.51)/(x_{UQ} - x_{LQ}) = 0.104$

$\approx 0.478 \times 0.51(1 - 0.51) = 0.119$.

The approximation will be better if we use $\pi(\bar{c}, \bar{x}) = 0.674$ for 0.51:

$0.478 \times 0.674(1 - 0.674) = 0.105$. 
5 Building and Applying Logistic Regression Models

I Strategies in Model Selection

I.1 Num of $x$’s in a logistic regression model

- # of $x$’s can be entered in the model:
  - Rule of thumb: # of events (both $[Y = 1]$ and $[Y = 0]$) per $x \geq 10$.

- Need to be aware of collinearity in $x$’s.
I.2 Crab data revisited

- If we throw all indep variables to the logistic regression:

\[
\text{logit}\{\pi\} = \alpha + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 s_1 + \beta_5 s_2 + \beta_6 wt + \beta_7 width
\]

The LRT for \(H_0\): all \(\beta\)'s = 0 is 40.6 with \(df = 7\) (p-value < 0.0001).

- However, only \(\beta_2\) is significantly from 0! Something is wrong.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-9.273</td>
<td>3.838</td>
</tr>
<tr>
<td>Color(1)</td>
<td>1.609</td>
<td>0.936</td>
</tr>
<tr>
<td>Color(2)</td>
<td>1.506</td>
<td>0.567</td>
</tr>
<tr>
<td>Color(3)</td>
<td>1.120</td>
<td>0.593</td>
</tr>
<tr>
<td>Spine(1)</td>
<td>-0.400</td>
<td>0.503</td>
</tr>
<tr>
<td>Spine(2)</td>
<td>-0.496</td>
<td>0.629</td>
</tr>
<tr>
<td>Weight</td>
<td>0.826</td>
<td>0.704</td>
</tr>
<tr>
<td>Width</td>
<td>0.263</td>
<td>0.195</td>
</tr>
</tbody>
</table>

- Collinearity is an issue! Wt, width and color are correlated.
I.3 Variable selection

- Use traditional model selection procedures (used when $p << n$)
  1. Forward selection (simple one + variant)
  2. Backward elimination
  3. Better to use LRT for variable selection
  4. Can consider interactions (usually 2-way interactions)

- Use modern model selection procedures, usually in the form of penalized likelihood (can handle $p > n$); New research area.
I.4 Backward elimination for crab data

Table 5.2. Results of Fitting Several Logistic Regression Models to Horseshoe Crab Data

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>Deviance</th>
<th>df</th>
<th>AIC</th>
<th>Models Compared</th>
<th>Deviance Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$C \times S + C \times W + S \times W$</td>
<td>173.7</td>
<td>155</td>
<td>209.7</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$C + S + W$</td>
<td>186.6</td>
<td>166</td>
<td>200.6</td>
<td>(2)–(1)</td>
<td>12.9 (df = 11)</td>
</tr>
<tr>
<td>3a</td>
<td>$C + S$</td>
<td>208.8</td>
<td>167</td>
<td>220.8</td>
<td>(3a)–(2)</td>
<td>22.2 (df = 1)</td>
</tr>
<tr>
<td>3b</td>
<td>$S + W$</td>
<td>194.4</td>
<td>169</td>
<td>202.4</td>
<td>(3b)–(2)</td>
<td>7.8 (df = 3)</td>
</tr>
<tr>
<td>3c</td>
<td>$C + W$</td>
<td>187.5</td>
<td>168</td>
<td>197.5</td>
<td>(3c)–(2)</td>
<td>0.9 (df = 2)</td>
</tr>
<tr>
<td>4a</td>
<td>$C$</td>
<td>212.1</td>
<td>169</td>
<td>220.1</td>
<td>(4a)–(3c)</td>
<td>24.6 (df = 1)</td>
</tr>
<tr>
<td>4b</td>
<td>$W$</td>
<td>194.5</td>
<td>171</td>
<td>198.5</td>
<td>(4b)–(3c)</td>
<td>7.0 (df = 3)</td>
</tr>
<tr>
<td>5</td>
<td>$C = \text{dark} + W$</td>
<td>188.0</td>
<td>170</td>
<td>194.0</td>
<td>(5)–(3c)</td>
<td>0.5 (df = 2)</td>
</tr>
<tr>
<td>6</td>
<td>None</td>
<td>225.8</td>
<td>172</td>
<td>227.8</td>
<td>(6)–(5)</td>
<td>37.8 (df = 2)</td>
</tr>
</tbody>
</table>

*Note: $C =$ color, $S =$ spine condition, $W =$ width.*

The table indicates that model 5 ($M_3$ on slide 241) may be considered the final model.
I.5 Use AIC or BIC for model selection

- AIC formula (smaller, the better):
  \[ AIC = -2 \text{ (log likelihood - \# of parameters in the model)} \]

- AIC “penalizes a bigger model” by its complexity/size.

- For model 5 in Table 5.2, the SAS program and output:

```sas
data crab;
  input color spine width satell weight;
  weight=weight/1000;
  color=color-1;
  y=(satell>0);
  n=1;
  if color<4 then c=1;
  else c=0;

datalines;
3 3 28.3 8 3050
4 3 22.5 0 1550
2 1 26.0 9 2300
....
```

...
proc genmod descending;
  model y/n = width c / dist=bin;
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>170</td>
<td>187.9579</td>
<td>1.1056</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>170</td>
<td>187.9579</td>
<td>1.1056</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>170</td>
<td>167.4557</td>
<td>0.9850</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>170</td>
<td>167.4557</td>
<td>0.9850</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>-93.9789</td>
<td></td>
</tr>
<tr>
<td>Full Log Likelihood</td>
<td></td>
<td>-93.9789</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>193.9579</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>194.0999</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>203.4178</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AIC = $-2(-93.98 - 3) = 193.96 \approx 194$.

• **Note:** Now Proc Genmod and Proc Logistic do not produce Pearson $\chi^2$ and deviance for binary data anymore, unless aggregate=(width c) is used, in which case their df=# of distinct settings determined by width and c - # of parameters in the model.

In the above program, we tricked proc genmod by using y/n so the procedure does not think the data is binary.
I.6 Summarizing predictive power, classification tables and ROC curves

- Suppose we have binary response $Y_i = 1/0$ (success/failure), $x_i$ a vector of covariates.

\[
\pi(x_i) = P[Y_i = 1|x_i]
\]
\[
\text{logit}\{\pi(x_i)\} = x_i^T \beta \quad \text{(can have more than 1 $x$)}
\]

After we fit the model, we got $\hat{\beta} \Rightarrow$ we got $\hat{\pi}_i$ as

\[
\hat{\pi}_i = \frac{e^{x_i^T \hat{\beta}}}{1 + e^{x_i^T \hat{\beta}}}. 
\]

- Choose a known value $\pi_0$ (e.g., $\pi_0 = 0.5$), and conduct prediction $\hat{Y}_i$ as

\[
\hat{Y}_i = \begin{cases} 
1 & \text{if } \hat{\pi}_i > \pi_0 \\
0 & \text{otherwise}
\end{cases}
\]
and then construct the table (classification table)

\[
\begin{array}{c|cc}
\hat{Y} & 0 & 1 \\
\hline 
Y & & \\
\hline 
1 & n_{11} & n_{12} \\
0 & n_{21} & n_{22} \\
\end{array}
\]

The following two quantities tell us how good the prediction is:

sensitivity = \( \frac{n_{11}}{n_{11} + n_{12}} \)

specificity = \( \frac{n_{22}}{n_{21} + n_{22}} \)

- Using only one table with one \( \pi_0 \) loses information.
- Solution: use many different values of \( \pi_0 \) ⇒ many classification tables
  ⇒ many pairs of sensitivity and specificity ⇒ plot sensitivity \( v.s.1 - \) specificity ⇒ ROC (receiver operating characteristic curve) ⇒ Area under the ROC curve summarizes the predictive power of the model, often called the \( c \)-index.
• An example:

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$\hat{\pi}$</th>
<th>$\hat{Y}_{0.3}$</th>
<th>$\hat{Y}_{0.4}$</th>
<th>$\hat{Y}_{0.5}$</th>
<th>$\hat{Y}_{0.6}$</th>
<th>$\hat{Y}_{0.7}$</th>
<th>$\hat{Y}_{0.8}$</th>
<th>$\hat{Y}_{0.8+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\hat{Y}$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

$se = \frac{3}{3}$  $sp = \frac{0}{3}$

$se = \frac{3}{3}$  $sp = \frac{1}{3}$

$se = \frac{2}{3}$  $sp = \frac{1}{3}$

$se = \frac{2}{3}$  $sp = \frac{2}{3}$

$se = \frac{1}{3}$  $sp = \frac{2}{3}$

$se = \frac{1}{3}$  $sp = \frac{3}{3}$

$se = \frac{0}{3}$  $sp = \frac{3}{3}$
ROC curve for the example
• The AUC for the above ROC curve:

\[ 1 - \frac{3}{9} = \frac{2}{3} \]

= proportion of *concordant* pairs in \((Y_i, \hat{\pi}_i)\) among all pairs with different outcome \(Y_i\).

# of pairs with different outcomes: \(3 \times 3 = 9\).
# of concordant pairs: \(3 + 2 + 1 = 6\).
If there are ties in $\hat{\pi}_i$'s, need to do some adjustment. For example, suppose two $\hat{\pi}_i$ for a $Y_i = 1$ and a $Y_i = 0$ are the same (0.4):

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$\hat{\pi}$</th>
<th>$\hat{Y}_{0.4-}$</th>
<th>$\hat{Y}_{0.5-}$</th>
<th>$\hat{Y}_{0.6-}$</th>
<th>$\hat{Y}_{0.7-}$</th>
<th>$\hat{Y}_{0.8-}$</th>
<th>$\hat{Y}_{0.8+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The corresponding classification tables are:

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$\hat{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>1 0</td>
</tr>
<tr>
<td>1</td>
<td>3 0</td>
</tr>
<tr>
<td>0</td>
<td>3 0</td>
</tr>
</tbody>
</table>

$se = \frac{3}{3}$  $se = \frac{2}{3}$  $se = \frac{2}{3}$  $se = \frac{1}{3}$  $se = \frac{1}{3}$  $se = \frac{0}{3}$

$sp = \frac{0}{3}$  $sp = \frac{1}{3}$  $sp = \frac{2}{3}$  $sp = \frac{2}{3}$  $sp = \frac{3}{3}$  $sp = \frac{3}{3}$
**ROC curve when there are tied predictive probs**

![ROC Curve Diagram](image)
\begin{itemize}
\item AUC = \frac{5.5}{9}
\end{itemize}

9 = \# of pairs with diff outcomes
5.5 = \# of concordant pairs (5) + 0.5 \times \# of ties in \hat{\pi}_i’s with diff. outcomes (1).

\begin{itemize}
\item \textbf{Note:} For binomial data, we need to decompose them as binary data.
There will be a lot tied predicted probabilities.
\end{itemize}

\begin{itemize}
\item The program to get \(\hat{\pi}_i\), ROC curve and c-index:
\end{itemize}

\begin{verbatim}
Proc logistic; * may need descending for binary y;
   model y/n = x / outroc=roc;
   output out=outpred predicted=pihat;
run;

title "ROC Plot";
symbol1 v=dot i=join;
proc gplot data=roc;
   plot _sensit_*_1mspec_; 
   run;

here variable _1mspec_ means 1 minus specificity.
\end{verbatim}
• SAS program and output for the logistic model for crab data:

\[ M_3 : \text{logit}\{\pi(x, c)\} = \alpha + \beta_1 c + \beta_2 x \]

```sas
title "ROC Curve and c-index";
proc logistic descending;
   model y = width c / link=logit outroc=roc;
   output out=outpred predicted=pihat;
run;
proc plot data=roc;
   plot _sensit_*_1mspec_; 
run;
```

*************************************************************************
Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-12.9795</td>
<td>2.7272</td>
<td>22.6502</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>width</td>
<td>1</td>
<td>0.4782</td>
<td>0.1041</td>
<td>21.0841</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>1.3005</td>
<td>0.5259</td>
<td>6.1162</td>
<td>0.0134</td>
</tr>
</tbody>
</table>

Association of Predicted Probabilities and Observed Responses

<table>
<thead>
<tr>
<th>Percent Concordant</th>
<th>76.7</th>
<th>Somers’ D</th>
<th>0.544</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Discordant</td>
<td>22.3</td>
<td>Gamma</td>
<td>0.549</td>
</tr>
<tr>
<td>Percent Tied</td>
<td>0.9</td>
<td>Tau-a</td>
<td>0.252</td>
</tr>
<tr>
<td>Pairs</td>
<td>6882</td>
<td>c</td>
<td>0.772</td>
</tr>
</tbody>
</table>
ROC curve from the model:

Plot of _SENSIT_* _1MSPEC_. Legend: A = 1 obs, B = 2 obs, etc.

1 - Specificity
II Model Checking for Logistic Models

II.1 LRT testing current model to more complex models

• Suppose we would like to see if the logistic model (with only one $x$):

$$\log\{\pi(x)\} = \alpha + \beta x$$

fits the data well, we can fit a more complex model such as

$$\log\{\pi(x)\} = \alpha + \beta_1 x + \beta_2 x^2.$$ 

and test $H_0 : \beta_2 = 0$ using the Wald, score and LRT tests. LRT is usually preferred.
II.2 Goodness of fit using deviance and Pearson $\chi^2$ for grouped data

- For binomial data like the Snoring/Heart disease example:

<table>
<thead>
<tr>
<th>$x$</th>
<th>Yes ($y_i$)</th>
<th>No</th>
<th>$n_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Never</td>
<td>24</td>
<td>1355</td>
<td>1379</td>
</tr>
<tr>
<td>Snoring</td>
<td>35</td>
<td>605</td>
<td>638</td>
</tr>
<tr>
<td>2 Occasionally</td>
<td>21</td>
<td>192</td>
<td>213</td>
</tr>
<tr>
<td>4 Nearly every night</td>
<td>30</td>
<td>224</td>
<td>254</td>
</tr>
<tr>
<td>5 Every night</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

where $n_i \to \infty$, we can use the deviance or Pearson $\chi^2$ to check the goodness of fit of the logistic model

$$\text{logit}\{(\pi(x))\} = \alpha + \beta x.$$
• Treat the data as if from $I \times 2$ table, the deviance $G^2(M)$ of the current model $M$ can be shown to have the form:

$$G^2(M) = 2 \sum \text{obs} \times \log \left\{ \frac{\text{obs}}{\text{fitted}} \right\}$$

and the Pearson $\chi^2$ have the form:

$$\chi^2 = \sum \frac{(\text{obs} - \text{fitted})^2}{\text{fitted}}$$

where the summation is over $2I$ cells (8 cells for the previous example)

• For snoring/HD example, we know that linear probability model has a better fit than the logistic model.
II.3 Goodness of fit for ungrouped data, Hosmer-Lemeshow test

- After fitting the logistic regression model for binary data (can be recovered for binomial data), group data into $g$ groups of approximately the same size based on the estimated success probabilities:

<table>
<thead>
<tr>
<th>$Y_1$, $Y_2$, $\cdots$, $Y_{1n_1}$</th>
<th>$\hat{\pi}_1$, $\hat{\pi}<em>2$, $\cdots$, $\hat{\pi}</em>{1n_1}$</th>
<th>$n_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_2$, $Y_2$, $\cdots$, $Y_{2n_2}$</td>
<td>$\hat{\pi}_2$, $\hat{\pi}<em>2$, $\cdots$, $\hat{\pi}</em>{2n_2}$</td>
<td>$n_2$</td>
</tr>
<tr>
<td>$\cdots$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_g$, $Y_g$, $\cdots$, $Y_{gn_g}$</td>
<td>$\hat{\pi}_g$, $\hat{\pi}<em>g$, $\cdots$, $\hat{\pi}</em>{gn_g}$</td>
<td>$n_g$</td>
</tr>
</tbody>
</table>
• Then construct the following stat

\[
\sum_{i=1}^{g} \frac{(\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \hat{\pi}_{ij})^2}{(\sum_{j=1}^{n_i} \hat{\pi}_{ij})(n_i - \sum_{j=1}^{n_i} \hat{\pi}_{ij})/n_i} \overset{H_0}{\sim} \chi^2_{g-2} \text{(roughly)},
\]

when the \# of distinct covariate patterns is large.

• This is the Hosmer-Lemeshow test of goodness-of-fit.

• The test can be obtained using

```
Proc Logistic;
   model y/n = x1 x2 / lackfit;
Run;
```
II.4 Residuals from the logistic models

- With data \( y_i \) from Bin\((n_i, \pi_i)\) and we fit the logistic model

\[
\text{logit}(\pi_i) = \alpha + \beta x_i.
\]

After we got \( \hat{\alpha}, \hat{\beta} \Rightarrow \hat{\pi}_i:\)

\[
\hat{\pi}_i = \frac{e^{\hat{\alpha} + x_i \hat{\beta}}}{1 + e^{\hat{\alpha} + x_i \hat{\beta}}}.
\]

- Pearson Residual:

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

- Standardized Pearson residual

\[
e_{st}^i = \frac{y_i - n_i \hat{\pi}_i}{SE} = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)(1 - h_i)}} = \frac{e_i}{\sqrt{1 - h_i}}
\]

where \( h_i \) is the \( i \)th element of the hat matrix.
• $E(e^*_i) \approx 0$, $\text{var}(e^*_i) \approx 1$ for large $n_i$. So $e^*_i$ behaves like a $N(0,1)$ random variable. Large $e^*_i$ ($|e^*_i| > 2$) indicates potential outlier.

• Plots of $e^*_i$ v.s. $x_i$ or $x_i\hat{\beta}$ may detect lack of fit.

• When $n_i = 1$ (binary data), $e^*_i$ is not very informative.

• **Note**: Proc Logistic does not report $e^*_i$. Need to use Proc GenMod to get $e^*_i$. 
Example 1: Residual plot for the crab data:

Model: \( \text{logit}(P[Y = 1|x, c]) = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 x \)

```plaintext
data crab;
    input color spine width satell weight;
    weight=weight/1000;
    color=color-1;
    satbin=(satell>0);
    c1 = (color=1);
    c2 = (color=2);
    c3 = (color=3);
    c4 = (color=4);
    s1 = (spine=1);
    s2 = (spine=2);
    datalines;
    3 3 28.3 8 3050
    4 3 22.5 0 1550
    2 1 26.0 9 2300
    4 3 24.8 0 2100
    4 3 ...

   proc genmod data=crab descending;
       model satbin = width c1 c2 c3 / dist=bin link=logit;
       output out=resid ResRaw=ResRaw ResChi=ResChi StdReschi=StdReschi;
   run;

   data _null_; set resid;
       file "crab_res";
       put stdreschi width;
   run;
```
Standardized Pearson Residual Plot for Crab Data
Example 2: Admission to Graduate School at UF in 1997-1998 (Table 5.5)

Let $\pi(k, g) = P[\text{admission}|D = k, G = g]$ for department $D = k$ and gender $G = g$. We consider three models:

1. $\pi(k, g) = D_k$: Admission is independent of gender at each department.

2. $\pi(k, g) = D_k + G_g$: Admission-Gender association is the same across departments ($\iff \logit{\pi(k, g)} = D_k + G_g$).

3. $\pi(k, g) = G_g$: Get the marginal Admission-Gender association collapsed over departments.

```
options ls=75 ps=100;

data admit;
  input dept $ gender y yno;
  n = y+yno;
  male=gender-1;
cards;
  anth 1 32 81
  anth 2 21 41
  astr 1 6 0
  astr 2 3 8
  chem 1 12 43
  chem 2 34 110
```
... 

title "Model 1: Logistic model assuming gender and admission are";
title2 "conditional independent given department";
proc genmod;
  class dept;
  model y/n = dept /dist=bin link=logit;
  output out=resid Resraw=Resraw Reschi=Reschi StdReschi=StdReschi;
run;

data resid; set resid;
  keep dept male Resraw Reschi StdReschi;
run;

title "Residuals from Model 1";
proc print data=resid;
run;

title "Model 2: Logistic model with homogeneous GA and DA association";
proc genmod data=admit;
  class dept;
  model y/n = dept male;
run;

title "Model 3: Logistic model for marginal GA association";
proc genmod data=admit;
  model y/n = male;
run;
Model 1: Logistic model assuming gender and admission are conditional independent given department.

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>23</td>
<td>44.7352</td>
<td>1.9450</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>23</td>
<td>44.7352</td>
<td>1.9450</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>23</td>
<td>40.8523</td>
<td>1.7762</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>23</td>
<td>40.8523</td>
<td>1.7762</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>dept</th>
<th>male</th>
<th>Reschi</th>
<th>Resraw</th>
<th>Std Reschi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>anth</td>
<td>0</td>
<td>-0.45509</td>
<td>-2.22286</td>
<td>-0.76457</td>
</tr>
<tr>
<td>2</td>
<td>anth</td>
<td>1</td>
<td>0.61438</td>
<td>2.22286</td>
<td>0.76457</td>
</tr>
<tr>
<td>3</td>
<td>astr</td>
<td>0</td>
<td>2.30940</td>
<td>2.82353</td>
<td>2.87096</td>
</tr>
<tr>
<td>4</td>
<td>astr</td>
<td>1</td>
<td>-1.70561</td>
<td>-2.82353</td>
<td>-2.87096</td>
</tr>
<tr>
<td>5</td>
<td>chem</td>
<td>0</td>
<td>-0.22824</td>
<td>-0.71357</td>
<td>-0.26830</td>
</tr>
<tr>
<td>6</td>
<td>chem</td>
<td>1</td>
<td>0.14105</td>
<td>0.71357</td>
<td>0.26830</td>
</tr>
<tr>
<td>7</td>
<td>clas</td>
<td>0</td>
<td>-0.75593</td>
<td>-0.50000</td>
<td>-1.06904</td>
</tr>
<tr>
<td>8</td>
<td>clas</td>
<td>1</td>
<td>0.75593</td>
<td>0.50000</td>
<td>1.06904</td>
</tr>
<tr>
<td>9</td>
<td>comm</td>
<td>0</td>
<td>-0.16670</td>
<td>-1.04167</td>
<td>-0.63260</td>
</tr>
<tr>
<td>10</td>
<td>comm</td>
<td>1</td>
<td>0.61024</td>
<td>1.04167</td>
<td>0.63260</td>
</tr>
<tr>
<td>11</td>
<td>comp</td>
<td>0</td>
<td>0.85488</td>
<td>1.63636</td>
<td>1.15752</td>
</tr>
<tr>
<td>12</td>
<td>comp</td>
<td>1</td>
<td>-0.78040</td>
<td>-1.63636</td>
<td>-1.15752</td>
</tr>
<tr>
<td>13</td>
<td>engl</td>
<td>0</td>
<td>0.67452</td>
<td>3.32130</td>
<td>0.94209</td>
</tr>
<tr>
<td>14</td>
<td>engl</td>
<td>1</td>
<td>-0.65769</td>
<td>-3.32130</td>
<td>-0.94209</td>
</tr>
<tr>
<td>15</td>
<td>geog</td>
<td>0</td>
<td>1.79629</td>
<td>2.75000</td>
<td>2.16641</td>
</tr>
<tr>
<td>16</td>
<td>geog</td>
<td>1</td>
<td>-1.21106</td>
<td>-2.75000</td>
<td>-2.16641</td>
</tr>
<tr>
<td>17</td>
<td>geol</td>
<td>0</td>
<td>-0.21822</td>
<td>-0.30000</td>
<td>-0.26082</td>
</tr>
<tr>
<td>18</td>
<td>geol</td>
<td>1</td>
<td>0.14286</td>
<td>0.30000</td>
<td>0.26082</td>
</tr>
<tr>
<td>19</td>
<td>germ</td>
<td>0</td>
<td>0.89974</td>
<td>0.77273</td>
<td>1.88730</td>
</tr>
<tr>
<td>20</td>
<td>germ</td>
<td>1</td>
<td>-1.65903</td>
<td>-0.77273</td>
<td>-1.88730</td>
</tr>
</tbody>
</table>
Model 2: Logistic model with homogeneous GA and DA association  

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>22</td>
<td>42.3601</td>
<td>1.9255</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>22</td>
<td>42.3601</td>
<td>1.9255</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>22</td>
<td>38.9908</td>
<td>1.7723</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>22</td>
<td>38.9908</td>
<td>1.7723</td>
</tr>
</tbody>
</table>

Slide 276
## Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-2.0323</td>
<td>0.2877</td>
<td>-2.5962 -1.4685</td>
<td>49.91</td>
</tr>
<tr>
<td>dept</td>
<td>1</td>
<td>1.2585</td>
<td>0.3277</td>
<td>0.6162 1.9008</td>
<td>14.75</td>
</tr>
<tr>
<td>dept</td>
<td>1</td>
<td>2.2622</td>
<td>0.5631</td>
<td>1.1586 3.3659</td>
<td>16.14</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male</td>
<td>1</td>
<td>-0.1730</td>
<td>0.1123</td>
<td>-0.3932 0.0472</td>
<td>2.37</td>
</tr>
</tbody>
</table>

Model 3: Logistic model for marginal GA association

### 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>44</td>
<td>449.3122</td>
<td>10.2116</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>44</td>
<td>449.3122</td>
<td>10.2116</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>44</td>
<td>409.4050</td>
<td>9.3047</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>44</td>
<td>409.4050</td>
<td>9.3047</td>
</tr>
</tbody>
</table>

Models 2 & 3 show *Simpson’s Paradox.*
Example 3: Heart disease and blood pressure (Table 5.6, P. 151)

```sas
data HD;
  input bp $ n y;
  if bp="<117" then x=111.5;
  else if bp="117-126" then x=121.5;
  else if bp="127-136" then x=131.5;
  else if bp="137-146" then x=141.5;
  else if bp="147-156" then x=151.5;
  else if bp="157-166" then x=161.5;
  else if bp="167-186" then x=176.5;
  else x=191.5;
cards;
  <117 156 3
  117-126 252 17
  127-136 284 12
  137-146 271 16
  147-156 139 12
  157-166 85  8
  167-186 99  16
  >186  43  8
;

proc genmod;
  model y/n = x /dist=bin link=logit residual;
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>6</td>
<td>5.9092</td>
<td>0.9849</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>6</td>
<td>5.9092</td>
<td>0.9849</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>6</td>
<td>6.2899</td>
<td>1.0483</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>6</td>
<td>6.2899</td>
<td>1.0483</td>
</tr>
</tbody>
</table>

# Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-6.0820</td>
<td>0.7243</td>
<td>-7.5017 -4.6624</td>
<td>70.51</td>
</tr>
<tr>
<td>x</td>
<td>1</td>
<td>0.0243</td>
<td>0.0048</td>
<td>0.0148 0.0338</td>
<td>25.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation</th>
<th>Raw Residual</th>
<th>Pearson Residual</th>
<th>Deviance Residual</th>
<th>Std Deviance</th>
<th>Std Pearson Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.194866</td>
<td>-0.979434</td>
<td>-1.061683</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.198648</td>
<td>-1.105788</td>
<td>-1.179257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6.3932374</td>
<td>2.0057053</td>
<td>1.8501072</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.1903838</td>
<td>2.3745999</td>
<td>2.2447199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-3.072737</td>
<td>-0.813338</td>
<td>-0.841966</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.978546</td>
<td>-0.945274</td>
<td>-0.970016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-2.081617</td>
<td>-0.50673</td>
<td>-0.51623</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.583485</td>
<td>-0.572747</td>
<td>-0.581169</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.3836399</td>
<td>0.1175816</td>
<td>0.1170016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1254648</td>
<td>0.1260868</td>
<td>0.1255461</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-0.856987</td>
<td>-0.304247</td>
<td>-0.308775</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.330927</td>
<td>-0.326074</td>
<td>-0.330303</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.791237</td>
<td>0.5134723</td>
<td>0.5049657</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.6411542</td>
<td>0.651955</td>
<td>0.6452766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>-0.361958</td>
<td>-0.139464</td>
<td>-0.140243</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.178337</td>
<td>-0.177346</td>
<td>-0.177959</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
III Sparse Data

III.1 Complete separation and quasi-complete separation

- Consider the following data set:

<table>
<thead>
<tr>
<th>Obs</th>
<th>x1</th>
<th>x2</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

There is a complete separation in $x_1$, and quasi-complete separation in $x_2$.

- What would happen if we fit

$$M_1 : \text{logit}(\pi_i) = \alpha + \beta x_{1i}$$

and

$$M_2 : \text{logit}(\pi_i) = \alpha + \beta x_{2i}?$$
Complete separation in $x_1$

If we fit $M_1$, $\alpha \to -\infty$, $\beta \to \infty$.

How about $M_2$?
III.2 Sparse $2 \times 2 \times K$ tables

Table 5.7. Clinical Trial Relating Treatment ($X$) to Response ($Y$) for Five Centers ($Z$), with $XY$ and $YZ$ Marginal Tables

<table>
<thead>
<tr>
<th>Center ($Z$)</th>
<th>Treatment ($X$)</th>
<th>Response ($Y$)</th>
<th>$YZ$ Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Success</td>
<td>Failure</td>
</tr>
<tr>
<td>1</td>
<td>Active drug</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Active drug</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Active drug</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Active drug</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Active drug</td>
<td>5</td>
<td>9</td>
</tr>
</tbody>
</table>

$XY$ Marginal: Active drug: 12, 36
Placebo: 4, 42

Source: Diane Connell, Sandoz Pharmaceuticals Corp.
As we see before, we may not be interested in $XY$ marginal association. Instead should focus on conditional association.

Consider logistic model for $\pi(x, z) = P[Y = 1|x, z]$:

$$\text{logit}\{\pi(x, z)\} = \beta x + \beta_k^Z$$

$x = 1/0$ for active drug/placebo, $k = 1, 2, 3, 4, 5$ for 5 centers. Common odds-ratio $\theta_{XY|Z} = e^{\beta}$ across centers.

SAS program and part of the output:

```sas
data fungal;
  input center trt y y0;
  n=y+y0;
  control=1-trt;
cards;
  1 1 0 5
  1 0 0 9
  2 1 1 12
  2 0 0 10
  3 1 0 7
  3 0 0 5
  4 1 6 3
  4 0 2 6
  5 1 5 9
  5 0 2 12;
```

Slide 283
**From the output, we know that for centers 1 & 3, $\hat{\beta}_k^Z = -\infty$.**

- $\hat{\beta} = 1.546, \overline{SE}(\hat{\beta}) = 0.702$, p-value from Wald test = 0.0276. May not be valid!
IV Conditional Logistic Models and Exact Inference

IV.1 Conditional logistic regression for $2 \times 2 \times K$ tables

- If the number of centers $K$ is large in the previous common odds-ratio example:
  \[
  \text{logit}\{\pi(x, z)\} = \beta x + \beta^Z_k, \quad z = 1, 2, ..., K
  \]
  then there will be too many $\beta^Z_k$'s and the ML inference on $\beta$ may not be valid.

- Idea: find out sufficient statistics of $\beta_k$ and conduct inference on $\beta$ based on the conditional distribution of the data given those sufficient statistics.
- Data from center $k$:

\[
\begin{array}{ccc}
Y \\
S & F \\
X \text{ trt} & n_{11k} & n_{12k} & n_{1+k} \\
control & n_{21k} & n_{22k} & n_{2+k} \\
Z = k
\end{array}
\]

- It can be shown that $n_{+1k} = n_{11k} + n_{21k}$ (total # of successes at center $k$) is a sufficient statistic for $\beta_k$.

\[
\Rightarrow L_k(\beta, \beta_k|n_{+1k}) = L_k(\beta|n_{+1k}) \text{ should be free of } \beta_k - \text{non-central hypergeometric dist.}
\]

When $\beta = 0 (X \perp Y|Z)$, $L_k(\beta|n_{+1k})$ is the standard hypergeometric dist. with no unknown parameter.
• The conditional logistic inference (on $\beta$) is based on the conditional likelihood:

$$L_c(\beta|\{n_{+1k}\}) = \prod_{k=1}^{K} L_k(\beta, \beta_k|n_{+1k}),$$

which only has one parameter $\beta$ no matter how large $K$ is!

Treat this as a regular likelihood function, we can estimate $\beta$ by maximizing $L_c(\beta|\{n_{+1k}\})$. We can also conduct the Wald, score and LRT for testing $H_0 : \beta = 0$. 
• SAS program and output:

```sas
title "Use a conditional logistic regression to assess treatment effect";
proc logistic data=fungal;
   class center;
   model y/n = trt;
   strata center;
run;
```

The LOGISTIC Procedure

Conditional Analysis

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>5.2269</td>
<td>1</td>
<td>0.0222</td>
</tr>
<tr>
<td>Score</td>
<td>5.0170</td>
<td>1</td>
<td>0.0251</td>
</tr>
<tr>
<td>Wald</td>
<td>4.6507</td>
<td>1</td>
<td>0.0310</td>
</tr>
</tbody>
</table>

Analysis of Conditional Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>1</td>
<td>1.4706</td>
<td>0.6819</td>
<td>4.6507</td>
<td>0.0310</td>
</tr>
</tbody>
</table>

• However, since the tables are sparse, all three tests may not be valid ⇒ exact conditional inference!
IV.2 Exact conditional inference for $2 \times 2 \times K$ tables

- With common odds-ratio model for $2 \times 2 \times K$ tables

\[
\text{logit}\{\pi(x, z)\} = \beta x + \beta^Z_k, \quad z = 1, 2, \ldots, K
\]

The conditional likelihood of $\beta$ only depends on $\beta$.

- Under $H_0 : \beta = 0(X \perp Y|Z)$, the conditional likelihood $L_k(\beta|n_{+1k})$ is completely known, and is equal to the conditional distribution of $n_{11k}$ given all the margins — hypergeometric dist.

- We can conduct exact inference for $H_0 : \beta = 0(X \perp Y|Z)$ using this hypergeometric dist.
**SAS program and part of the output:**

```sas
proc logistic data=fungal;
   class center / param=ref;
   model y/n = center trt;
   exact trt;
run;
```

The LOGISTIC Procedure

Exact Conditional Tests

<table>
<thead>
<tr>
<th>Effect</th>
<th>Test</th>
<th>Statistic</th>
<th>Exact</th>
<th>Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>trt</td>
<td>Score</td>
<td>5.0170</td>
<td>0.0333</td>
<td>0.0235</td>
</tr>
<tr>
<td>Probability</td>
<td></td>
<td>0.0197</td>
<td>0.0333</td>
<td>0.0235</td>
</tr>
</tbody>
</table>

**Note:** Since the above exact test is based on the conditional dist. of $n_{11k}$ given margins, which is the dist that CMH test is based, it can be shown that the above exact score test is actually the exact CMH test! Compare this to the large-sample CMH test on the next slide.

```sas
data y1; set fungal;
   count=y;
   drop y0;
   y=1;
run;
```
```sas
data y0; set fungal;
count=y0;
drop y0;
y=0;
run;
data new; set y1 y0;
run;
title "MH test for conditional independence and MH common OR";
proc freq data=new order=data;
   weight count;
   tables center*trt*y/nopercent norow nocol cmh;
run;
```

MH test for conditional independence and MH common OR

The FREQ Procedure

Summary Statistics for trt by y
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>5.0170</td>
<td>0.0251</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>1</td>
<td>5.0170</td>
<td>0.0251</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>1</td>
<td>5.0170</td>
<td>0.0251</td>
</tr>
</tbody>
</table>
IV.3 Other exact conditional test in logistic models

- For a logistic model:

\[
\logit\{\pi(x)\} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p
\]

We can find out suff. stat. for each \(\beta_k\), denoted by \(T_k\). Suppose we would like to make exact conditional inference on, \(\beta_p\), say, then the exact inference can be based on

\[
f(y_1, y_2, \ldots, y_n | T_1, T_2, \ldots, T_{p-1}) = \tilde{L}(\beta_p).
\]

For exact test of \(H_0 : \beta_p = 0\), the cond. dist. of data \((Y_1, Y_2, \ldots, Y_n)\) given \(T_1, T_2, \ldots, T_{p-1}\) is completely known. We can do exact score test based on \(\tilde{L}(\beta_p)\).

We can also construct an exact CI for \(\beta_p\) based on \(\tilde{L}(\beta_p)\).

Software:

Proc Logistic; *may use "descending" for binary response;
  model y/n = x1 x2 x3 / link=logit;
  exact x3;
run;
• **Fisher’s Exact Test:** We can consider a logistic model

\[
\logit(P[Y = 1]) = \alpha + \beta x
\]

for the following $2 \times 2$ table:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>$y_1$</td>
<td>$n_1 - y_1$</td>
</tr>
<tr>
<td>0</td>
<td>$y_2$</td>
<td>$n_2 - y_2$</td>
</tr>
</tbody>
</table>

It can be shown that a sufficient statistic of $\alpha$ is $y_1 + y_2$ – the column margin. Then the Fisher’s exact test can be achieved by

```r
Proc Logistic;
  model y/n = x / link=logit;
  exact x;
run;
```
• **Exact Cochran-Armitage trend test**: If there is only one ordinal $x$ (with score denoted by $x$), then we conduct the exact test for $\beta = 0$ in the following logistic regression:

$$\logit\{\pi(x)\} = \alpha + \beta x.$$ 

It can be shown that the resulting exact score test is the exact Cochran-Armitage trend test.

• **Example**: Mother’s alcohol consumption and infant malformation

<table>
<thead>
<tr>
<th>Alcohol Consumption</th>
<th>Malformation</th>
<th>Present ($Y = 1$)</th>
<th>Absent ($Y = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0)</td>
<td></td>
<td>48</td>
<td>17,066</td>
</tr>
<tr>
<td>&lt; 1 (0.5)</td>
<td></td>
<td>38</td>
<td>14,464</td>
</tr>
<tr>
<td>1 – 2 (1.5)</td>
<td></td>
<td>5</td>
<td>788</td>
</tr>
<tr>
<td>3 – 5 (4)</td>
<td></td>
<td>1</td>
<td>126</td>
</tr>
<tr>
<td>≥ 6 (7)</td>
<td></td>
<td>1</td>
<td>37</td>
</tr>
</tbody>
</table>
• SAS program and part of the output:

```sas
data table2_7;
  input alcohol malform count @@;
datalines;
  0 1 48 0 0 17066
  0.5 1 38 0.5 0 14464
  1.5 1 5 1.5 0 788
  4 1 1 4 0 126
  7 1 1 7 0 37
;
title "Exact Cochran-Armitage trend test";
proc logistic;
  freq count;
  model malform (event="1") = alcohol / link=logit;
  * equivalent to model malform (ref="0") = alcohol / link=logit;
  exact alcohol;
run;
```

The LOGISTIC Procedure

Exact Conditional Tests

<table>
<thead>
<tr>
<th>Effect</th>
<th>Test</th>
<th>Statistic</th>
<th>Exact</th>
<th>Mid</th>
</tr>
</thead>
<tbody>
<tr>
<td>alcohol</td>
<td>Score</td>
<td>6.5699</td>
<td>0.0172</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>Probability</td>
<td>0.00291</td>
<td>0.0217</td>
<td>0.0202</td>
</tr>
</tbody>
</table>

The exact Cochran-Armitage trend test has p-value = 0.0172 (mid p-value=0.0158) ⇒ significant evidence for alcohol effect on infant malformation!

Slide 295
Sample Size Calculation for Comparing Two Proportions

- Sample size calculation is usually posed as a hypothesis testing problem. For comparing two success probabilities $\pi_1$ and $\pi_2$ from two groups, the null hypothesis is $H_0 : \pi_1 = \pi_2$ and the alternative is $H_a : \pi_1 \neq \pi_2$.

- Suppose we have data: $y_1 \sim \text{Bin}(n_1, \pi_1)$ and $y_2 \sim \text{Bin}(n_2, \pi_2)$, we would construct a test statistic

$$T = \frac{p_1 - p_2}{\sqrt{p_1 (1 - p_1)/n_1 + p_2 (1 - p_2)/n_2}},$$

where $p_1 = y_1/n_1, p_2 = y_2/n_2$, and reject $H_0 : \pi_1 = \pi_2$ at level $\alpha$ if

$$|T| \geq z_{\alpha/2},$$

when both $n_1$ and $n_2$ are large.
• If we would like to have power $1 - \beta$ to detect a difference $\delta = \pi_1 - \pi_2$ (w.l.o.g, assume $\delta > 0$), then we need

$$P[T \geq z_{\alpha/2}|H_a : \pi_1 - \pi_2 = \delta] = 1 - \beta.$$ 

• Assume equal sample size for each group: $n_1 = n_2$, then the above power statement leads to (approximately)

$$P\left[ \frac{p_1 - p_2 - \delta}{\sqrt{\pi_1(1 - \pi_1)/n_1 + \pi_2(1 - \pi_2)/n_1}} \geq z_{\alpha/2} - \frac{\delta}{\sqrt{\pi_1(1 - \pi_1)/n_1 + \pi_2(1 - \pi_2)/n_1}} \right] = 1 - \beta$$

$\Rightarrow$

$$P[Z \geq z_{\alpha/2} - \delta \sqrt{n_1}/\sqrt{\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)}] = 1 - \beta$$

where $Z \sim N(0, 1)$. 
\[ z_{\alpha/2} - \delta \sqrt{n_1} / \sqrt{\pi_1 (1 - \pi_1) + \pi_2 (1 - \pi_2)} = -z_\beta \]

\[ n_1 = n_2 = \frac{(z_{\alpha/2} + z_\beta)^2 [\pi_1 (1 - \pi_1) + \pi_2 (1 - \pi_2)]}{(\pi_1 - \pi_2)^2}. \]

- For example, if we would like to detect \( H_a : \pi_1 = 0.3, \pi_2 = 0.2 \) with 90\% power at level 0.05, then

\[
n_1 = n_2 = \frac{(z_{0.05/2} + z_{0.1})^2 [0.3(1 - 0.3) + 0.2(1 - 0.2)]}{(0.3 - 0.2)^2} = \frac{(1.96 + 1.28)^2 [0.3(1 - 0.3) + 0.2(1 - 0.2)]}{(0.3 - 0.2)^2} = 388.4 = 389. \]

- **Note:** The textbook also discussed the sample size calculation in detecting \( \beta \) for a logistic regression model (p.161-162).
6 Multicategory Logit Models

I Logit Models for Nominal Response $Y$

I.1 Baseline-category logit models

- Nominal response $Y$ has $J > 2$ levels:

$$
\begin{array}{cccc}
Y \\
1 & 2 & \cdots & J \\
\end{array}
$$

- Given data $(x_i, y_i)$, let

$$
\begin{align*}
\pi_1(x_i) &= P[Y_i = 1 | x_i] \\
\pi_2(x_i) &= P[Y_i = 2 | x_i] \\
\vdots \\
\pi_J(x_i) &= P[Y_i = J | x_i] \\
\pi_1(x_i) + \pi_2(x_i) + \cdots + \pi_J(x_i) &= 1 \text{ for any } x_i.
\end{align*}
$$
• We would like to model the relationship between
\[ \{\pi_1(x_i), \pi_2(x_i), \cdots, \pi_J(x_i)\} \] and \( x_i \).

• We need to pick up a cat. as a reference cat. We can pick anyone. Let
us pick cat. J as the ref. cat. and model \( \frac{\pi_j(x)}{\pi_J(x)} \) as:

\[
\log \left\{ \frac{\pi_1(x_i)}{\pi_J(x_i)} \right\} = \alpha_1 + \beta_1 x_i \\
\log \left\{ \frac{\pi_2(x_i)}{\pi_J(x_i)} \right\} = \alpha_2 + \beta_2 x_i \\
\log \left\{ \frac{\pi_{J-1}(x_i)}{\pi_J(x_i)} \right\} = \alpha_{J-1} + \beta_{J-1} x_i
\]

– Baseline-category logit model.

**Note:** Each quantity on the LHS is a *generalized logit*. \( \frac{\pi_1(x_i)}{\pi_J(x_i)} \) is the conditional odds that \( Y_i \) is in cell 1 *v.s.* that \( Y_i \) is in cell \( J \) given that \( Y_i \) is in either cell 1 or cell \( J \).
- Given the baseline-category logit model, we can compare any 2 categories. For example,

\[
\log \left\{ \frac{\pi_1(x_i)}{\pi_2(x_i)} \right\} = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)x_i
\]

- We can also find out \(\pi_j(x)\) for any \(j\) with any \(x\):

\[
\pi_1(x) = \pi_J(x)e^{\alpha_1 + \beta_1 x}
\]

\[
\pi_2(x) = \pi_J(x)e^{\alpha_2 + \beta_2 x}
\]

\[
\cdots
\]

\[
\pi_{J-1}(x) = \pi_J(x)e^{\alpha_{J-1} + \beta_{J-1} x}
\]

\[
\pi_1(x) + \pi_2(x) + \cdots + \pi_J(x) = 1
\]

\[
\Rightarrow \quad \pi_j(x) = \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}}
\]

\[
\Rightarrow \quad \pi_j(x) = \frac{e^{\alpha_j + \beta_j x}}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}} \quad j = 1, 2, \ldots, J - 1.
\]
• Data structure needed for fitting the baseline-category logit model using SAS:

At \( x_i \), suppose there are \( n_i = n_{i+} \) subjects such that

\[
\begin{array}{c|c|c|c}
Y & 1 & 2 & \cdots & J \\
\hline
n_{i1} & n_{i2} & \cdots & n_{iJ} \\
\end{array}
\]

\((n_{i1}, n_{i2}, \cdots, n_{iJ})^T \sim \text{Multinomial}\{n_i, \pi_1(x_i), \pi_2(x_i), \ldots, \pi_J(x_i)\}\]

where \( \pi_j(x_i)'s \) are determined by the baseline-category logit model (functions of \( \alpha_j's \) and \( \beta_j's \))
For example: $N = 7$, $J = 3$, $x = \text{age}$:

<table>
<thead>
<tr>
<th>$y$</th>
<th>count</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

If $n_i = 1$, then we don’t need the variable count.
• **Software:**

```plaintext
Proc Logistic;
  freq count;
  model y (ref="1") = x / link=glogit aggregate=(x) scale=none;
run;
```

**Note:** We can use other category as the reference.

• When \( I \), the \# of settings determined by \( x \) is fixed and \( n_i \to \infty \), we can use the Pearson \( \chi^2 \) or the deviance \( G^2 \) for the goodness-of-fit of the baseline-category logit model.

\( df \) for the Pearson \( \chi^2 \) or the deviance \( G^2 \):

\[
df = \# \text{ of free parameters under saturated model} - \# \text{ of free parameters under fitted model}
\]

\# of free parameters under saturated model = \( I \times (J - 1) \)

\# of free parameters under fitted model = \( (J - 1) + (J - 1) \times \text{dim}(x) \)

\( df \) of the Pearson \( \chi^2 \) or \( G^2 \) = \( (J - 1) \times (I - 1 - \text{dim}(x)) \).
I.2 Example: Alligator food choice

- Alligators’ food choice: Fish (F), Invertebrates (I), Others (O)

<table>
<thead>
<tr>
<th>Alligator Size (Meters)</th>
<th>Primary Food Choice, (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.24 I</td>
<td>1.30 I</td>
</tr>
<tr>
<td>1.45 I</td>
<td>1.45 O</td>
</tr>
<tr>
<td>1.63 I</td>
<td>1.65 O</td>
</tr>
<tr>
<td>1.78 I</td>
<td>1.78 I</td>
</tr>
<tr>
<td>1.98 I</td>
<td>2.03 F</td>
</tr>
<tr>
<td>2.36 F</td>
<td>2.39 F</td>
</tr>
<tr>
<td>2.79 F</td>
<td>2.84 F</td>
</tr>
<tr>
<td>3.68 O</td>
<td>3.71 F</td>
</tr>
</tbody>
</table>

\(a\) \(F = \text{Fish}, I = \text{Invertebrates}, O = \text{Other.}\)

*Source:* Thanks to M. F. Delany and Clint T. Moore for these data.

- Want to see how alligators’ size (length) affects their food choice.
• Consider baseline-category logit model with food=others as the reference category:

```r
data gator;
  input length food $ @@;
datalines;
1.24 I 1.30 I 1.30 I 1.32 F 1.32 F 1.40 F 1.42 I 1.42 F
1.45 I 1.45 0 1.47 I 1.47 F 1.50 I 1.52 I 1.55 I 1.60 I
1.63 I 1.65 0 1.65 I 1.65 F 1.65 F 1.68 F 1.70 I 1.73 0
1.78 I 1.78 I 1.78 0 1.80 I 1.80 F 1.85 F 1.88 I 1.93 I
1.98 I 2.03 F 2.03 F 2.16 F 2.26 F 2.31 F 2.31 F 2.36 F
2.36 F 2.39 F 2.41 F 2.44 F 2.46 F 2.56 0 2.67 F 2.72 I
2.79 F 2.84 F 3.25 0 3.28 0 3.33 F 3.56 F 3.58 F 3.66 F
3.68 0 3.71 F 3.89 F
;
proc logistic;
  model food (ref="0") = length / link=glogit aggregate scale=none;
run;
```

• Since “O” is the last category, by default it is the reference category. So ref=’’0’’ is not needed. We keep it in the program to make it more specific.
The LOGISTIC Procedure

Model Information

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>food</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>I</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>O</td>
<td>8</td>
</tr>
</tbody>
</table>

Logits modeled use food='O' as the reference category.

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>75.1140</td>
<td>86</td>
<td>0.8734</td>
<td>0.7929</td>
</tr>
<tr>
<td>Pearson</td>
<td>80.1879</td>
<td>86</td>
<td>0.9324</td>
<td>0.6563</td>
</tr>
</tbody>
</table>

Number of unique profiles: 45

Type 3 Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>2</td>
<td>8.9360</td>
<td>0.0115</td>
</tr>
</tbody>
</table>

- \( df = (45 - 1 - \dim(x)) \times (J - 1) = 43 \times 2 = 86 \). Too large so cannot do goodness of fit test.
CHAPTER 6

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>food</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>F</td>
<td>1</td>
<td>1.6177</td>
<td>1.3073</td>
<td>1.5314</td>
<td>0.2159</td>
</tr>
<tr>
<td>Intercept</td>
<td>I</td>
<td>1</td>
<td>5.6974</td>
<td>1.7938</td>
<td>10.0881</td>
<td>0.0015</td>
</tr>
<tr>
<td>length</td>
<td>F</td>
<td>1</td>
<td>-0.1101</td>
<td>0.5171</td>
<td>0.0453</td>
<td>0.8314</td>
</tr>
<tr>
<td>length</td>
<td>I</td>
<td>1</td>
<td>-2.4654</td>
<td>0.8997</td>
<td>7.5101</td>
<td>0.0061</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>food</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>length</td>
<td>F</td>
<td>0.896</td>
<td>0.325, 2.468</td>
</tr>
<tr>
<td>length</td>
<td>I</td>
<td>0.085</td>
<td>0.015, 0.496</td>
</tr>
</tbody>
</table>

• From the output, we have:

\[
\log(\hat{\pi}_F/\hat{\pi}_O) = 1.618 - 0.110x
\]

\[
\log(\hat{\pi}_I/\hat{\pi}_O) = 5.697 - 2.465x
\]

where \( x \) is the alligator’s length in meters. \( \Rightarrow \)

\[
\log(\hat{\pi}_F/\hat{\pi}_I) = (1.618 - 5.697) + (2.465 - 0.110)x = -4.079 + 2.355x
\]

Among fish and invertebrates, the odds-ratio of choosing fish over invertebrates is \( e^{2.355} = 10.5 \) with one meter increase in length.
• The estimated food choice probabilities as functions of alligator’s length:

\[
\hat{\pi}_F = \frac{e^{1.618-0.110x}}{1 + e^{1.618-0.110x} + e^{5.697-2.465x}}
\]

\[
\hat{\pi}_I = \frac{e^{5.697-2.465x}}{1 + e^{1.618-0.110x} + e^{5.697-2.465x}}
\]

\[
\hat{\pi}_O = \frac{1}{1 + e^{1.618-0.110x} + e^{5.697-2.465x}}
\]
Figure 6.1. Estimated probabilities for primary food choice.
Belief in afterlife from another GSS:

Table 6.4. Belief in Afterlife by Gender and Race

<table>
<thead>
<tr>
<th>Race</th>
<th>Gender</th>
<th>Yes</th>
<th>Undecided</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Female</td>
<td>371</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>250</td>
<td>45</td>
<td>71</td>
</tr>
<tr>
<td>Black</td>
<td>Female</td>
<td>64</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>25</td>
<td>5</td>
<td>13</td>
</tr>
</tbody>
</table>

Source: General Social Survey.

Independence of belief in afterlife \((Y)\) and race, gender \((X)\) can be tested by the Pearson \(\chi^2\) and LRT for contingency table:

Pearson \(\chi^2 = 10.21\) (df=6), p-value=0.12
LRT \(G^2 = 9.60\), (df=6), p-value=0.14.
SAS program and part of output:

```sas
data afterlife;
  input race $ gender $ count1 count2 count3;
  female=(gender="Female");
  white=(race="White");
  racesex=race||gender;
  datalines;
  White Female 371 49 74
  White Male 250 45 71
  Black Female 64 9 15
  Black Male 25 5 13
;

data afterlife; set afterlife;
  array temp {3} count1-count3;
  do y=1 to 3;
    count=temp(y);
    output;
  end;
run;

proc freq data=afterlife;
  weight count;
  tables racesex*y / nocol nopercent chisq;
run;
```
### Table of racesex by y

<table>
<thead>
<tr>
<th>racesex</th>
<th>y</th>
<th>Frequency</th>
<th>Row Pct</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Female</td>
<td></td>
<td>64</td>
<td>72.73</td>
<td>9</td>
<td>15</td>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23</td>
<td></td>
<td>7.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black Male</td>
<td></td>
<td>25</td>
<td>58.14</td>
<td>5</td>
<td>13</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.63</td>
<td></td>
<td>30.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Female</td>
<td></td>
<td>371</td>
<td>75.10</td>
<td>49</td>
<td>74</td>
<td></td>
<td>494</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.98</td>
<td></td>
<td>14.98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Male</td>
<td></td>
<td>250</td>
<td>68.31</td>
<td>45</td>
<td>71</td>
<td></td>
<td>366</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.30</td>
<td></td>
<td>19.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>710</td>
<td>108</td>
<td>173</td>
<td></td>
<td></td>
<td>991</td>
</tr>
</tbody>
</table>

### Statistics for Table of racesex by y

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>6</td>
<td>10.2056</td>
<td>0.1163</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>6</td>
<td>9.5975</td>
<td>0.1427</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>0.2569</td>
<td>0.6123</td>
</tr>
</tbody>
</table>

- **Note:** Mantel-Haenszel $M^2$ is not appropriate.
• Consider baseline-category logit model with main effects only:

\[
\log \left( \frac{\pi_j}{\pi_3} \right) = \alpha_j + \beta_j^G x_1 + \beta_j^R x_2, \quad j = 1, 2,
\]

where \( x_1 \) is the dummy for female, \( x_2 \) is dummy for white.

• SAS program:

```
title "Baseline-category logit model for afterlife data"; proc logistic data=afterlife; freq count; model y (ref="3") = female white / link=glogit aggregate scale=none; run;
```

• Part of the output:

<table>
<thead>
<tr>
<th>Response Profile</th>
<th>Ordered Value</th>
<th>y</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>710</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>173</td>
</tr>
</tbody>
</table>

Logits modeled use \( y=3 \) as the reference category.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.
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>0.8539</td>
<td>2</td>
<td>0.4269</td>
<td>0.6525</td>
</tr>
<tr>
<td>Pearson</td>
<td>0.8609</td>
<td>2</td>
<td>0.4304</td>
<td>0.6502</td>
</tr>
</tbody>
</table>

Number of unique profiles: 4

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>1560.197</td>
<td>1559.453</td>
</tr>
<tr>
<td>SC</td>
<td>1569.994</td>
<td>1588.845</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>1556.197</td>
<td>1547.453</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>8.7437</td>
<td>4</td>
<td>0.0678</td>
</tr>
<tr>
<td>Score</td>
<td>8.8498</td>
<td>4</td>
<td>0.0650</td>
</tr>
<tr>
<td>Wald</td>
<td>8.7818</td>
<td>4</td>
<td>0.0668</td>
</tr>
</tbody>
</table>

Type 3 Analysis of Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>2</td>
<td>7.2074</td>
<td>0.0272</td>
</tr>
<tr>
<td>white</td>
<td>2</td>
<td>2.0824</td>
<td>0.3530</td>
</tr>
</tbody>
</table>
Baseline-category logit model for afterlife data

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>y</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1</td>
<td>0.8828</td>
<td>0.2426</td>
<td>13.2390</td>
<td>0.0003</td>
</tr>
<tr>
<td>Intercept</td>
<td>2</td>
<td>1</td>
<td>-0.7582</td>
<td>0.3614</td>
<td>4.4031</td>
<td>0.0359</td>
</tr>
<tr>
<td>female</td>
<td>1</td>
<td>1</td>
<td>0.4186</td>
<td>0.1713</td>
<td>5.9737</td>
<td>0.0145</td>
</tr>
<tr>
<td>female</td>
<td>2</td>
<td>1</td>
<td>0.1051</td>
<td>0.2465</td>
<td>0.1817</td>
<td>0.6699</td>
</tr>
<tr>
<td>white</td>
<td>1</td>
<td>1</td>
<td>0.3420</td>
<td>0.2370</td>
<td>2.0814</td>
<td>0.1491</td>
</tr>
<tr>
<td>white</td>
<td>2</td>
<td>1</td>
<td>0.2712</td>
<td>0.3541</td>
<td>0.5863</td>
<td>0.4438</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>y</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>1</td>
<td>1.520</td>
<td>1.086 2.126</td>
</tr>
<tr>
<td>female</td>
<td>2</td>
<td>1.111</td>
<td>0.685 1.801</td>
</tr>
<tr>
<td>white</td>
<td>1</td>
<td>1.408</td>
<td>0.885 2.240</td>
</tr>
<tr>
<td>white</td>
<td>2</td>
<td>1.311</td>
<td>0.655 2.625</td>
</tr>
</tbody>
</table>

- Compared to the saturated model, this model has a good fit (small deviance and Pearson $\chi^2$ - valid for non-sparse contingency tables.
- Gender has a significant overall effect, race is not significant!
• We can estimate the probabilities for the combination of race and gender:

\[
\log \left( \frac{\hat{\pi}_1}{\hat{\pi}_3} \right) = 0.883 + 0.419x_1 + 0.342x_2 \\
\log \left( \frac{\hat{\pi}_2}{\hat{\pi}_3} \right) = -0.758 + 0.105x_1 + 0.271x_2
\]

\[
\hat{\pi}_1 = \frac{e^{0.883+0.419x_1+0.342x_2}}{1 + e^{0.883+0.419x_1+0.342x_2} + e^{-0.758+0.105x_1+0.271x_2}} \\
\hat{\pi}_2 = \frac{e^{-0.758+0.105x_1+0.271x_2}}{1 + e^{0.883+0.419x_1+0.342x_2} + e^{-0.758+0.105x_1+0.271x_2}} \\
\hat{\pi}_3 = \frac{1}{1 + e^{0.883+0.419x_1+0.342x_2} + e^{-0.758+0.105x_1+0.271x_2}}
\]

For example, for white females, \( x_1 = x_2 = 1 \), then

\[
\hat{\pi}_1 = \frac{e^{0.883+0.419+0.342}}{1 + e^{0.883+0.419+0.342} + e^{-0.758+0.105+0.271}} = 0.76.
\]
These estimated probabilities are very close to the sample proportions.

- **Note:** The covariates $x$’s in the baseline-category logit model are not related to the category of $Y$. In economics, $x$’s may be category specific (price to type of cars, cost to transport mode, etc). This is *discrete choice model*. Need to use *Proc Phreg*.

**Table 6.6. Estimated Probabilities for Belief in Afterlife**

<table>
<thead>
<tr>
<th>Race</th>
<th>Gender</th>
<th>Yes</th>
<th>Undecided</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Female</td>
<td>0.76</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.68</td>
<td>0.12</td>
<td>0.20</td>
</tr>
<tr>
<td>Black</td>
<td>Female</td>
<td>0.71</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>0.62</td>
<td>0.12</td>
<td>0.26</td>
</tr>
</tbody>
</table>
II Cumulative Logit Models for Ordinal Response Y

II.1 Cumulative logit models

• Ordinal response $Y$ has $J > 2$ levels (assume $1 < 2 < \cdots < J$):

$$Y \text{ at } x$$

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>$\cdots$</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1(x)$</td>
<td>$\pi_2(x)$</td>
<td>$\cdots$</td>
<td>$\pi_J(x)$</td>
</tr>
</tbody>
</table>

• Of course, we can fit the Baseline Category Logit model by treating $Y$ as a nominal variable. But we want to take the ordinal scale into account for a better power.

• One way is to model the cumulative probabilities:

$$\tau_j(x) = P[Y \leq j|x] = \pi_1(x) + \pi_2(x) + \cdots + \pi_j(x), \quad j = 1, 2, \ldots, J-1,$$
and consider a logistic model for $\tau_j(x)$:

$$\log \left\{ \frac{\tau_j(x)}{1 - \tau_j(x)} \right\} = \alpha_j + \beta x, \quad j = 1, 2, \ldots, J - 1$$

This is called a cumulative logit model.

- **Note 1**: We have a logistic model for each cumulative probability $\tau_j$ ($j = 1, 2, \ldots, J - 1$) with different intercepts and the same $\beta$. So a cumulative logit model actually consists of $J - 1$ logistic models.

- **Note 2**: If the above model is correct, then we can pick any $j$ and define a success $\Leftrightarrow [Y \leq j]$, then we can fit a logistic model to the reduced data to make inference on $\beta$. This approach is less efficient.

- Since $\tau_1(x) < \tau_2(x) < \ldots < \tau_{J-1}(x)$ for any $x$, so the intercepts $\alpha_j$'s have to satisfy

$$\alpha_1 < \alpha_2 < \cdots < \alpha_{J-1}.$$
II.2 Interpretation of $\beta$, proportional odds, probability expression

- Interpretation of $\beta$ – similar to a regular logistic regression:
  The odds of the event $[Y \leq j]$ at $x + 1$ is $e^\beta$ times the odds of event $[Y \leq j]$ at $x$ (while other covariates held fixed) for any cut point $j$:

$$\frac{\tau_j(x + 1)/\{1 - \tau_j(x + 1)\}}{\tau_j(x)/\{1 - \tau_j(x)\}} = e^\beta, \quad j = 1, 2, ..., J - 1.$$

$\Rightarrow$ proportional odds model.

- Data structure: the data is organized in exactly the same way as for a nominal response, or each record can represent one subject’s information ($n_i = 1$).

- Software (assume $1 < 2 < \cdots < J$ for $Y$, model $P[Y \leq j]$):
  
  ```
  Proc Logistic; * default is cumulative probs over lower cat;
  freq count; * you dont need this line if ni=1;
  model y = x; * y is the values for categories;
  run;
  ```
The expression of $\tau_j(x)$ and $\pi_j(x)$:

$$
\tau_j(x) = \frac{e^{\alpha_j + \beta x}}{1 + e^{\alpha_j + \beta x}}, \quad j = 1, 2, \ldots, J - 1
$$

$$
\Rightarrow
$$

$$
\pi_1(x) = \tau_1(x)
$$

$$
\pi_2(x) = \tau_2(x) - \tau_1(x)
$$

$$
\pi_j(x) = \tau_j(x) - \tau_{j-1}(x)
$$

$$
\pi_{J-1}(x) = \tau_{J-1}(x) - \tau_{J-2}(x)
$$

$$
\pi_J(x) = 1 - \tau_{J-1}(x)
$$
### II.3 Example: Political ideology and party affiliation

- Table 6.7 from a GSS:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Political Party</th>
<th>Very Liberal</th>
<th>Slightly Liberal</th>
<th>Moderate</th>
<th>Slightly Conservative</th>
<th>Very Conservative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Democratic</td>
<td>44</td>
<td>47</td>
<td>118</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Republican</td>
<td>18</td>
<td>28</td>
<td>86</td>
<td>39</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>Democratic</td>
<td>36</td>
<td>34</td>
<td>53</td>
<td>18</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Republican</td>
<td>12</td>
<td>18</td>
<td>62</td>
<td>45</td>
<td>51</td>
</tr>
</tbody>
</table>

*Source: General Social Survey.*
• Let $Y = 1 < 2 < 3 < 4 < 5$ for 5 categories of political ideology. Define $x = 1/0$ for Democrat/Republican, $z = 1/0$ for male/female and consider cumulative logit model:

$$\text{logit}\{\tau_j(x, z)\} = \alpha_j + \beta_1 x + \beta_2 z + \beta_3 x \times z, \quad j = 1, 2, 3, 4.$$ 

• SAS program and output:

```sas
data ideology;
  input gender $ party $ y1-y5;
  partysex=gender || party;
  x=(party="Democrat");
  z=(gender="Male");
datalines;
Femal Democratic 44 47 118 23 32
Femal Republican 18 28 86 39 48
Male Democratic 36 34 53 18 23
Male Republican 12 18 62 45 51;

data ideology; set ideology;
  array temp {5} y1-y5;
  do y=1 to 5;
    count=temp(y);
    output;
  end;
run;
```
proc freq data=ideology;
  weight count;
  tables partysex*y / nocol nopercent chisq;
run;

***************************************************************************

The FREQ Procedure

Table of partysex by y

<table>
<thead>
<tr>
<th>partysex</th>
<th>y</th>
<th>Frequency</th>
<th>Row Pct</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femal Democrat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>44</td>
<td>47</td>
<td>118</td>
<td>23</td>
<td>32</td>
<td>264</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.67</td>
<td>17.80</td>
<td>44.70</td>
<td>8.71</td>
<td>12.12</td>
<td></td>
</tr>
<tr>
<td>Femal Republic</td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>28</td>
<td>86</td>
<td>39</td>
<td>48</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.22</td>
<td>12.79</td>
<td>39.27</td>
<td>17.81</td>
<td>21.92</td>
<td></td>
</tr>
<tr>
<td>Male Democrat</td>
<td></td>
<td></td>
<td></td>
<td>36</td>
<td>34</td>
<td>53</td>
<td>18</td>
<td>23</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.95</td>
<td>20.73</td>
<td>32.32</td>
<td>10.98</td>
<td>14.02</td>
<td></td>
</tr>
<tr>
<td>Male Republic</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>18</td>
<td>62</td>
<td>45</td>
<td>51</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.38</td>
<td>9.57</td>
<td>32.98</td>
<td>23.94</td>
<td>27.13</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>110</td>
<td>127</td>
<td>319</td>
<td>125</td>
<td>154</td>
<td>835</td>
</tr>
</tbody>
</table>

Statistic  | DF | Value | Prob
----------|----|-------|------
Chi-Square | 12 | 74.2418 | <.0001
Likelihood Ratio Chi-Square | 12 | 74.5433 | <.0001

Slide 325
title "Cumulative logit model for political ideology data"
proc logistic data=ideology;
   freq count;
   model y = x z x*z / aggregate scale=none;
run;

*************************************************************************

The LOGISTIC Procedure

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>y</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>110</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>127</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>319</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>125</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>154</td>
</tr>
</tbody>
</table>

Probabilities modeled are cumulated over the lower Ordered Values.

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.3986</td>
<td>9</td>
<td>0.2494</td>
</tr>
</tbody>
</table>

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>11.0634</td>
<td>9</td>
<td>1.2293</td>
<td>0.2714</td>
</tr>
<tr>
<td>Pearson</td>
<td>11.0876</td>
<td>9</td>
<td>1.2320</td>
<td>0.2698</td>
</tr>
</tbody>
</table>

Number of unique profiles: 4

Slide 326
Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>2541.630</td>
<td>2484.150</td>
</tr>
<tr>
<td>SC</td>
<td>2560.540</td>
<td>2517.242</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>2533.630</td>
<td>2470.150</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>63.4800</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>61.4897</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>61.8399</td>
<td>3</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-2.3082</td>
<td>0.1536</td>
<td>225.8239</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>-1.3112</td>
<td>0.1350</td>
<td>94.3605</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>0.4084</td>
<td>0.1265</td>
<td>10.4257</td>
<td>0.0012</td>
</tr>
<tr>
<td>Intercept 4</td>
<td>1</td>
<td>1.2450</td>
<td>0.1356</td>
<td>84.3507</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x</td>
<td>1</td>
<td>0.7562</td>
<td>0.1669</td>
<td>20.5270</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>z</td>
<td>1</td>
<td>-0.3660</td>
<td>0.1797</td>
<td>4.1495</td>
<td>0.0416</td>
</tr>
<tr>
<td>x*z</td>
<td>1</td>
<td>0.5089</td>
<td>0.2541</td>
<td>4.0111</td>
<td>0.0452</td>
</tr>
</tbody>
</table>
What we see from the output:

1. Without model, the Pearson $\chi^2 = 74.24$ and LRT $G^2 = 74.53$ with $df = (4 - 1)(5 - 1) = 12$ for testing $H_0 : Y \perp \text{gender and party}$.

2. With the model, $H_0 : Y \perp \text{gender and party}$
   $\Leftrightarrow H_0 : \beta_1 = \beta_2 = \beta_3 = 0$. LRT=63.48, Score=61.49, Wald=61.84 with $df = 3$.

3. Fitted model:

   \[
   \logit\{\tau_j(x, z)\} = \hat{\alpha}_j + 0.756x - 0.366z + 0.509x \times z, \quad j = 1, 2, 3, 4
   \]

   \[
   \hat{\alpha}_1 = -2.308,
   \]

   \[
   \hat{\alpha}_2 = -1.311,
   \]

   \[
   \hat{\alpha}_3 = 0.408,
   \]

   \[
   \hat{\alpha}_4 = 1.245.
   \]
4. From the fitted model, the odds-ratio of \( Y \leq j \) (more liberal) between males and females:

\[
\theta_j(x) = e^{-0.366+0.509x}
\]

\[
= \begin{cases} 
  e^{-0.366+0.509} = 1.15 & \text{for Democrats (} x = 1 \text{)} \\
  e^{-0.366+0} = 0.69 & \text{for Republicans (} x = 0 \text{)}
\end{cases}
\]

⇒ Male Democrats tend to be more liberal than female democrats. However, male Republicans are less liberal than female republicans.
5. With fitted model, we can estimate 4 cumulative probabilities:
   
   Female Democrats: $x = 1, z = 0 : \tau'_j s = 0.174, 0.365, 0.762, 0.881$
   
   $\Rightarrow$ cell probs: $\pi'_j s : 0.174, 0.190, 0.397, 0.119, 0.119$

   Female Republicans: $x = 0, z = 0 : \tau'_j s = 0.090, 0.212, 0.601, 0.776$
   
   $\Rightarrow$ cell probs: $\pi'_j s : 0.090, 0.122, 0.388, 0.176, 0.234$

   Male Democrats: $x = 1, z = 1 : \tau'_j s = 0.196, 0.398, 0.787, 0.895$
   
   $\Rightarrow$ cell probs: $\pi'_j s : 0.196, 0.202, 0.389, 0.108, 0.105$

   Male Republicans: $x = 0, z = 1 : \tau'_j s = 0.065, 0.157, 0.510, 0.707$
   
   $\Rightarrow$ cell probs: $\pi'_j s : 0.065, 0.093, 0.353, 0.196, 0.293$

   These cumulative probabilities can also be obtained from proc logistic using statement output out= predicted;
II.4 Model checking for cumulative logit models

- For data in the form of contingency tables with large row margins, the Pearson $\chi^2$ and Deviance statistics can be used to test the goodness of fit of the cumulative logit models. For the political ideology example, the Pearson $\chi^2$ and Deviance are about 11 with $df$

$$ df = I \times (J - 1) - (J - 1 + \text{dim}(x)) $$

$$ = (I - 1)(J - 1) - \text{dim}(x) = (4 - 1)(5 - 1) - 3 = 9. \Rightarrow P\text{-value} = 0.27, \text{reasonably good fit!} $$
• We can also consider a more complicated model with different $\beta$’s for different category $j$ for the same $x$ and conduct a score test. For example, for the political ideology example,

$$H_0 : \text{logit}\{\tau_j(x, z)\} = \alpha_j + \beta_1 x + \beta_2 z + \beta_3 x \times z, \quad j = 1, 2, 3, 4.$$  

$$H_a : \text{logit}\{\tau_j(x, z)\} = \alpha_j + \beta_{1j} x + \beta_{2j} z + \beta_{3j} x \times z, \quad j = 1, 2, 3, 4.$$  

The score statistic is 11.40 with $df$:

$$df = (J - 1) \times \text{dim}(x) - \text{dim}(x) = (J - 2) \times \text{dim}(x) = (5 - 2) \times 3 = 9.$$
II.5 Example with continuous/categorical $x$’s

- Mental impairment example (Table 6.9): 40 subjects.
  - $Y =$ mental impairment, has 4 levels:
    \[
    \begin{array}{cccc}
    & 1 & 2 & 3 & 4 \\
    \hline
    \text{Well} & \text{Mild} & \text{Moderate} & \text{Impaired} \\
    \end{array}
    \]
  - $x_1 =$ life event index (composite # of important life event)
  - $x_2 =$ social-economic status (ses)

Want to study the impact of $x_1$ and $x_2$ on $Y$ using:

\[
\log \frac{P[Y \leq j]}{1 - P[Y \leq j]} = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2, 3.
\]
SAS program and output:

```sas
* SAS program and output:

data mental;
  input mental ses life;
  cards;
  1 1 1
  1 1 9
  1 1 4
  1 1 3
  1 0 2
  1 1 0
  1 0 1
  1 1 3
  1 1 3
  1 1 7
  1 0 1
  1 0 2
  2 1 5
  2 0 6
  2 1 3
  2 0 1
  2 1 8
  2 1 2
  2 0 5
  2 1 5
  2 1 9
  2 0 3
  2 1 3
  2 1 1
  3 0 0
  ...
;
  title "Cumulative logistic model for mental impairment example with main effects only";
  proc logistic; * we use default, may put order=data or descending here;
    * we can put a freq statement here;
    model mental = life ses / aggregate scale=none;
  run;
```
Cumulative logistic model for mental impairment example with main effects

The LOGISTIC Procedure

Probabilities modeled are cumulated over the lower Ordered Values.

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3255</td>
<td>4</td>
<td>0.6761</td>
</tr>
</tbody>
</table>

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>57.6833</td>
<td>52</td>
<td>1.1093</td>
<td>0.2732</td>
</tr>
<tr>
<td>Pearson</td>
<td>57.0248</td>
<td>52</td>
<td>1.0966</td>
<td>0.2937</td>
</tr>
</tbody>
</table>

Number of unique profiles: 19

Model Fit Statistics

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Intercept Only</th>
<th>Intercept and Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>115.042</td>
<td>109.098</td>
</tr>
<tr>
<td>SC</td>
<td>120.109</td>
<td>117.542</td>
</tr>
<tr>
<td>-2 Log L</td>
<td>109.042</td>
<td>99.098</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>9.9442</td>
<td>2</td>
<td>0.0069</td>
</tr>
<tr>
<td>Score</td>
<td>9.1431</td>
<td>2</td>
<td>0.0103</td>
</tr>
<tr>
<td>Wald</td>
<td>8.5018</td>
<td>2</td>
<td>0.0143</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-0.2818</td>
<td>0.6231</td>
<td>0.2045</td>
<td>0.6511</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>1.2129</td>
<td>0.6511</td>
<td>3.4700</td>
<td>0.0625</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>2.2095</td>
<td>0.7171</td>
<td>9.4932</td>
<td>0.0021</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.3189</td>
<td>0.1194</td>
<td>7.1294</td>
<td>0.0076</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>1.1111</td>
<td>0.6143</td>
<td>3.2719</td>
<td>0.0705</td>
</tr>
</tbody>
</table>

Odds Ratio Estimates

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>life</td>
<td>0.727</td>
<td>0.575</td>
</tr>
<tr>
<td>ses</td>
<td>3.038</td>
<td>0.911</td>
</tr>
</tbody>
</table>
Fitted model:

$$\logit P[Y \leq j] = \alpha_j - 0.3189 \times Life + 1.1111 \times SES.$$ 

⇒ The odds for subjects with higher SES to have better mental health is $$e^{1.1111} = 3.038$$ times the odds for subjects lower SES to have better mental health.

⇒ The odds for subjects with one less life event index to have better mental health is $$e^{0.3189} = 1.38$$ times the odds for subjects with one more life event index to have better mental health.
We can estimate all probs for a population defined by \( x_0 \). For example, let us take \( x_1 = \bar{x}_1 = 4.275, \ x_2 = 0 \):

\[
\pi_1 = \frac{e^{-0.2818-0.3189\times4.275}}{1 + e^{-0.2818-0.3189\times4.275}} = 0.1617
\]

\[
\pi_1 + \pi_2 = \frac{e^{1.2129-0.3189\times4.275}}{1 + e^{1.2129-0.3189\times4.275}} = 0.4625
\]

\[
\pi_1 + \pi_2 + \pi_3 = \frac{e^{2.2095-0.3189\times4.275}}{1 + e^{2.2095-0.3189\times4.275}} = 0.7
\]

\[\Rightarrow \pi_4 = 0.3\]

\[\pi_3 = 0.7 - 0.4625 = 0.2375\]

\[\pi_2 = 0.4625 - 0.1617 = 0.3008\]

\[\pi_1 = 0.1617\]
• **Note 1:** The score GOF test for the cumulative logit model

\[
\log \frac{P[Y \leq j]}{1 - P[Y \leq j]} = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2, 3,
\]

has test statistic = 2.33 with \(df\):

\[
df = (J - 2) \times \text{dim}(x) = (4 - 2) \times 2 = 4.
\]

\(\Rightarrow\) P-value = 0.675, good fit!

• **Note 2:** We can also use **Proc GenMod** to fit the above model:

```plaintext
title "Fitting the above cumulative logistic model using proc genmod";
proc genmod; * default is ascending, may put order=data or descending here;
* we can put a freq statement here;
   model mental = life ses / dist=multinomial link=cumlogit
      aggregate=(life ses);
run;
```
Fitting the above cumulative logistic model using proc genmod

The GENMOD Procedure

PROC GENMOD is modeling the probabilities of levels of mental having LOWER Ordered Values in the response profile table. One way to change this to model the probabilities of HIGHER Ordered Values is to specify the DESCENDING option in the PROC statement.

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>52</td>
<td>57.6833</td>
<td>1.1093</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>52</td>
<td>57.6833</td>
<td>1.1093</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>52</td>
<td>57.0245</td>
<td>1.0966</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>52</td>
<td>57.0245</td>
<td>1.0966</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept1</td>
<td>1</td>
<td>-0.2819</td>
<td>0.6423</td>
<td>-1.5407 0.9769</td>
<td>0.19</td>
</tr>
<tr>
<td>Intercept2</td>
<td>1</td>
<td>1.2128</td>
<td>0.6607</td>
<td>-0.0821 2.5076</td>
<td>3.37</td>
</tr>
<tr>
<td>Intercept3</td>
<td>1</td>
<td>2.2094</td>
<td>0.7210</td>
<td>0.7963 3.6224</td>
<td>9.39</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.3189</td>
<td>0.1210</td>
<td>-0.5560 -0.0817</td>
<td>6.95</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>1.1112</td>
<td>0.6109</td>
<td>-0.0861 2.3085</td>
<td>3.31</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

\[ df = (\# \text{of} \{\text{life} \times \text{ses}\} - 1) \times (4 - 1) - 2 = 18 \times 3 - 2 = 52. \]
Note 3: We can also consider the interaction between $x_1$ and $x_2$ and test the significance of $x_1 \times x_2$ using Score, LRT and Wald tests.

```latex
\text{title} "\text{Cumulative logistic model for mental impairment example with interaction}";
\text{proc logistic};
\text{model mental = life ses life*ses;}
\text{run;}
```

***************************************************************************

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>0.0981</td>
<td>0.8110</td>
<td>0.0146</td>
<td>0.9037</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>1.5925</td>
<td>0.8372</td>
<td>3.6186</td>
<td>0.0571</td>
</tr>
<tr>
<td>Intercept 3</td>
<td>1</td>
<td>2.6066</td>
<td>0.9097</td>
<td>8.2111</td>
<td>0.0042</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.4204</td>
<td>0.1903</td>
<td>4.8811</td>
<td>0.0272</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>0.3709</td>
<td>1.1302</td>
<td>0.1077</td>
<td>0.7428</td>
</tr>
<tr>
<td>life*ses</td>
<td>1</td>
<td>0.1813</td>
<td>0.2361</td>
<td>0.5896</td>
<td>0.4426</td>
</tr>
</tbody>
</table>

Wald Test: $\chi^2 = 0.5896$, P-value = 0.4426. Not significant!
• **Note**: The cumulative logit model can be obtained by assuming that there is a (underlying) latent (unobservable) variable $Y^*$ such that

$$Y^* = -\beta x + \varepsilon,$$

where $\varepsilon$ is the error that has a cdf $G(\cdot)$.

* Assume that there are $J - 1$ cut-off points:

$$-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{J-1} < \alpha_J = \infty$$

such that

$$[Y = j] \iff \alpha_{j-1} < Y^* \leq \alpha_j$$
Then
\[
\tau_j(x) = P[Y \leq j|x] \\
= P[Y^* \leq \alpha_j|x] \\
= P[Y^* + \beta x \leq \alpha_j + \beta x|x] \\
= P[\varepsilon \leq \alpha_j + \beta x|x] \\
= G(\alpha_j + \beta x).
\]

If we assume \( \varepsilon \) has a standard logistic distribution, then
\[
G(z) = \frac{e^z}{1+e^z}
\]
and we have
\[
\text{logit}\{\tau_j(x)\} = \alpha_j + \beta x, \quad j = 1, 2, \ldots, J-1.
\]

If we assume \( \varepsilon \) has a standard normal distribution, then
\[
G(z) = \Phi(z)
\]
and we have a cumulative probit model:
\[
\Phi^{-1}\{\tau_j(x)\} = \alpha_j + \beta x, \quad j = 1, 2, \ldots, J-1.
\]
II.6 Invariance to choice of response categories

- If the original cumulative logit model is true for ordinal response $Y = 1 < 2 < \cdots < J$:

$$\text{logit}(\tau_j) = \alpha_j + \beta x,$$

then we can group adjacent categories to form a new category. The resulting ordinal response also has a cumulative logit model with the same $\beta$. A little less efficient.

- For the mental health example

<table>
<thead>
<tr>
<th>$Y$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>Mild</td>
<td>Moderate</td>
<td>Impaired</td>
<td></td>
</tr>
</tbody>
</table>

assume the model:

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta_1 x_1 + \beta_2 x_2, j = 1, 2, 3.$$
Suppose we group the middle 2 categories to form a new category MM:

\[ \tilde{Y} \]

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>MM: Mild or Moderate</td>
<td>Impaired</td>
</tr>
</tbody>
</table>

Then

\[
\logit(P[\tilde{Y} \leq 1]) = \alpha_1 + \beta_1 x_1 + \beta_2 x_2
\]

\[
\logit(P[\tilde{Y} \leq 2]) = \alpha_3 + \beta_1 x_1 + \beta_2 x_2.
\]

So we can fit a cumulative logit model to \( \tilde{Y} \) and will get similar estimates of \( \alpha_1, \alpha_3, \beta_1, \beta_2 \). We cannot estimate \( \alpha_2 \) in the original model.
• SAS program and part of the output:

```sas
data mental2; set mental;
   mental2 = mental;
   if mental2 = 3 then mental2 = 2;
run;

title "Cumulative logit model with middle 2 categories combined";
proc logistic data=mental2;
   * we can put a freq statement here;
   model mental2 = life ses / aggregate scale=none;
run;
```

*********************************************************************************

Score Test for the Proportional Odds Assumption

<table>
<thead>
<tr>
<th>Chi-Square</th>
<th>DF</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1794</td>
<td>2</td>
<td>0.9142</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-0.0468</td>
<td>0.6424</td>
<td>0.0053</td>
<td>0.9420</td>
</tr>
<tr>
<td>Intercept 2</td>
<td>1</td>
<td>2.4812</td>
<td>0.7829</td>
<td>10.0456</td>
<td>0.0015</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.3546</td>
<td>0.1287</td>
<td>7.5916</td>
<td>0.0059</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>0.9326</td>
<td>0.6404</td>
<td>2.1206</td>
<td>0.1453</td>
</tr>
</tbody>
</table>
What we observed:

\[ \hat{\alpha}_1 = -0.0468 (SE = 0.642), \] compared to \(-0.282 (SE = 0.623)\) from the original model.
\[ \hat{\alpha}_3 = 2.482 (SE = 0.783), \] compared to \(2.210 (SE = 0.717)\) from the original model.
\[ \hat{\beta}_1 = -0.355 (SE = 0.129), \] compared to \(-0.319 (SE = 0.119)\) from the original model.
\[ \hat{\beta}_2 = 0.933 (SE = 0.640), \] compared to \(1.111 (SE = 0.614)\) from the original model.

Overall, the original model is more efficient (with smaller SE’s for model parameter estimates), even though the model with combined categories has a better fit! (P-value from score test is 0.9142)
III Paired-Category Logistic Models for Ordinal Response

III.1 Adjacent-category logistic models

- Ordinal response $Y$ has $J > 2$ levels (assume $1 < 2 < \cdots < J$):

  $$Y \text{ at } x$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>$\cdots$</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_1(x)$</td>
<td>$\pi_2(x)$</td>
<td>$\pi_J(x)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- We may consider modeling adjacent logits through

  $$\log \left\{ \frac{\pi_{j+1}(x)}{\pi_j(x)} \right\} = \alpha_j + \beta_j x, \quad j = 1, 2, \ldots, J - 1.$$

  This is equivalent to the baseline-category logit model. We can obtain $\alpha_j, \beta_j$ by running a baseline-category logit model with the $j$th category as the reference category, treating $Y$ as a nominal categorical variable.
• In the above adjacent-category logit model, the slopes $\beta_j$’s are different. We can consider the model with equal slopes:

$$\log \left\{ \frac{\pi_{j+1}(x)}{\pi_j(x)} \right\} = \alpha_j + \beta x, \quad j = 1, 2, \ldots, J - 1.$$ 

$\Rightarrow$ The odds (relative to the adjacent categories) is proportional ($e^\beta$) with one unit increase in $x$.

• Software (currently not available yet):

```latex
proc logistic data=;
   freq count;
   model y = x / link=algorit aggregate scale=none;
run;
```
III.2 Continuation-ratio logistic models

- Ordinal response $Y$ has $J > 2$ levels (assume $1 < 2 < \cdots < J$):

\[
\begin{array}{c|c|c|c}
   & 1 & 2 & \cdots & J \\
\hline
\pi_1(x) & \pi_2(x) & \cdots & \pi_J(x) \\
\end{array}
\]

- We may consider modeling continuation-ratio logits through

\[
\log \left\{ \frac{\pi_1(x)}{\pi_2(x) + \cdots + \pi_J(x)} \right\} = \alpha_1 + \beta_1 x
\]

\[
\log \left\{ \frac{\pi_2(x)}{\pi_3(x) + \cdots + \pi_J(x)} \right\} = \alpha_2 + \beta_2 x
\]

\[
\vdots
\]

\[
\log \left\{ \frac{\pi_{J-1}(x)}{\pi_J(x)} \right\} = \alpha_{J-1} + \beta_{J-1} x
\]
• It can be shown that the MLEs of $\alpha_j$'s and $\beta_j$'s can be obtained by running $J - 1$ separate logistic regression models. The model fit statistic Deviance is the sum of the Deviances from individual models.

• Using mental heath example, we illustrate how to fit a continuation-ratio logit model:

$$\log \left\{ \frac{\pi_1}{\pi_2 + \pi_3 + \pi_4} \right\} = \alpha_1 + \beta_{11} x_1 + \beta_{12} x_2$$

$$\log \left\{ \frac{\pi_2}{\pi_3 + \pi_4} \right\} = \alpha_2 + \beta_{21} x_1 + \beta_{22} x_2$$

$$\log \left\{ \frac{\pi_3}{\pi_4} \right\} = \alpha_3 + \beta_{31} x_1 + \beta_{32} x_2$$
• SAS Program and output:

```sas
data mental; set mental;
  y1 = mental;
  if y1>1 then y1=2;
  y2 = mental;
  if y2>2 then y2=3;
  y3 = mental;
  if y3>3 then y3=4;
run;

title "Model 1: cat 1 vs higher";
proc logistic data=mental;
  model y1=life ses / aggregate scale=none;
run;

title "Model 2: cat 2 vs higher";
proc logistic data=mental;
  where y2 in (2,3);
  model y2=life ses / aggregate scale=none;
run;

title "Model 3: cat 3 vs higher";
proc logistic data=mental;
  where y3 in (3,4);
  model y3=life ses / aggregate scale=none;
run;
```
### Model 1: cat 1 vs higher

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>21.3446</td>
<td>16</td>
<td>1.3340</td>
<td>0.1656</td>
</tr>
<tr>
<td>Pearson</td>
<td>18.3443</td>
<td>16</td>
<td>1.1465</td>
<td>0.3041</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.1729</td>
<td>0.7481</td>
<td>0.0534</td>
<td>0.8173</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.3275</td>
<td>0.1637</td>
<td>4.0029</td>
<td>0.0454</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>1.0064</td>
<td>0.7839</td>
<td>1.6482</td>
<td>0.1992</td>
</tr>
</tbody>
</table>

### Model 2: cat 2 vs higher

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>21.1683</td>
<td>14</td>
<td>1.5120</td>
<td>0.0974</td>
</tr>
<tr>
<td>Pearson</td>
<td>16.8073</td>
<td>14</td>
<td>1.2005</td>
<td>0.2666</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.0660</td>
<td>0.9020</td>
<td>0.0054</td>
<td>0.9417</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.1984</td>
<td>0.1665</td>
<td>1.4204</td>
<td>0.2333</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>1.3782</td>
<td>0.8487</td>
<td>2.6374</td>
<td>0.1044</td>
</tr>
</tbody>
</table>
Model 3: cat 3 vs higher

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>12.8261</td>
<td>8</td>
<td>1.6033</td>
<td>0.1180</td>
</tr>
<tr>
<td>Pearson</td>
<td>10.0481</td>
<td>8</td>
<td>1.2560</td>
<td>0.2617</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.4826</td>
<td>1.2829</td>
<td>1.3356</td>
<td>0.2478</td>
</tr>
<tr>
<td>life</td>
<td>1</td>
<td>-0.3045</td>
<td>0.2264</td>
<td>1.8099</td>
<td>0.1785</td>
</tr>
<tr>
<td>ses</td>
<td>1</td>
<td>-0.4614</td>
<td>1.1496</td>
<td>0.1611</td>
<td>0.6882</td>
</tr>
</tbody>
</table>

- The *Deviance* goodness-of-fit statistics is

\[
deviance = 21.3446 + 21.1683 + 12.8261 = 55.34
\]

\[
df = 16 + 14 + 8 = 38
\]

- Note The adjacent-category logit model and the continuation-ratio logit model are less popular than the cumulative logit model.
IV Tests of Independence & Conditional independence

IV.1 Tests of $X \perp Y$

- Case 1: $X, Y$ – ordinal. Use Table 2.13 as an example:

<table>
<thead>
<tr>
<th></th>
<th>Below average</th>
<th>Average</th>
<th>Above Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not too happy</td>
<td>94</td>
<td>249</td>
<td>83</td>
</tr>
<tr>
<td>Pretty happy</td>
<td>53</td>
<td>372</td>
<td>221</td>
</tr>
<tr>
<td>Very happy</td>
<td>21</td>
<td>159</td>
<td>110</td>
</tr>
</tbody>
</table>

We can test $H_0 : X \perp Y$ using Mental-Haenszel (MH) test. Assign scores 1, 2, 3 for $X$ and 1, 2, 3 for $Y$, say, then we use $M^2 = (n - 1)r^2$.

We can also consider a cumulative logit model:

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta x, \quad j = 1, 2$$

and test $H_0 : \beta = 0$ to test $H_0 : X \perp Y$. 

Slide 355
SAS program and output:

```sas
data table2_13;
  input x y1-y3 @@;
datalines;
1 94 249 83
2 53 372 221
3 21 159 110
;

data table2_13; set table2_13;
  array temp {3} y1-y3;
  do y=1 to 3;
    count=temp(y);
    output;
  end;
run;

proc freq;
  weight count;
  tables x*y/chisq cmh;
run;
```

```
<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>4</td>
<td>73.3525</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>4</td>
<td>71.3045</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

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>55.9258</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>2</td>
<td>67.9946</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>4</td>
<td>73.2986</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
```
The MH test for $H_0: X \perp Y$ is $M^2 = 55.9$. The Wald test for $H_0: \beta = 0$ is $\chi^2 = 53.8$. Both are compared to $\chi_1^2$. Very similar.
• Case 2: $Y$ – ordinal, $X$ – nominal (CMH2). For table 2.13, if we treat $X$ (income) as nominal, we may consider

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta_1 x_1 + \beta_2 x_2, \quad j = 1, 2$$

and test $H_0 : \beta_1 = 0, \beta_2 = 0$ to test $H_0 : X \perp Y$.

```r
proc logistic;
  freq count;
  class x / param=ref;
  model y = x / aggregate scale=none;
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>67.4166</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>64.6620</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>65.4019</td>
<td>2</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

All tests are very close to CMH2 ($\chi^2 = 67.99$) with $df = 2$. 
• Case 3: $X, Y$ — nominal (CMH3). For table 2.13, if we treat both $X, Y$ as nominal, we may consider the baseline-category logit model

$$\text{logit}(\pi_1/\pi_3) = \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2$$
$$\text{logit}(\pi_2/\pi_3) = \alpha_2 + \beta_{21}x_1 + \beta_{22}x_2$$

and test $H_0: \beta_{11} = 0, \beta_{12} = 0, \beta_{21} = 0, \beta_{22} = 0$ to test $H_0: X \perp Y$.

```plaintext
proc logistic;
  freq count;
  class x / param=ref;
  model y (ref="3") = x / aggregate scale=none link=glogit;
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>71.3045</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>73.3525</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>68.3455</td>
<td>4</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

All tests are similar to CMH3: $\chi^2 = 73.3$ or Pearson $\chi^2$, LRT, $df = 4$. 

Slide 359
IV.2 Tests of $X \perp Y | Z$

- Test independence between income ($X$) and job satisfaction ($Y$) given gender ($Z$). Data – 1991 GSS.

Table 6.12. Job Satisfaction and Income, Controlling for Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Income</th>
<th>Very Dissatisfied</th>
<th>A Little Satisfied</th>
<th>Moderately Satisfied</th>
<th>Very Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>&lt;5000</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5000–15,000</td>
<td>2</td>
<td>3</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>15,000–25,000</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>&gt;25,000</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Male</td>
<td>&lt;5000</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5000–15,000</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15,000–25,000</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>&gt;25,000</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>6</td>
</tr>
</tbody>
</table>

• We can use CMH to test $H_0 : X \perp Y \mid Z$:

```
data table6_12;
  input gender$ income$ incscore y1-y4;
cards;
Female <5000 3 1 3 11 2
Female 5000~15,000 10 2 3 17 3
Female 15,000~25,000 20 0 1 8 5
Female >25,000 35 0 2 4 2
Male <5000 3 1 1 2 1
Male 5000~15,000 10 0 3 5 1
Male 15,000~25,000 20 0 0 7 3
Male >25,000 35 0 1 9 6;
```

```
data table6_12; set table6_12;
array temp {4} y1-y4;
  do y=1 to 4;
    count=temp(y);
    if y=1 then jobsat=1; else jobsat=y+1; /* jobsat scores: 1,3,4,5 */
    output;
  end;
run;
```

```
data table6_12; set table6_12;
  array temp {4} y1-y4;
  do y=1 to 4;
    count=temp(y);
    if y=1 then jobsat=1; else jobsat=y+1; /* jobsat scores: 1,3,4,5 */
    output;
  end;
run;
```

```
proc freq order=data;
  weight count;
  tables gender*incscore*jobsat / cmh;
run;
```

\begin{tabular}{lcc}
\hline
Statistic & Alternative Hypothesis & DF & Value & Prob \\
\hline
1 & Nonzero Correlation & 1 & 6.1563 & 0.0131 \\
2 & Row Mean Scores Differ & 3 & 9.0342 & 0.0288 \\
3 & General Association & 9 & 10.2001 & 0.3345 \\
\hline
\end{tabular}

Slide 361
• We can also adjust for $z$ in the previous 3 models.

Case 1: Treat $X, Y$ as ordinal and consider cumulative logit model:

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta x + \beta_z z, \ j = 1, 2, 3.$$ 

```latex
proc logistic;
    freq count;
    class gender / param=ref;
    model y = gender / aggregate=(income gender) scale=none;
run;
```

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>19.6230</td>
<td>20</td>
<td>0.9812</td>
<td>0.4817</td>
</tr>
<tr>
<td>Pearson</td>
<td>20.9457</td>
<td>20</td>
<td>1.0473</td>
<td>0.4003</td>
</tr>
</tbody>
</table>

```latex
proc logistic;
    freq count;
    class gender / param=ref;
    model y = incscore gender / aggregate=(income gender) scale=none;
run;
```

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>13.9519</td>
<td>19</td>
<td>0.7343</td>
<td>0.7865</td>
</tr>
<tr>
<td>Pearson</td>
<td>14.3128</td>
<td>19</td>
<td>0.7533</td>
<td>0.7652</td>
</tr>
</tbody>
</table>

$LRT \ H_0 : \beta = 0 (X \perp Y | Z)$ is $G^2 = 19.6230 - 13.9519 = 5.67$, with $df = 1$, $p$-value=0.0173. Similar to CMH1.
Case 2: Treat $Y$ as ordinal, $X$ as nominal:

$$\text{logit}(P[Y \leq j]) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_z z, \quad j = 1, 2, 3$$

and test $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ to test $H_0 : X \perp Y | Z$.

```latex
proc logistic;
  freq count;
  class gender income / param=ref;
  model y = income gender / aggregate=(income gender) scale=none;
run;
```

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>10.5051</td>
<td>17</td>
<td>0.6179</td>
<td>0.8811</td>
</tr>
<tr>
<td>Pearson</td>
<td>10.5691</td>
<td>17</td>
<td>0.6217</td>
<td>0.8781</td>
</tr>
</tbody>
</table>

The LRT for $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ is

$$G^2 = 19.6230 - 10.5051 = 9.12$$

with $df = 3$, p-value=0.0277. Very similar to CMH2.
Case 3: $Y$-nominal, $X$-ordinal. Consider baseline-category logit model:

$$\text{logit}(\pi_j / \pi_4) = \alpha_j + \beta_j x + \beta_z z, \quad j = 1, 2, 3$$

and test $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ to test $H_0 : X \perp Y | Z$.

```
proc logistic;
    freq count;
    class gender / param=ref;
    model y (ref="4") = gender / link=glogit aggregate=(income gender) scale=none;
run;

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>19.3684</td>
<td>18</td>
<td>1.0760</td>
<td>0.3695</td>
</tr>
<tr>
<td>Pearson</td>
<td>21.0545</td>
<td>18</td>
<td>1.1697</td>
<td>0.2767</td>
</tr>
</tbody>
</table>
```

```
proc logistic;
    freq count;
    class gender / param=ref;
    model y (ref="4") = incscore gender / link=glogit aggregate=(income gender) scale=none;
run;

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>11.7448</td>
<td>15</td>
<td>0.7830</td>
<td>0.6982</td>
</tr>
<tr>
<td>Pearson</td>
<td>11.3182</td>
<td>15</td>
<td>0.7545</td>
<td>0.7297</td>
</tr>
</tbody>
</table>
```

The LRT for $H_0 : \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ is $G^2 = 7.62$ with $df = 3$, p-value=0.055. Similar to CMH2.
Case 4: Treat $X, Y$ as nominal, Consider baseline-category logit model:

$$\text{logit}(\pi_j / \pi_4) = \alpha_j + \beta_{j1}x_1 + \beta_{j2}x_2 + \beta_{j3}x_3 + \beta_{zj}z, \ j = 1, 2, 3$$

and test $H_0 : \beta_{ij} = 0 (i, j = 1, 2, 3)$ to test $H_0 : X \perp Y | Z$.

```plaintext
proc logistic;
freq count;
class gender income / param=ref;
model y (ref="4") = income gender / link=glogit aggregate=(income gender) scale=none;
run;
```

**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>7.0935</td>
<td>9</td>
<td>0.7882</td>
<td>0.6274</td>
</tr>
<tr>
<td>Pearson</td>
<td>6.6050</td>
<td>9</td>
<td>0.7339</td>
<td>0.6782</td>
</tr>
</tbody>
</table>

The LRT for $H_0 : \beta_{ij} = 0 (i, j = 1, 2, 3)$ is

$$G^2 = 19.3684 - 7.0935 = 12.27 \text{ with } df = 9, \ p\text{-value}=0.199. \text{ Similar to CMH3.}$$
8 Models for Matched Pairs

I Comparing Two Probabilities Using Dependent Proportions

- Example: Opinion relating to environment (Table 8.1 from 2000 GSS)

<table>
<thead>
<tr>
<th>Cut living standard ($Y_2$)</th>
<th>Yes (1)</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay higher taxes ($Y_1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes (1)</td>
<td>227</td>
<td>132</td>
</tr>
<tr>
<td>No (0)</td>
<td>107</td>
<td>678</td>
</tr>
</tbody>
</table>

$n = 1144$ Americans. Here each subject is matched with himself/herself to get $Y_1$ and $Y_2$.

We are interested in comparing $\pi_1 = P[Y_1 = 1]$ and $\pi_2 = P[Y_2 = 1]$.

We are not very interested in testing $Y_1 \perp Y_2$. 
If we convert table to

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay higher taxes</td>
<td>359</td>
<td>785</td>
</tr>
<tr>
<td>Cut living standard</td>
<td>334</td>
<td>810</td>
</tr>
</tbody>
</table>

\[
P[Y_1 = 1]: \hat{\pi}_1 = \frac{359}{1144} = 0.314
\]

\[
P[Y_2 = 1]: \hat{\pi}_2 = \frac{334}{1144} = 0.292
\]

Difference \[\hat{\pi}_1 - \hat{\pi}_2 = 0.022\]

\[
\text{var}(\hat{\pi}_1 - \hat{\pi}_2)
\]

No way to get \[\text{var}(\hat{\pi}_1 - \hat{\pi}_2)\] if data is summarized using this table.

Need to go back to the original table!
I.1 Proportion difference using a matched sample

- Data and probability structure

\[
\begin{array}{c|cc}
Y_2 & 1 & 0 \\
\hline
Y_1 & 1 & n_{11} & n_{12} \\
& 0 & n_{21} & n_{22} \\
1 & 0 & \pi_{11} & \pi_{12} \\
\pi_{21} & \pi_{22} \\
\end{array}
\]

\[\pi_1 = P[Y_1 = 1] = \pi_{11} + \pi_{12},\]
\[\pi_2 = P[Y_2 = 1] = \pi_{11} + \pi_{21}.\]

Difference \(\delta = \pi_1 - \pi_2 = \pi_{12} - \pi_{21}.\)

Given data, the MLE of \(\pi_{ij}\)’s: \(\hat{\pi}_{ij} = n_{ij}/n\)

\[\Rightarrow \quad \hat{\delta} = \hat{\pi}_{12} - \hat{\pi}_{21} = \frac{n_{12} - n_{21}}{n}.\]
$\text{var}(\hat{\delta}) = \frac{\pi_{12}(1 - \pi_{12})}{n} + \frac{\pi_{21}(1 - \pi_{21})}{n} + \frac{2\pi_{12}\pi_{21}}{n}$

$\text{var}(\hat{\delta}) = \frac{\hat{\pi}_{12}(1 - \hat{\pi}_{12})}{n} + \frac{\hat{\pi}_{21}(1 - \hat{\pi}_{21})}{n} + \frac{2\hat{\pi}_{12}\hat{\pi}_{21}}{n}$

$= \frac{n_{12}(n - n_{12}) + n_{21}(n - n_{21}) + 2n_{12}n_{21}}{n^3}$

$= \frac{(n_{12} + n_{21}) - (n_{12} - n_{21})^2/n}{n^2}$

- For our example,

$\hat{\delta} = 0.022$

$\text{var}(\hat{\delta}) = \frac{(132 + 107) - (132 - 107))^2/1144}{1144^2} = \frac{238.45}{1144^2}$

$\text{SE}(\hat{\delta}) = \frac{\sqrt{238.45}}{1144} = 0.0135$

$\chi^2 = (0.022/0.0135)^2 = 2.66$

$95\% \text{ Wald CI of } \delta : \ 0.022 \pm 1.96 \times 0.0135 = [-0.005, 0.048]$
I.2 McNemar’s Test

- If we calculate \( \text{var}(\hat{\delta}) \) under \( H_0 : \delta = 0 \Leftrightarrow H_0 : \pi_{21} = \pi_{12} \), then

\[
\text{var}(\hat{\delta}) = \frac{\pi_{12}(1 - \pi_{12})}{n} + \frac{\pi_{21}(1 - \pi_{21})}{n} + \frac{2\pi_{21}\pi_{12}}{n}
\]

\[
= \frac{\pi_{12}(1 - \pi_{12})}{n} + \frac{\pi_{12}(1 - \pi_{12})}{n} + \frac{2\pi_{12}\pi_{12}}{n}
\]

\[
= \frac{2\pi_{12}}{n}.
\]

- It can be shown the MLE of \( \pi_{12} \) under \( H_0 : \pi_{12} = \pi_{21} \) is that

\[
\hat{\pi}_{12} = \frac{n_{12} + n_{21}}{2n}
\]
\[
\text{var}(\hat{\delta})_{H_0} = \frac{2}{n} \times \frac{n_{12} + n_{21}}{2n} = \frac{n_{12} + n_{21}}{n^2}
\]

\[
\chi^2 = \frac{\hat{\delta}^2}{\text{var}(\hat{\delta})_{H_0}} = \frac{(n_{12} - n_{21})^2/n^2}{(n_{12} + n_{21})/n^2} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \sim \chi_1^2
\]

This is the McNemar’s test.

- For our example, McNemar’s \( \chi^2 = (132 - 107)^2/(132 + 107) = 2.615 \). Do not reject \( H_0 : \pi_{12} = \pi_{21} \) at level 0.05.
• SAS program and output

```sas
data table8_1;
  input pay_ht y1 y2;
cards;
  1 227 132
  0 107 678;
;
proc print;
  var pay_ht cut_ls count;
run;
```

```
Obs  pay_ht  cut_ls  count
1    1      1      227
2    1      0      132
3    0      1      107
4    0      0      678
```
proc freq order=data;
    weight count;
    tables pay_ht*cut_ls / ;
    test agree;
run;

******************************************************************************************************

Statistics for Table of pay_ht by cut_ls

McNemar’s Test

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic (S)</td>
<td>2.6151</td>
</tr>
<tr>
<td>DF</td>
<td>1</td>
</tr>
<tr>
<td>Pr &gt; S</td>
<td>0.1059</td>
</tr>
</tbody>
</table>

******************************************************************************************************
Note: The McNemar’s test can be derived from the Pearson $\chi^2$ test.

Under $H_0 : \pi_{12} = \pi_{21}$, the MLE’s of $\pi_{ij}$ are

$$\hat{\pi}_{11} = \frac{n_{11}}{n}, \hat{\pi}_{12} = \hat{\pi}_{21} = \frac{n_{12} + n_{21}}{2n}, \hat{\pi}_{22} = \frac{n_{22}}{n}.$$ 

The Pearson $\chi^2$ test for $H_0 : \pi_{12} = \pi_{21}$ is

$$\chi^2 = \frac{(n_{11} - n\hat{\pi}_{11})^2}{n\hat{\pi}_{11}} + \frac{(n_{12} - n\hat{\pi}_{12})^2}{n\hat{\pi}_{12}} + \frac{(n_{21} - n\hat{\pi}_{21})^2}{n\hat{\pi}_{21}} + \frac{(n_{22} - n\hat{\pi}_{22})^2}{n\hat{\pi}_{22}}$$

$$= 0 + \frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})} + \frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})} + 0$$

$$= \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}},$$

with $df = 3 - 2 = 1$. This is the same as the McNemar’s test.
GLM/Logistic Model for Matched Data

II.1 Marginal probabilities, population-level odds-ratio

- Risk difference from the converted table:

<table>
<thead>
<tr>
<th></th>
<th>$Y$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td></td>
<td>Yes (1)</td>
<td>No (0)</td>
</tr>
<tr>
<td>Pay higher taxes (1)</td>
<td>359</td>
<td>785</td>
<td>1144</td>
</tr>
<tr>
<td>Cut living standard (0)</td>
<td>334</td>
<td>810</td>
<td>1144</td>
</tr>
</tbody>
</table>

Let $\pi(x) = P[Y = 1 | X = x]$. If we fit a GLM link to $\pi(x)$ with the identity

$$\pi(x) = \alpha + \beta x,$$

then $\beta = \delta$, the risk difference.

As we indicated before, $\hat{\text{var}}(\hat{\delta})$ cannot be derived from this table and we need to go back to the original table.
• The formula $\hat{\text{var}}(\hat{\delta})$ can be obtained by fitting the above GLM to the data by recovering the original data at subject level and recognizing the dependence of two observations from the same subjects.

• Each subject has two binary data points $y_{i1}, y_{i2}$

\[
\begin{array}{ccc}
\hline
X & \text{Yes (1)} & \text{No (0)} \\
\hline
\text{Pay higher taxes (1)} & y_{i1} & 1 - y_{i1} & 1 \\
\text{Cut living standard (0)} & y_{i2} & 1 - y_{i2} & 1 \\
\hline
\end{array}
\]

• There are only 4 types of such tables:

\[
\begin{array}{cccc}
\hline
& 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
X & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
\hline
\end{array}
\]

Type I: 227 II: 132 III: 107 IV: 678
• SAS program and part of output:

```sas
title "Recover the individual data";
data newdata; set table8_1;
  retain id;
  if _n_=1 then id=0;
  do i=1 to count;
    id = id+1;
    do question=1 to 2;
      x = 2-question;
      if question=1 then
        y=pay_ht;
      else
        y=cut_ls;
      output;
    end;
  end;
run;

proc genmod data=newdata descending;
  class id;
  model y = x / dist=bin link=identity;
  repeated subject=id / type=un;
run;
```

******************************************************************************

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z| |
|-----------|----------|----------------|-----------------------|--------|--------|
| Intercept | 0.2920   | 0.0134         | 0.2656                | 0.3183 | 21.72  | <.0001 |
| x         | 0.0219   | 0.0135         | -0.0046               | 0.0483 | 1.62   | 0.1055 |

Slide 377
- The approach we used to account for the dependence of observations from the same subjects is called GEE (for generalized estimating equation). We will talk about GEE in more detail in Chapter 9.

- The point estimate of $\beta$ and its standard error using GEE with the identity link are the same as those obtained before (slide 359).

- Odds-ratio from the converted table:

<table>
<thead>
<tr>
<th>$X$</th>
<th>Yes (1)</th>
<th>No (0)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay higher taxes (1)</td>
<td>359</td>
<td>785</td>
<td>1144</td>
</tr>
<tr>
<td>Cut living standard (0)</td>
<td>334</td>
<td>810</td>
<td>1144</td>
</tr>
</tbody>
</table>
The odds-ratio estimate of responding *Yes* between *paying higher taxes* \((X = 1)\) and *cutting living standard* \((X = 0)\) is

\[
\hat{\theta}_{XY} = \frac{359 \times 810}{334 \times 789} = 1.11
\]

which can be obtained by fitting the logit model to the data \((\theta_{XY} = e^{\beta})\):

\[
\logit\{\pi(x)\} = \alpha + \beta x.
\]

However, we cannot use the following formula:

\[
\var\left(\log \hat{\theta}_{XY}\right) = \frac{1}{359} + \frac{1}{785} + \frac{1}{334} + \frac{1}{810} = 0.00829,
\]

since two samples defined by two rows are identical! This will be the formula used for \(\var(\hat{\beta})\) if we fit a regular logit model to the data.

We can get the correct \(\var(\hat{\beta})\) if we take the dependence of two observations from the same subject into account with GEE.
• SAS program and part of the output:

```sas
proc genmod data=newdata descending;
   class id;
   model y = x / dist=bin link=logit;
   repeated subject=id / type=un;
run;
```

********************************************************************************

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z| |
|-----------|----------|----------------|-----------------------|--------|---|
| Intercept | -0.8859  | 0.0650         | -1.0133 -0.7584 -13.62 <.0001 |
| x         | 0.1035   | 0.0640         | -0.0219 0.2289 1.62 0.1056   |

95% CI for \( \log(\theta_{XY}) \): \(0.1035 \pm 1.96 \times 0.0640 = [-0.022, 0.229]\).

95% CI for \( \theta_{XY} \): \([e^{-0.022}, e^{0.229}] = [0.978, 1.257]\).

• **Note:** In our example, the correct \( \hat{\text{var}}(\hat{\beta}) = 0.0640^2 = 0.0041 < 0.00829 = \) the estimate from the incorrect variance formula!

• We can also adjust for other covariates in the above GLMs.

• **Note:** The estimator \( \hat{\theta}_{XY} \) estimates an underlying true-odds ratio.

That odds-ratio is in the population level. Therefore it is called
population-averaged odds-ratio.

- We can also consider models at the individual level

\[
\begin{array}{c|cc}
X & Y_{\text{Yes}} (1) & Y_{\text{No}} (0) \\
\hline
\text{Pay higher taxes (1)} & y_{i1} & 1 - y_{i1} & 1 \\
\text{Cut living standard (0)} & y_{i2} & 1 - y_{i2} & 1 \\
\end{array}
\]

Let \( \pi_i(x) = P[Y_{ij} = 1|x, \alpha_i] \) the individual probability of responding “Yes” to question \( j \) and consider the logit model:

\[
\text{logit}\{\pi_i(x)\} = \alpha_i + \beta_s x,
\]

where \( \alpha_i \) is specific to subject \( i \), usually assumed to be random.

- The parameter \( \beta_s \) is subject-specific, and \( e^{\beta_s} \) is the subject-specific odds-ratio. It compares the response probs between questions 1 and 2 for a particular subject \( i \). If we assume \( \alpha_i \) a random variable, the above model is called a random effects model. Will be discussed more later.
II.2 Conditional logistic regression for matched data from prospective studies

- If we assume the subject-specific logit model for the opinion data
  \[
  \text{logit}\{\pi_i(x)\} = \alpha_i + \beta_s x, \quad i = 1, 2, \ldots, n.
  \]

  Since there are \(n\) many \(\alpha_i\)'s, we do not want to conduct the ML analysis.

- Conditional approach: find out sufficient stat for \(\alpha_i\)'s and use the conditional distribution of data given the suff. stat.

- It can be shown that the conditional likelihood of \(\beta_s\) is
  \[
  L_c(\beta_s) = \frac{e^{\beta_s n_{12}}}{(1 + e^{\beta_s})^{n_{21}+n_{12}}}
  \]

  The conditional ML estimate: \(\hat{\beta}_s = \log\left(\frac{n_{12}}{n_{21}}\right)\). The variance estimate of \(\hat{\beta}_s\) can be shown to be \(1/n_{12} + 1/n_{21}\).
• For our data, the subject-specific odds-ratio estimate is

\[ e^{\hat{\beta}_s} = \frac{n_{12}}{n_{21}} = \frac{132}{107} = 1.23. \]

Note that this subject-specific odds-ratio estimate is greater than the population-averaged odds-ratio estimate \( \hat{\theta}_{XY} = 1.11 \).

• SAS program and part of the output:

```sas
proc logistic data=newdata descending;
class id;
model y = x / link=logit;
strata id;
run;

*******************************************************************
Analysis of Conditional Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1</td>
<td>0.2100</td>
<td>0.1301</td>
<td>2.6055</td>
<td>0.1065</td>
</tr>
</tbody>
</table>

We can check that 0.21 = log(132/107), \( SE(\hat{\beta}_s) = \sqrt{1/132 + 1/107} \).

• Note: We can put more covariates in the conditional logistic regression model to adjust their effects.
II.3 Conditional logistic regression for matched case-control studies

- The conditional logistic regression model can also be applied to data obtained from matched case-control studies. For example, matched case-control study on association between diabetes and MI (case):

Table 8.3. Previous Diagnoses of Diabetes for Myocardial Infarction Case–Control Pairs

<table>
<thead>
<tr>
<th>MI Controls</th>
<th>Diabetes</th>
<th>No diabetes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>9</td>
<td>16</td>
<td>25</td>
</tr>
<tr>
<td>No diabetes</td>
<td>37</td>
<td>82</td>
<td>119</td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>98</td>
<td>144</td>
</tr>
</tbody>
</table>

• Let $Y_{ij} = 1/0$ for MI/control for subject $j$ in pair $i$, $x = 1/0$ for diabetes/no diabetes. There are 144 tables like the following:

\[
\begin{array}{cccc}
Y \\
\text{Type} & \text{I: 9} & \text{III: 16} & \text{II: 37} & \text{IV: 82} \\
X & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\
\end{array}
\]

• Treat data as if from a prospective study and fit

$$\text{logit}\{P(Y_{ij} = 1)\} = \alpha_i + \beta_s x, \quad i = 1, 2, \cdots, n \text{ pair, } j = 1, 2.$$  

• The conditional MLE of $\beta_s$ is

$$\hat{\beta}_s = \log(n_{21}/n_{12}) = \log(37/16) = 0.838 \text{ with variance estimate:}$$

$$\text{var}(\hat{\beta}_s) = 1/37 + 1/16 = 0.09, \quad \text{SE}(\hat{\beta}_s) = \sqrt{0.09} = 0.3$$
• The above analysis can be obtained using `proc logistic`. It is especially useful if other covariates (except the matching ones) are available:

• **SAS program and part of output:**

```sas
data table8_3;
  input condiab y1 y2;
cards;
  1 9 16
  0 37 82
;

data table8_3; set table8_3;
  array temp {2} y1-y2;
  do j=1 to 2;
    count=temp(j);
    casediab = 2-j;
    output;
  end;
run;

proc print;
  var condiab casediab count;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>condiab</th>
<th>casediab</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>82</td>
</tr>
</tbody>
</table>
title "Recover individual pair data";
data newdata; set table8_3;
  retain pair;
  if _n_=1 then pair=0;
  do i=1 to count;
    pair = pair+1;
    do mi=0 to 1;
      if mi=0 then
        diab = condiab; /* for MI=0, the diab info is the control diab info */
      else
        diab = casediab; /* for MI=1, the diab info is the case diab info */
      output;
    end;
  end;
run;

proc logistic descending;
  class pair;
  model mi = diab / link=logit;
  strata pair;
run;

*************************************************************************
Analysis of Conditional Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>diab</td>
<td>1</td>
<td>0.8383</td>
<td>0.2992</td>
<td>7.8501</td>
<td>0.0051</td>
</tr>
</tbody>
</table>
II.4 Connection between McNemar test and CMH test

- The table given at the beginning can be viewed as a summary of 1144 partial $2 \times 2$ tables, one for each subject:

\[
\begin{array}{c|c|c|c}
Y & 1 & 0 \\
\hline
1 & y_1 & 1 - y_1 & 1 \\
0 & y_2 & 1 - y_2 & 1 \\
\end{array}
\]

- There are only 4 types of such tables:

\[
\begin{array}{c|c|c}
Y & 1 & 0 \\
\hline
1 & n_{11} & n_{12} \\
0 & n_{21} & n_{22} \\
\end{array}
\]

Type: I: $n_{11}$ II: $n_{12}$ III: $n_{21}$ IV: $n_{22}$
Let us construct the CMH test for $H_0 : X$ and $Y$ are conditional independent given each subject:

$$E(y_{i1}|\text{margins}, H_0) = \begin{cases} 1 & \text{for type I tables} \\ 1/2 & \text{for type II or III tables} \\ 0 & \text{for type IV tables} \end{cases}$$

$$\text{var}(y_{i2}|\text{margins}, H_0) = \begin{cases} 0 & \text{for type I or IV tables} \\ \frac{1 \times 1 \times 1}{2^2 \times (2-1)} = \frac{1}{4} & \text{for type II or III tables} \end{cases}$$

$$\Rightarrow \chi^2_{CMH} = \frac{[n_{11}(1 - 1) + n_{12}(1 - 0.5) + n_{21}(0 - 0.5) + n_{22}(0 - 0)]^2}{n_{11} \times 0 + n_{12} \times 0.25 + n_{21} \times 0.25 + n_{22} \times 0}$$

$$= \frac{(n_{12} - n_{21})^2}{n_{21} + n_{12}},$$

the same as the McNemar’s test!
III Comparing Margins of Square Tables

III.1 Comparing margins for nominal response

- Example (Table 8.5) Coffee brand choice between 1st and 2nd purchases:

<table>
<thead>
<tr>
<th>First Purchase</th>
<th>High Point</th>
<th>Taster’s</th>
<th>Sanka</th>
<th>Nescafe</th>
<th>Brim</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Point</td>
<td>93 (93)</td>
<td>17 (13.2)</td>
<td>44 (32.5)</td>
<td>7 (6.1)</td>
<td>10 (7.8)</td>
</tr>
<tr>
<td>Taster’s choice</td>
<td>9 (12.7)</td>
<td>46 (46)</td>
<td>11 (10.5)</td>
<td>0 (0.0)</td>
<td>9 (9.1)</td>
</tr>
<tr>
<td>Sanka</td>
<td>17 (26.0)</td>
<td>11 (11.6)</td>
<td>155 (155)</td>
<td>9 (11.3)</td>
<td>12 (12.8)</td>
</tr>
<tr>
<td>Nescafe</td>
<td>6 (7.0)</td>
<td>4 (3.5)</td>
<td>9 (7.5)</td>
<td>15 (15)</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Brim</td>
<td>10 (14.0)</td>
<td>4 (4.0)</td>
<td>12 (11.3)</td>
<td>2 (2.3)</td>
<td>27 (27)</td>
</tr>
</tbody>
</table>

• Let
  \( Y_1 = \) coffee brand choice at first purchase,
  \( Y_2 = \) coffee brand choice at second purchase.

We are interested in testing \( H_0 : P[Y_1 = k] = P[Y_2 = k] \)
\((k = 1, 2, 3, 4, 5)\).

• We can test the above \( H_0 \) by comparing sample marginal proportions
  \( p_{i+} \) to \( p_{+i} \):

  \[
d = \begin{pmatrix}
p_{1+} - p_{+1} \\
p_{2+} - p_{+2} \\
\vdots \\
p_{I-1,+} - p_{+I-1}
\end{pmatrix}
\]

Then construct

\[
\chi^2 = d^T \{\text{var}(d)\}^{-1} d \overset{H_0}{\sim} \chi_{I-1}^2.
\]
• We can conduct the above test using `proc catmod`.

• **SAS program and part of output:**

```sas
data table8_5;
  input firstbuy y1-y5;
  cards;
  1  93  17  44  7  10
  2  9   46  11  0  9
  3  17  11  155 9  12
  4  6   4   9  15  2
  5  10  4   12 2  27
;

data table8_5; set table8_5;
  array temp {5} y1-y5;
  do secbuy=1 to 5;
    count=temp(secbuy);
    output;
  end;
run;

proc print;
  var firstbuy secbuy count;
run;
```
<table>
<thead>
<tr>
<th>Obs</th>
<th>firstbuy</th>
<th>secbuy</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>155</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>5</td>
<td>27</td>
</tr>
</tbody>
</table>

```plaintext
proc freq;
  weight count;
  tables firstbuy*secbuy / norow nocol;
  test agree;
run;
```
### Table of firstbuy by secbuy

<table>
<thead>
<tr>
<th>firstbuy</th>
<th>secbuy</th>
<th>Frequency</th>
<th>Percent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>93</td>
<td>17.19%</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>17</td>
<td>3.14%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>44</td>
<td>8.13%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
<td>1.29%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>1.85%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>9</td>
<td>1.66%</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>46</td>
<td>8.50%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>11</td>
<td>2.03%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0.00%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>1.66%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>17</td>
<td>3.14%</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11</td>
<td>2.03%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>155</td>
<td>28.65%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9</td>
<td>1.66%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>12</td>
<td>2.22%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1.11%</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0.74%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9</td>
<td>1.66%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>15</td>
<td>2.77%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>0.37%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>10</td>
<td>1.85%</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0.74%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>12</td>
<td>2.22%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>0.37%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>27</td>
<td>4.99%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>135</td>
<td>24.95%</td>
<td>541</td>
</tr>
<tr>
<td></td>
<td></td>
<td>82</td>
<td>15.16%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>231</td>
<td>42.70%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>6.10%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
<td>11.09%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>541</td>
<td>100.00%</td>
<td></td>
</tr>
</tbody>
</table>

Statistics for Table of firstbuy by secbuy

Test of Symmetry

<table>
<thead>
<tr>
<th>Statistic (S)</th>
<th>DF</th>
<th>Pr &gt; S</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.4124</td>
<td>10</td>
<td>0.0256</td>
</tr>
</tbody>
</table>
proc catmod data=table8_5;;
weight count;
response marginals;
model firstbuy*secbuy = _response_;
  repeated time 2;
run;

****************************************************************

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4</td>
<td>6471.41</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>time</td>
<td>4</td>
<td>12.58</td>
<td>0.0135</td>
</tr>
</tbody>
</table>

The Wald test for marginal homogeneity is $\chi^2 = 12.6$ with $df = 4$, p-value$=0.0135$. We reject the marginal homogeneity at level 0.05. That is, we conclude that customers’ coffee brand choices between their first and second buys are not the same.
III.2 Comparing margins for ordinal response

- Example (Table 8.6): Response to recycling and driving less to help environment

<table>
<thead>
<tr>
<th>Recycle</th>
<th>Drive Less</th>
<th>Always</th>
<th>Often</th>
<th>Sometimes</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always</td>
<td>12 (12)</td>
<td>43 (43.1)</td>
<td>163 (165.6)</td>
<td>233 (232.8)</td>
<td></td>
</tr>
<tr>
<td>Often</td>
<td>4 (3.9)</td>
<td>21 (21)</td>
<td>99 (98.0)</td>
<td>185 (184.5)</td>
<td></td>
</tr>
<tr>
<td>Sometimes</td>
<td>4 (1.4)</td>
<td>8 (9.0)</td>
<td>77 (77)</td>
<td>230 (227.3)</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0 (0.2)</td>
<td>1 (1.5)</td>
<td>18 (20.7)</td>
<td>132 (132)</td>
<td></td>
</tr>
</tbody>
</table>

- Let $Y_{i1}$ be the subject $i$’s response to “How often do you make a special effort to sort ...”, $Y_{i2}$ be the subject $i$’s response to “How often do you cut back on driving ...".
• Use 1, 2, 3, 4 for four values: never/sometimes/often/always and consider cumulative logit model:

\[
\text{logit}\{P[Y_{i1} \geq j]\} = \alpha_j + \beta,
\]

\[
\text{logit}\{P[Y_{i2} \geq j]\} = \alpha_j.
\]

Then \( H_0 : \beta = 0 \Rightarrow \text{marginal homogeneity.} \)

• We can fit the above model using \texttt{proc genmod} by taking into account the correlation between 2 obs from the same subject using GEE (this analysis is different from the one given in the textbook).

• SAS program and part of output:

```sas
data table8_6;
  input recycle y1-y4;
  cards;
  4 12 43 163 233
  3 4 21 99 185
  2 4 8 77 230
  1 0 1 18 132
;```

data table8_6; set table8_6;
array temp {4} y1-y4;

do j=1 to 4;
    driveles=5-j;
    count=temp(j);
    output;
end;
run;

proc print;
    var recycle driveles count;
run;

<table>
<thead>
<tr>
<th>Obs</th>
<th>recycle</th>
<th>driveles</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
<td>163</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>233</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>99</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>2</td>
<td>77</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1</td>
<td>230</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>1</td>
<td>132</td>
</tr>
</tbody>
</table>
title "Recover individual data"

data newdata; set table8_6;
    retain id;
    if _n_=1 then id=0;

do i=1 to count;
    id = id+1;
    do question=1 to 2;
        x = 2-question;
        if question=1 then y=recycle;
        if question=2 then y=driveles;
        output;
    end;
end;
run;

proc genmod data=newdata descending;
    class id;
    model y = x / dist=multinomial link=clogit;
    repeated subject=id / type=ind;
run;
Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>y</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>471</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>382</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>676</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>931</td>
</tr>
</tbody>
</table>

PROC GENMOD is modeling the probabilities of levels of $y$ having LOWER Ordered Values in the response profile table.

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter     | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z| |
|---------------|----------|----------------|-----------------------|--------|------------------|
| Intercept1    | -3.3511  | 0.0829         | -3.5136 -3.1886       | -40.43 | <.0001           |
| Intercept2    | -2.2767  | 0.0743         | -2.4224 -2.1311       | -30.64 | <.0001           |
| Intercept3    | -0.5849  | 0.0588         | -0.7002 -0.4696       | -9.94  | <.0001           |
| x             | 2.7536   | 0.0815         | 2.5939 2.9133         | 33.80  | <.0001           |

The Wald test for $H_0 : \beta = 0$ is $z = 33.80$, p-value $< 0.0001$. Since $\beta > 0$, people are willing to put more effort in recycling than driving less to help environment.
IV Symmetry and Quasi-Symmetry for Square Tables

IV.1 Symmetry for nominal square tables

• Suppose $Y_1, Y_2$ are 2 categorical variables taking the same values $1, 2, \ldots, I$ with the probability structure as (assuming $I = 3$):

\[
\begin{array}{ccc}
 & 1 & 2 & 3 \\
1 & \pi_{11} & \pi_{12} & \pi_{13} \\
2 & \pi_{21} & \pi_{22} & \pi_{23} \\
3 & \pi_{31} & \pi_{32} & \pi_{33}
\end{array}
\]

We are interested in testing $H_0 : \pi_{ij} = \pi_{ji}$.

• Given data $\{n_{ij}\}$ from a multinomial sampling, the MLE’s of $\pi_{ij}$ under $H_0$ are:

\[
\hat{\pi}_{ii} = \frac{n_{ii}}{n}, \quad \hat{\pi}_{ij} = \frac{n_{ij} + n_{ji}}{2n}.
\]
The Pearson $\chi^2$ test and LRT for $H_0: \pi_{ij} = \pi_{ji}$ are

$$
\chi^2(S) = \sum_{i<j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} \overset{H_0}{\sim} \chi^2_{df}
$$

$$
G^2(S) = 2 \sum_{i<j} n_{ij} \log \left( \frac{2n_{ij}}{n_{ij} + n_{ji}} \right) + n_{ji} \log \left( \frac{2n_{ji}}{n_{ij} + n_{ji}} \right) \overset{H_0}{\sim} \chi^2_{df}
$$

with $df = I(I - 1)/2$.

The above Pearson $\chi^2$ test is an extension of the McNemar’s test.

For the coffee data, $\chi^2 = 20.4$, $G^2 = 22.5$ with $df = 5(5 - 1)/2 = 10$. The Pearson $\chi^2 = 20.4$ can be obtained using test agree in proc freq.
IV.2 Quasi-symmetry for nominal square tables

- The symmetry (⇒marginal homogeneity) model seldom fits data well. A more general model is the quasi-symmetry model that allows marginal heterogeneity:

\[
\log(\pi_{ij}/\pi_{ji}) = \beta_i - \beta_j \quad (i < j).
\]

Of course, only \(I - 1\) many \(\beta_i\)'s are needed. We can set \(\beta_I = 0\).

- If \(\beta_i = 0\) \((i = 1, 2, ..., I - 1)\), then we have a marginal symmetry model.

- The fitting of the above model can be realized by fitting a logistic model to the paired data \((n_{ij}, n_{ji}) \ (i < j)\) treating \(n_{ij}\) as the total # of success and \(n_{ji}\) as the total number of failure with no intercept.

- We need to delete the diagonal elements \(n_{ii}\)'s.
SAS program for the coffee data:

```sas
data table8_5; set table8_5;
  if firstbuy=secbuy then delete;
  if firstbuy<secbuy then do;
    y=1; ind1=firstbuy; ind2=secbuy;
    end;
  else do;
    y=0; ind1=secbuy; ind2=firstbuy;
    end;

  array x {5};
  do k=1 to 5;
    if k=ind1 then x[k]=1;
    else if k=ind2 then x[k]=-1;
    else x[k]=0;
  end;

  drop y1-y5 k;
run;

proc sort;
  by ind1 ind2 descending y;
run;

proc print;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>firstbuy</th>
<th>secbuy</th>
<th>count</th>
<th>y</th>
<th>ind1</th>
<th>ind2</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>17</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>44</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>17</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>5</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>3</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
</tbody>
</table>
title "Quasi-symmetry model";
proc genmod descending;
  freq count;
  model y = x1 x2 x3 x4 / dist=bin link=logit aggregate 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>6</td>
<td>9.9740</td>
<td>1.6623</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>6</td>
<td>9.9740</td>
<td>1.6623</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>6</td>
<td>8.5303</td>
<td>1.4217</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>6</td>
<td>8.5303</td>
<td>1.4217</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>x1</td>
<td>1</td>
<td>0.5954</td>
<td>0.2937</td>
<td>0.0199</td>
<td>4.11</td>
<td>0.0426</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>-0.0040</td>
<td>0.3294</td>
<td>-0.6495</td>
<td>0.00</td>
<td>0.9903</td>
</tr>
<tr>
<td>x3</td>
<td>1</td>
<td>-0.1133</td>
<td>0.2851</td>
<td>-0.6720</td>
<td>0.16</td>
<td>0.6911</td>
</tr>
<tr>
<td>x4</td>
<td>1</td>
<td>0.3021</td>
<td>0.4016</td>
<td>-0.4850</td>
<td>1.0892</td>
<td>0.4519</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

- **Note:** There is a weight statement in proc genmod. But it is not for the count $n_{ij}$'s!
- We can also use Proc Logistic to fit the above model and get a test of symmetry under the *Quasi-symmetry model*.

```sql
title "Quasi-symmetry model using proc logistic";
proc logistic descending;
  freq count;
  model y = x1 x2 x3 x4 / link=logit noint;
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>12.4989</td>
<td>4</td>
<td>0.0140</td>
</tr>
<tr>
<td>Score</td>
<td>12.2913</td>
<td>4</td>
<td>0.0153</td>
</tr>
<tr>
<td>Wald</td>
<td>11.8742</td>
<td>4</td>
<td>0.0183</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1</td>
<td>0.5954</td>
<td>0.2937</td>
<td>4.1105</td>
<td>0.0426</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>-0.00401</td>
<td>0.3294</td>
<td>0.0001</td>
<td>0.9903</td>
</tr>
<tr>
<td>x3</td>
<td>1</td>
<td>-0.1133</td>
<td>0.2851</td>
<td>0.1579</td>
<td>0.6911</td>
</tr>
<tr>
<td>x4</td>
<td>1</td>
<td>0.3021</td>
<td>0.4016</td>
<td>0.5659</td>
<td>0.4519</td>
</tr>
</tbody>
</table>
• From the output, we know that the GOF stats:

\[ \chi^2(QS) = 8.5, G^2(QS) = 10.0, \]

with \( df = 6 \). Reasonably good fit.

• We know the GOF for symmetry model

\[ \chi^2(S) = 20.4, G^2(S) = 22.5, \]

with \( df = 10 \).

• Assuming \textit{quasi-symmetry} model, \textit{symmetry} model can be tested using LRT

\[ LRT = 22.5 - 10.0 = 12.5, \]

with \( df = 10 - 6 = 4 \), \( \Rightarrow \) Reject \textit{symmetry} model under \textit{quasi-symmetry} model.
IV.3 Quasi-symmetry for ordinal square tables

- For square tables formed with two ordinal variables with the same levels, we can assign scores $u_i$ to the $i$th level and consider the following ordinal quasi-symmetry model:

$$\log(\pi_{ij}/\pi_{ji}) = \beta(u_j - u_i), \quad (i < j).$$

- Similar to the quasi-symmetry model for nominal square tables, we can fit the above model by fitting a logistic model to the paired data $(n_{ij}, n_{ji})$ $(i < j)$ treating $n_{ij}$ as the total # of success and $n_{ji}$ as the total number of failure and $x = u_j - u_i$ as the covariate with no intercept.

- We need to delete the diagonal elements $n_{ii}$’s.

- $\beta = 0 \Rightarrow$ symmetry. So we can test $H_0 : \beta = 0$ to test symmetry.
Let us use the recycle example to illustrate this above model. SAS program and part of output:

data table8_6; set table8_6;
  if recycle=driveles then delete;
  if recycle>driveles then do;
    y=1;
    x=recycle-driveles;
    ind1=driveles;
    ind2=recycle;
  end;
  else do;
    y=0;
    x=driveles-recycle;
    ind1=recycle;
    ind2=driveles;
  end;
array z {4};
do k=1 to 4;
  if k=ind1 then
    z[k]=1;
  else if k=ind2 then
    z[k]=-1;
  else
    z[k]=0;
end;
run;
title "Ordinal quasi-symmetry model"
proc logistic data=table8_6;
   freq count;
   model y (ref="0") = x / link=glogit aggregate scale=none noint;
run;

-----------------------------------------------------------------------------------

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>2.0309</td>
<td>2</td>
<td>1.0155</td>
<td>0.3622</td>
</tr>
<tr>
<td>Pearson</td>
<td>2.1029</td>
<td>2</td>
<td>1.0514</td>
<td>0.3494</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>1101.7102</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Score</td>
<td>762.6001</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Wald</td>
<td>252.0238</td>
<td>1</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>y</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1</td>
<td>1</td>
<td>2.3936</td>
<td>0.1508</td>
<td>252.0238</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

• GOF: Pearson $\chi^2 = 2.1$, $G^2 = 2.0$ with $df = 2$. Good fit. Based on this model, reject $H_0 : \beta = 0$, so reject symmetry.
• From the output, we got

\[
\log(\hat{\pi}_{12}/\hat{\pi}_{21}) = 2.3936(2 - 1) = 2.3936
\]
\[
\log(\hat{\pi}_{13}/\hat{\pi}_{31}) = 2.3936(3 - 1) = 4.78
\]
\[
\log(\hat{\pi}_{14}/\hat{\pi}_{41}) = 2.3936(4 - 1) = 7.18
\]
\[
\log(\hat{\pi}_{23}/\hat{\pi}_{32}) = 2.3936(3 - 2) = 2.3936
\]
\[
\log(\hat{\pi}_{24}/\hat{\pi}_{42}) = 2.3936(4 - 2) = 4.78
\]
\[
\log(\hat{\pi}_{34}/\hat{\pi}_{43}) = 2.3936
\]

For example,

\[
\hat{\pi}_{12} = \hat{\pi}_{21} e^{2.3936} = 11\hat{\pi}_{21}
\]

That is,

\[
P[\text{Recycle}={\text{Always}}, \text{Drive-less}={\text{often}}] = 11 \times P[\text{Recycle}={\text{Often}}, \text{Drive-less}={\text{Always}}]
\]
title "Quasi-symmetry model treating ordinal as nominal";
proc genmod data=table8_6 descending;
    freq count;
    model y = z1 z2 z3 / dist=bin link=logit aggregate noint;
run;
************************************************************************

<table>
<thead>
<tr>
<th>Criteria For Assessing Goodness Of Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Deviance</td>
</tr>
<tr>
<td>Scaled Deviance</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>.</td>
</tr>
<tr>
<td>z1</td>
<td>1</td>
<td>6.9269</td>
<td>0.4708</td>
<td>6.0040</td>
<td>7.8497</td>
<td>216.43</td>
</tr>
<tr>
<td>z2</td>
<td>1</td>
<td>4.3452</td>
<td>0.4223</td>
<td>3.5175</td>
<td>5.1729</td>
<td>105.87</td>
</tr>
<tr>
<td>z3</td>
<td>1</td>
<td>1.9937</td>
<td>0.3822</td>
<td>1.2447</td>
<td>2.7428</td>
<td>27.22</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>.</td>
</tr>
</tbody>
</table>

- Treating table as a nominal table, the quasi-symmetry has GOF:
  - Pearson $\chi^2 = 2.68$, $G^2 = 2.71$ with $df = 3$, again reasonably good fit.
• From nominal quasi-symmetry model fit, we know that

\[
\begin{align*}
\log(\hat{\pi}_{12}/\hat{\pi}_{21}) &= 6.9269 - 4.3452 = 2.58 \\
\log(\hat{\pi}_{13}/\hat{\pi}_{31}) &= 6.9269 - 1.9937 = 4.93 \\
\log(\hat{\pi}_{14}/\hat{\pi}_{41}) &= 6.9269 \\
\log(\hat{\pi}_{23}/\hat{\pi}_{32}) &= 4.3452 - 1.9937 = 2.35 \\
\log(\hat{\pi}_{24}/\hat{\pi}_{42}) &= 4.3452 \\
\log(\hat{\pi}_{34}/\hat{\pi}_{43}) &= 1.9937
\end{align*}
\]

Very similar to the results from the ordinal quasi-symmetry model fit.

• **Note:** Pearson GOF and LRT for symmetry: \( \chi^2 = 856, G^2 = 1093, df = 6 \). Very poor fit!
Analyzing Rater Agreement

- Example (Table 8.7): Diagnoses of carcinoma by two pathologists

<table>
<thead>
<tr>
<th>Pathologist X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>(8.5)</td>
<td>(-0.5)</td>
<td>(-5.9)</td>
<td>(-1.8)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
<td>14</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>(-0.5)</td>
<td>(3.2)</td>
<td>(-0.5)</td>
<td>(-1.8)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>36</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>(-4.1)</td>
<td>(-1.2)</td>
<td>(5.5)</td>
<td>(-2.3)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>(-3.3)</td>
<td>(-1.3)</td>
<td>(0.3)</td>
<td>(5.9)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>12</td>
<td>69</td>
<td>10</td>
<td>118</td>
</tr>
</tbody>
</table>

• Usually, the diagnoses \((Y_1, Y_2)\) between two raters are correlated (not independent). So if we use Pearson \(\chi^2\) or LRT \(G^2\), we would reject independence. Indeed,

\[
\chi^2 = 120, \quad G^2 = 118, \quad df = 9,
\]
even without taking into the ordinal scale. See the program and output on the next slide.

• However, \((Y_1, Y_2)\) being dependent does not mean \(Y_1\) agrees well with \(Y_2\). That is, association is not the same as agreement.

• Pearson \(\chi^2\) for symmetry \(H_0: \pi_{ij} = \pi_{ji}\) is \(\chi^2 = 30.3\) with \(df = 6\). Symmetry model not good either!

• We may consider models that captures agreement and disagreement.
data table8_7;
  input rater1 y1-y4;
cards;
  1  22  2  2  0
  2  5  7 14  0
  3  0  2 36  0
  4  0  1 17 10
;

data table8_7; set table8_7;
array temp {4} y1-y4;
do rater2=1 to 4;
  count=temp(rater2);
  output;
end;
run;

proc freq;
  weight count;
  tables rater1*rater2 / norow nocol chisq;
  test agree;
run;

********************************************************************************
Statistics for Table of rater1 by rater2

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>9</td>
<td>120.2635</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>9</td>
<td>117.9569</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>73.4843</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Test of Symmetry

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic (S)</td>
<td>30.2857</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Pr &gt; S</td>
<td>6</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Slide 417
CHAPTER 8

Simple Kappa Coefficient
--------------------------------

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa</td>
<td>0.4930</td>
</tr>
<tr>
<td>ASE</td>
<td>0.0567</td>
</tr>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.3818</td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>0.6042</td>
</tr>
</tbody>
</table>

Test of H0: Kappa = 0

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASE under H0</td>
<td>0.0501</td>
</tr>
<tr>
<td>Z</td>
<td>9.8329</td>
</tr>
<tr>
<td>One-sided Pr &gt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-sided Pr &gt;</td>
<td>Z</td>
</tr>
</tbody>
</table>

Weighted Kappa Coefficient
--------------------------------

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted Kappa</td>
<td>0.6488</td>
</tr>
<tr>
<td>ASE</td>
<td>0.0477</td>
</tr>
<tr>
<td>95% Lower Conf Limit</td>
<td>0.5554</td>
</tr>
<tr>
<td>95% Upper Conf Limit</td>
<td>0.7422</td>
</tr>
</tbody>
</table>

Test of H0: Weighted Kappa = 0

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASE under H0</td>
<td>0.0631</td>
</tr>
<tr>
<td>Z</td>
<td>10.2891</td>
</tr>
<tr>
<td>One-sided Pr &gt; Z</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Two-sided Pr &gt;</td>
<td>Z</td>
</tr>
</tbody>
</table>

Sample Size = 118

Slide 418
V.1 Quasi-independence model for rater agreement

- Treat \( \{n_{ij}\} \)'s as independent Poisson data with mean \( \mu_{ij} \)'s, we can fit the following *quasi-independence* model to the agreement data:

  \[
  \log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \delta_i I(i = j).
  \]

- **Note**: Without \( \delta_i \), the above model reduces to the independence model between \( Y_1 \) and \( Y_2 \). So the name *quasi-independence* model.

- Interpretation of *quasi-independence* model: For a pair of subjects, consider the event that each rater put one subject in category \( a \) and the other subject in category \( b \). Then the conditional odds that two raters agree rather than disagree on which subject is cat \( a \) and which one in cat \( b \) is

  \[
  \tau_{ab} = \frac{\pi_{aa}\pi_{bb}}{\pi_{ab}\pi_{ba}} = e^{\delta_a + \delta_b}.
  \]

  So if \( \delta_i > 0 \), then two raters tend to agree rather than disagree.
• SAS program and output for the quasi-independence model:

```sas
data table8_7; set table8_7;
  if rater1=rater2 then
    qi=rater1;
  else
    qi=5;
run;

title "Quasi-independence model";
proc genmod data=table8_7;
  class rater1 rater2 qi;
  model count = rater1 rater2 qi / dist=poi link=log;
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>5</td>
<td>13.1781</td>
<td>2.6356</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>5</td>
<td>13.1781</td>
<td>2.6356</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>5</td>
<td>11.5236</td>
<td>2.3047</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>5</td>
<td>11.5236</td>
<td>2.3047</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>qi</td>
<td>1</td>
<td>3.8611</td>
<td>0.7297</td>
<td>2.4308 - 5.2913</td>
<td>28.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>qi</td>
<td>2</td>
<td>0.6042</td>
<td>0.6900</td>
<td>-0.7481 - 1.9566</td>
<td>0.77</td>
<td>0.3812</td>
</tr>
<tr>
<td>qi</td>
<td>3</td>
<td>1.9025</td>
<td>0.8367</td>
<td>0.2625 - 3.5425</td>
<td>5.17</td>
<td>0.0230</td>
</tr>
<tr>
<td>qi</td>
<td>4</td>
<td>25.3775</td>
<td>0.0000</td>
<td>25.3775 - 25.3775</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qi</td>
<td>5</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 - 0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• The GOF stats of the above model are

\[ \chi^2 = 11.5, \quad G^2 = 13.2, \quad df = 5. \]

Not a good fit!

• If we assume the model, then \( \hat{\delta}_1 = 3.86, \hat{\delta}_2 = 0.60, \hat{\delta}_3 = 1.90 \). All are positive. So two raters agree more than disagree.

• Consider the event that each rater put one subject in category 2 and the other subject in category 3, then the conditional odds that raters agree rather than disagree is

\[ \hat{\tau}_{23} = e^{\hat{\delta}_2 + \hat{\delta}_3} = e^{0.60 + 1.90} = 12.3. \]
V.2 Quasi-symmetry model for rater agreement

- We know that symmetry models does not fit the data well (slide 402).
- Consider quasi-symmetry model

\[
\log(\pi_{ij}/\pi_{ji}) = \beta_i - \beta_j, \ i < j.
\]

- Estimates: \( \hat{\beta}_1 = -27.1679, \hat{\beta}_2 = -26.495, \hat{\beta}_3 = -28.668. \Rightarrow \)

\[
\begin{align*}
\hat{\pi}_{12}/\hat{\pi}_{21} &= e^{\hat{\beta}_1-\hat{\beta}_2} = 0.51 \\
\hat{\pi}_{13}/\hat{\pi}_{31} &= e^{\hat{\beta}_1-\hat{\beta}_3} = 4.48 \\
\hat{\pi}_{14}/\hat{\pi}_{41} &= e^{\hat{\beta}_1} = 0 \\
\hat{\pi}_{23}/\hat{\pi}_{32} &= e^{\hat{\beta}_2-\hat{\beta}_3} = 8.78 \\
\hat{\pi}_{24}/\hat{\pi}_{42} &= e^{\hat{\beta}_2} = 0 \\
\hat{\pi}_{34}/\hat{\pi}_{43} &= e^{\hat{\beta}_3} = 0
\end{align*}
\]

\( \Rightarrow \) Rater 1 tends to rate higher (4) than rater 2.
• SAS program and part of output:

```sas
data table8_7; set table8_7;
  if rater1=rater2 then delete;
  if rater1<rater2 then y=1;
  else y=0;
  if rater1<rater2 then do;
    ind1=rater1; ind2=rater2;
  end;
  else do;
    ind1=rater2; ind2=rater1;
  end;
array x {4};
  do k=1 to 4;
    if k=ind1 then
      x[k]=1;
    else if k=ind2 then
      x[k]=-1;
    else
      x[k]=0;
  end;
drop y1-y4 k;
run;

proc sort;
  by ind1 ind2 descending y;
run;

proc print;
run;
```
<table>
<thead>
<tr>
<th>Obs</th>
<th>rater1</th>
<th>rater2</th>
<th>count</th>
<th>qi</th>
<th>y</th>
<th>ind1</th>
<th>ind2</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>3</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>17</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>
title "Quasi-symmetry model";
proc genmod descending;
  freq count;
  model y = x1 x2 x3 / dist=bin link=logit aggregate 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>2</td>
<td>0.9783</td>
<td>0.4892</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>2</td>
<td>0.9783</td>
<td>0.4892</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>2</td>
<td>0.6219</td>
<td>0.3109</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>2</td>
<td>0.6219</td>
<td>0.3109</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</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>.</td>
<td>.</td>
</tr>
<tr>
<td>x1</td>
<td>1</td>
<td>-27.1679</td>
<td>0.9731</td>
<td>-29.0752 -25.2606</td>
<td>779.42</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>-26.4950</td>
<td>0.7628</td>
<td>-27.9900 -24.9999</td>
<td>1206.44</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>x3</td>
<td>0</td>
<td>-28.6680</td>
<td>0.0000</td>
<td>-28.6680 -28.6680</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

• GOF: Pearson $\chi^2 = 0.63$, Deviance $G^2 = 0.98$, $df = 2$, good fit.
V.3 Kappa measure of rater agreement

- Cohen’s *Kappa*:

\[ \kappa = \frac{\sum \pi_{ii} - \sum \pi_i \pi_i}{1 - \sum \pi_i \pi_i}. \]

The numerator = agreement probabilities - agreement expected under independence.

The denominator = maximum difference.

- Perfect agreement ⇔ \( \kappa = 1 \)

Random agreement ⇔ \( \kappa = 0 \).

- Replacing \( \pi_{ij} \)'s by the sample proportions \( p_{ij} \)'s leads to an estimate of \( \kappa \).

- For ordinal tables, using scores to emphasizes the disagreement ⇒ weighted \( \kappa \).

VI Bradley-Terry Model for Paired Preferences

- Example:

<table>
<thead>
<tr>
<th>Winner</th>
<th>Agassi</th>
<th>Federer</th>
<th>Henman</th>
<th>Hewitt</th>
<th>Roddick</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agassi</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Federer</td>
<td>6</td>
<td>-</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Henman</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hewitt</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Roddick</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

*Source: www.atptennis.com.*
• Let

\[ \Pi_{ij} = P[\text{Player } i \text{ wins Player } j]. \]

Consider Bradley-Terry model for comparison:

\[ \log \{ \Pi_{ij}/(1 - \Pi_{ij}) \} = \log \{ \Pi_{ij}/\Pi_{ji} \} = \beta_i - \beta_j, \quad i < j = 1, \ldots, I. \]

Need to set \( \beta_I = 0 \).

• We can rank players based on \( \beta_i \)'s.

• The above model can be fit by treating it as a quasi-symmetry model.
data table8_9;
  input winner player $ y1-y5;
cards;
  1 Agassi 0 0 1 1
  2 Federer 6 . 3 9 5
  3 Henman 0 1 . 0 1
  4 Hewitt 0 0 2 . 3
  5 Roddick 0 0 1 2 .
;

data table8_9; set table8_9;
  array temp {5} y1-y5;
  do loser=1 to 5;
    count=temp(loser);
    output;
  end;
run;

data table8_9; set table8_9;
  if winner=loser then delete;
  if winner<loser then do;
    y=1; ind1=winner; ind2=loser;
  end;
  else do ;
    y=0; ind1=loser; ind2=winner;
  end;

  array x {5};
  do k=1 to 5;
    if k=ind1 then 
      x[k]=1;
    else if k=ind2 then 
      x[k]=-1;
    else 
      x[k]=0;
  end;
  drop y1-y5 k;
run;
### CHAPTER 8

```sas
proc sort;
    by ind1 ind2 descending y;
run;

proc print;
run;
```

<table>
<thead>
<tr>
<th>Obs</th>
<th>winner</th>
<th>player</th>
<th>loser</th>
<th>count</th>
<th>y</th>
<th>ind1</th>
<th>ind2</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Agassi</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Federer</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>Agassi</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Henman</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>Agassi</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>Hewitt</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>Agassi</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>Roddick</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>Federer</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>Henman</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>Federer</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>Hewitt</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>Federer</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>Roddick</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>Henman</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>Hewitt</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>Henman</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>Roddick</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>Hewitt</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>Roddick</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
</tbody>
</table>
```

Slide 430
title "Bradley-Terry Model for Tennis Matches";
proc genmod descending;
  freq count;
  model y = x1 x2 x3 x4 / dist=bin link=logit aggregate noint covb;
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>5</td>
<td>8.1910</td>
<td>1.6382</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>5</td>
<td>8.1910</td>
<td>1.6382</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>5</td>
<td>11.6294</td>
<td>2.3259</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>5</td>
<td>11.6294</td>
<td>2.3259</td>
</tr>
</tbody>
</table>

Estimated Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>Prm2</th>
<th>Prm3</th>
<th>Prm4</th>
<th>Prm5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prm2</td>
<td>1.93092</td>
<td>1.06655</td>
<td>0.27405</td>
<td>0.40015</td>
</tr>
<tr>
<td>Prm3</td>
<td>1.06655</td>
<td>1.73340</td>
<td>0.34535</td>
<td>0.42773</td>
</tr>
<tr>
<td>Prm4</td>
<td>0.27405</td>
<td>0.34535</td>
<td>1.10898</td>
<td>0.32444</td>
</tr>
<tr>
<td>Prm5</td>
<td>0.40015</td>
<td>0.42773</td>
<td>0.32444</td>
<td>0.63787</td>
</tr>
</tbody>
</table>

Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>.</td>
</tr>
<tr>
<td>x1</td>
<td>1</td>
<td>1.4489</td>
<td>1.3896</td>
<td>-1.2747</td>
<td>4.1724</td>
<td>1.09</td>
</tr>
<tr>
<td>x2</td>
<td>1</td>
<td>3.8815</td>
<td>1.3166</td>
<td>1.3011</td>
<td>6.4620</td>
<td>8.69</td>
</tr>
<tr>
<td>x3</td>
<td>1</td>
<td>0.1875</td>
<td>1.0531</td>
<td>-1.8765</td>
<td>2.2515</td>
<td>0.03</td>
</tr>
<tr>
<td>x4</td>
<td>1</td>
<td>0.5734</td>
<td>0.7987</td>
<td>-0.9920</td>
<td>2.1387</td>
<td>0.52</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>.</td>
</tr>
</tbody>
</table>

Slide 431
• The GOF: $\chi^2 = 11.6$, Deviance $G^2 = 8.2$ with $df = 5$. Not a very good fit.

• Estimates of $\beta_i$'s:
  $\hat{\beta}_1 = 1.45, \hat{\beta}_2 = 3.88, \hat{\beta}_3 = 0.19, \hat{\beta}_4 = 0.57, \hat{\beta}_5 = 0$.

  $\Rightarrow \hat{\beta}_2 > \hat{\beta}_1 > \hat{\beta}_4 > \hat{\beta}_3 > \hat{\beta}_5$.

• We can estimate the winning probability that Player \( i \) wins against Player \( j \) \( \Pi_{ij} \):

\[
\hat{\Pi}_{ij} = \frac{e^{\hat{\beta}_i - \hat{\beta}_j}}{1 + e^{\hat{\beta}_i - \hat{\beta}_j}}.
\]

For example, consider Federer v.s. Agassi:

\[
\hat{\Pi}_{21} = \frac{e^{\hat{\beta}_2 - \hat{\beta}_1}}{1 + e^{\hat{\beta}_2 - \hat{\beta}_1}} = \frac{e^{3.88 - 1.45}}{1 + e^{3.88 - 1.45}} = 0.92.
\]

\[
\hat{\text{var}}(\hat{\beta}_2 - \hat{\beta}_1) = \hat{\text{var}}(\hat{\beta}_2) + \hat{\text{var}}(\hat{\beta}_1) - 2\hat{\text{cov}}(\hat{\beta}_2, \hat{\beta}_1)
\]

\[
= 1.73340 + 1.93092 - 2 \times 1.06655 = 1.5312
\]

\[
\hat{\text{SE}}(\hat{\beta}_2 - \hat{\beta}_1) = 1.24
\]

A 95% CI for \( \beta_2 - \beta_1 \):

\[
\hat{\beta}_2 - \hat{\beta}_1 \pm 1.96\hat{\text{SE}}(\hat{\beta}_2 - \hat{\beta}_1) = 2.43 \pm 1.96 \times 1.24 = [0, 4.86].
\]
A 95% CI for $\Pi_{21}$:

$$
\left[ \frac{e^0}{1 + e^0}, \frac{e^{4.86}}{1 + e^{4.86}} \right] = [0.5, 0.99].
$$

- **Note:** We can estimate $\Pi_{ij}$ based on the model even though Player $i$ may not have played Player $j$. For example, Agassi (Player 1) and Henman (Player 3) did not play in 2004-2005. But we can estimate the winning probability for Agassi *v.s.* Henman $\Pi_{13}$.

- **Note:** The above model can also be applied to other settings such as wine tasting.
9 Modeling Correlated, Clustered, Longitudinal Categorical Data

I GEE Models for Correlated/Clustered/Longitudinal Categorical Data

- Data: $y_{ij}$ (can be continuous, binary/binomial, count, etc), $i = 1, \ldots, m$ (# of subjects), $j = 1, \ldots, n_i$ ($n_i \geq 1$) (# of obs. for subject $i$) with mean and variance

$$\mu_{ij} = \mathbb{E}(y_{ij} | x_{ij}), \text{var}(y_{ij} | x_{ij}) = v(\mu_{ij}) \text{ (may be wrong)}$$

Denote

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix}, \quad \mu_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{in_i} \end{pmatrix}.$$
• Suppose we correctly specify the mean structure for data $y_{ij}$:

$$
g(\mu_{ij}) = \alpha + x_{1ij}\beta_1 + \ldots + x_{pij}\beta_p,
$$

• A GEE (generalized estimating equation) solves for $\beta = (\alpha, \beta_1, \ldots, \beta_p)^T$:

$$
S_\beta(\rho, \beta) = \sum_{i=1}^{m} \left( \frac{\partial \mu_i}{\partial \beta} \right)^T V_i^{-1}(y_i - \mu_i) = 0, \quad (9.1)
$$

where $V_i$ is some matrix (intended to specify for $\text{var}(y_i|x_i)$) and $\rho$ is the possible parameters in the correlation structure.

• The above estimating equation is \textbf{unbiased} no matter what matrix $V_i$ we use as long as the mean structure is right. That is

$$
E[S_\beta(\rho, \beta)] = 0.
$$

• Under some regularity conditions, the solution $\hat{\beta}$ from the above GEE
equation has asymptotic distribution

\[ \hat{\beta} \overset{a}{\sim} N(\beta, \Sigma), \]

where

\[ \Sigma = I_0^{-1}I_1I_0^{-1} \]

\[ I_0 = \sum_{i=1}^{m} D_i^T V_i^{-1} D_i \]

\[ I_1 = \sum_{i=1}^{m} D_i^T V_i^{-1} \text{var}(y_i|x_i)V_i^{-1} D_i \]

\[ = \sum_{i=1}^{m} D_i^T V_i^{-1}(y_i - \mu_i(\hat{\beta}))(y_i - \mu_i(\hat{\beta}))^T V_i^{-1} D_i \]

\( \Sigma \) is called the empirical, robust or sandwich variance estimate.

- If \( V_i \) is correctly specified, then \( I_1 \approx I_0 \) and \( \Sigma \approx I_0^{-1} \) (model based).
  In this case, \( \hat{\beta} \) is the most efficient estimate. Otherwise, \( \Sigma \neq I_0^{-1} \).
• The working variance matrix $V_i$ for $y_i$ (at $x_i$), can be decomposed as

$$V_i = A_i^{1/2} R_i A_i^{1/2},$$

where

$$A_i = \begin{pmatrix}
\text{var}(y_{i1}|x_{i1}) & 0 & \cdots & 0 \\
0 & \text{var}(y_{i2}|x_{i2}) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \text{var}(y_{in_i}|x_{in_i})
\end{pmatrix},$$

and $R_i$ is the correlation structure.

• We may try to specify $R_i$ so that it is close to the “true”. This $R_i$ is called the working correlation matrix and may be mis-specified.
Some working correlation structures

1. **Independent** \((\text{ind})\): \(R_i(\alpha) = I_{n_i \times n_i}\). No \(\rho\) needs to be estimated.

2. **Exchangeable** (compound symmetric) \((\text{exch})\):

\[
R_i = \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{bmatrix}
\]

Let \(e_{ij} = y_{ij} - \hat{\mu}_{ij}\). Since \(\mathbb{E}(e_{ij}e_{ik}) = \phi \rho\) (at true \(\beta\)), \(\implies\)

\[
\hat{\rho} = \frac{1}{(N^* - p - 1)\phi} \sum_{i=1}^{m} \sum_{j<k} e_{ij}e_{ik},
\]

where \(N^* = \sum_{i=1}^{m} n_i(n_i - 1)/2\) (total \# of pairs), \(\phi\) is usually estimated using the Pearson \(\chi^2\).
3. **AR(1) (ar(1))**: 

\[
R_i = \begin{bmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{n_i-1} \\
\rho & 1 & \rho & \cdots & \rho^{n_i-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \cdots & 1
\end{bmatrix}
\]

Since \(E(e_{ij}e_{i,j+1}) = \phi\rho\) (at true \(\beta\)), \(\Rightarrow\)

\[
\hat{\rho} = \frac{1}{(N^{**} - p - 1)\phi} \sum_{i=1}^{m} \sum_{j=1}^{n_i-1} e_{ij}e_{i,j+1},
\]

where \(N^{**} = \sum_{i=1}^{m} (n_i - 1)\) (total # of adjacent pairs).

4. **Unstructured (un)**: Let data determine \(R_i\).

- Many more can be found in *Proc GenMod* of SAS.
Key features of GEEs for analyzing longitudinal data

1. We only need to **correctly** specify how the mean of the outcome variable is related to the covariates of interest.

2. The correlation among the observations from the same subject over time is not the major interest and is treated as nuisance.

3. We can specify a correlation structure. The validity of the inference does not depend on the whether or not the specification of the correlation structure is correct. GEE gives us a robust inference on the regression coefficients, which is valid regardless whether or not the correlation structure we specified is right.

4. GEE calculates correct SEs for the regression coefficient estimates using *sandwich* estimates that take into account the possibility that the correlation structure is misspecified.

5. The regression coefficients in GEE have a population-average interpretation.

6. A fundamental assumption on missing data is that missing data
mechanism has to be MCAR (missing completely at random), while a likelihood-based approach (such as mixed model approach) only requires MAR (missing at random). The GEE approach will also be less efficient than a likelihood-based approach if the likelihood can be correctly specified.
Some popular GEE Models

- Continuous (Normal):

\[ \mu(x) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \]

where \( \mu(x) = \mathbb{E}(y|x) \) is the mean of outcome variable at \( x = (x_1, ..., x_p) \), such as mean of cholesterol level.

- Proportion (Binomial, Binary):

\[ \logit\{\pi(x)\} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \]

\[ \pi(x) = P[y = 1|x] = \mathbb{E}(y|x) \] such as disease risk.

\[ \logit(\pi) = \log\{\pi/(1 - \pi)\} \] is the logit link function. Other link functions are possible.
• Count or rate (Poisson-type)

\[
\log\{\lambda(x)\} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p
\]

\(\lambda(x)\) is the rate (e.g. \(\lambda(x)\) is the incidence rate of a disease) for the count data (number of events) \(y\) over a (time, space) region \(T\) such that

\[y|x \sim \text{Poisson}\{\lambda(x)T\}\]

Here \(\log(.)\) link is used. Other link functions are possible.

**Note**: For count data, we usually have to be concerned about the possible over-dispersion in the data. That is

\[\text{var}(y|x) > \text{E}(y|x)\]

With GEE, the over-dispersion is automatically taken into account.
II GEE Analysis of Longitudinal Binary/Binomial Data

- Example: longitudinal study of treatment for depression

Table 9.1. Cross-classification of Responses on Depression at Three Times (N = Normal, A = Abnormal) by Treatment and Diagnosis Severity

<table>
<thead>
<tr>
<th>Diagnosis Severity</th>
<th>Treatment</th>
<th>Response at Three Times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NNN</td>
</tr>
<tr>
<td>Mild</td>
<td>Standard</td>
<td>16</td>
</tr>
<tr>
<td>Mild</td>
<td>New drug</td>
<td>31</td>
</tr>
<tr>
<td>Severe</td>
<td>Standard</td>
<td>2</td>
</tr>
<tr>
<td>Severe</td>
<td>New drug</td>
<td>7</td>
</tr>
</tbody>
</table>

• Proportion of normal response rates over time:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>New Drug</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>33%</td>
<td>34%</td>
</tr>
<tr>
<td>Week 2</td>
<td>63%</td>
<td>42%</td>
</tr>
<tr>
<td>Week 4</td>
<td>89%</td>
<td>56%</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>180</td>
</tr>
</tbody>
</table>

• We could analyze data at each time point using ML ⇒ multiple test issues, no way to assess time effect.

• Assessment of the treatment effect over time should take into account the correlation of 3 observations from each patient.
• Let $s = 1/0$ for severe/mild, $d = 1/0$ for new drug and standard, 
  $t = \log_2(\text{week})$ time in $\log_2$ scale, and 
  $$\pi(s, d, t) = P[Y_t = 1|s, d, t].$$

• Consider the following logistic model 
  $$\text{logit}\{\pi(s, d, t)\} = \alpha + \beta_1 s + \beta_2 d + \beta_3 t + \beta_4 (d \times t).$$ 

  The correlation is taken into account using GEE approach. Here we used \textit{unstructured} working correlation matrix. May use \textit{exchangeable} as in the textbook. Results are similar

• SAS program and part of output:

  ```sas
  data table9_1;
  input severity $ treatment $ y1-y8;
  cards;
  Mild Standard 16 13 9 3 14 4 15 6
  Mild Newdrug 31 0 6 0 22 2 9 0
  Severe Standard 2 2 8 9 9 15 27 28
  Severe Newdrug 7 2 5 2 31 5 32 6
  ;
  run;
  ```
title "Recover individual data";
data table9_1; set table9_1;
  array temp {8} y1-y8;
  
  trt = (treatment="Newdrug");
  sev = (severity="Severe");
  retain id;
  if _n_=1 then id=0;

  do k=1 to 8;
    do i=1 to temp(k);
      id = id + 1;
      do j=1 to 3;
        time=j-1;
        if k=1 then y = 1;
        if k=2 then y = (j ne 3);
        if k=3 then y = (j ne 2);
        if k=4 then y = (j = 1);
        if k=5 then y = (j ne 1);
        if k=6 then y = (j = 2);
        if k=7 then y = (j = 3);
        if k=8 then y = 0;
        output;
      end;
    end;
  end;
  run;

  title "Treatment for Depression: Table 9.1";
  proc genmod descending;
    class id;
    model y = sev trt time trt*time / dist=bin link=logit;
    repeated subject=id / type=un corrw;
  run;
Working Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.0747</td>
<td>-0.0277</td>
</tr>
<tr>
<td>Row2</td>
<td>0.0747</td>
<td>1.0000</td>
<td>-0.0573</td>
</tr>
<tr>
<td>Row3</td>
<td>-0.0277</td>
<td>-0.0573</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Analysis Of GEE Parameter Estimates

Empirical Standard Error Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z | Pr > |Z| |
|-----------|----------|----------------|-----------------------|---|------|---|
| Intercept | -0.0255  | 0.1726         | -0.3638 0.3128       | -0.15 | 0.8826 |
| sev       | -1.3048  | 0.1450         | -1.5890 -1.0206      | -9.00 | <.0001 |
| trt       | -0.0543  | 0.2271         | -0.4995 0.3908       | -0.24 | 0.8109 |
| time      | 0.4758   | 0.1190         | 0.2425 0.7091        | 4.00  | <.0001 |
| trt*time  | 1.0129   | 0.1865         | 0.6473 1.3785        | 5.43  | <.0001 |

- The odds-ratio $\theta(s,t)$ of having a normal response between patients receiving new drug and standard drug is

$$\logit\{\pi(s,d = 1,t)\} = \alpha + \beta_1 s + \beta_2 \times 1 + \beta_3 t + \beta_4 (1 \times t)$$

$$\logit\{\pi(s,d = 0,t)\} = \alpha + \beta_1 s + \beta_2 \times 0 + \beta_3 t + \beta_4 (0 \times t)$$

$$\logit\{\pi(s,d = 1,t)\} - \logit\{\pi(s,d = 0,t)\} = \beta_4 t + \beta_2$$

$$\theta(s,t) = e^{\beta_4 t + \beta_2}$$
• The estimated odds-ratios are:
  \[ e^{1.01 \times 0 - 0.05} = 0.95 \] at week 1,
  \[ e^{1.01 \times 1 - 0.05} = 2.61 \] at week 2,
  \[ e^{1.01 \times 2 - 0.05} = 7.17 \] at week 4.

The new drug is much better at week 4 than the standard drug.

• Working correlation: \( \hat{\rho}_{12} = 0.07, \hat{\rho}_{13} = -0.03, \hat{\rho}_{23} = 0.06. \)

• Note: If there is baseline response \( Y \), we can put it as part of the outcome \( Y \) and model the change since baseline.
III GEE Analysis of Clustered Binary/Binomial Data

• Example (Table 9.4): Low-iron rat study where iron-deficient female rats were assigned to 4 groups:
  Group 1: untreated (control)
  Group 2: injection of iron supplement on days 7, 10
  Group 3: injection on days 0, 7
  Group 4: injection weekly

• Data:
  \[ y_{ig} = \# \text{ of dead baby rats out of } n_{ig} \text{ baby rats in litter} \]
  \[ i = 1, 2, \ldots, k_g, \; g = 1, 2, 3, 4. \]

  \[ y_{ig} \sim \text{Bin}(n_{ig}, \pi_g) ? \]

  If \( E(y_{ig}) = n_{ig}\pi_g \), is \( \text{var}(y_{ig}) = n_{ig}\pi_g(1 - \pi_g) \) true?
Table 9.4. Response Counts of (Litter Size, Number Dead) for 58 Litters of Rats in a Low-Iron Teratology Study

<table>
<thead>
<tr>
<th>Group 1: untreated (low iron)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1) (11, 4) (12, 9) (4, 4) (10, 10) (11, 9) (9, 9) (11, 11) (10, 10) (10, 7) (12, 12)</td>
</tr>
<tr>
<td>(10, 9) (8, 8) (11, 9) (6, 4) (9, 7) (14, 14) (12, 7) (11, 9) (13, 8) (14, 5) (10, 10)</td>
</tr>
<tr>
<td>(12, 10) (13, 8) (10, 10) (14, 3) (13, 13) (4, 3) (8, 8) (13, 5) (12, 12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2: injections days 7 and 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1) (3, 1) (13, 1) (12, 0) (14, 4) (9, 2) (13, 2) (16, 1) (11, 0) (4, 0) (1, 0) (12, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 3: injections days 0 and 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(8, 0) (11, 1) (14, 0) (14, 1) (11, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 4: injections weekly</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 0) (13, 0) (9, 2) (17, 2) (15, 0) (2, 0) (14, 1) (8, 0) (6, 0) (17, 0)</td>
</tr>
</tbody>
</table>

• We could model binomial data, but need to account for over-dispersion (Table 9.5 under Binomial ML did not account for overdispersion):

data rat;
  input litter group n n1;
  gp1 = (group=1); gp2 = (group=2); gp3 = (group=3); gp4 = (group=4);
  n0 = n-n1;
datalines;
  1 1 10 1
  2 1 11 4
  3 1 12 9
  4 1 14 4
  5 1 10 10
  6 1 11 9
  7 1 9 9
...

proc genmod data=rat;
  class group;
  model n1/n = gp2 gp3 gp4 / dist=bin link=logit scale=pearson;
run;

************************************************************************************
Analysis Of Maximum Likelihood 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>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.1440</td>
<td>0.2187</td>
<td>0.7154 - 1.5726</td>
<td>27.37</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gp2</td>
<td>1</td>
<td>-3.3225</td>
<td>0.5600</td>
<td>-4.4201 - 2.2250</td>
<td>35.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>gp3</td>
<td>1</td>
<td>-4.4762</td>
<td>1.2375</td>
<td>-6.9017 - 2.0507</td>
<td>13.08</td>
<td>0.0003</td>
</tr>
<tr>
<td>gp4</td>
<td>1</td>
<td>-4.1297</td>
<td>0.8061</td>
<td>-5.7095 - 2.5498</td>
<td>26.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.6926</td>
<td>0.0000</td>
<td>1.6926 - 1.6926</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
- We could also model the original binary data, but need to account for correlation:

```sas
title "Recover individual rat’s data";
data rat2; set rat;
do i=1 to n1;
y=1;
output;
end;
do i=1 to n0;
y=0;
output;
end;
run;

title "GEE for individual rat’s data";
Proc Genmod data=rat2 descending;
class litter group;
model y = gp2 gp3 gp4 / dist=bin link=logit;
repeated subject=litter / type=exch corrw;
run;
```
• Working correlation: $\hat{\rho} = 0.19$. Estimates of regression coefficients are similar to before.

• $e^{\hat{\beta}_2} = e^{-3.3692} = 0.034 \Rightarrow$ the odds of death for group 2 is about 0.034 times the odds of death for group 1.
IV GEE Analysis of Longitudinal Count Data

- Example: progabide trial on epileptic seizure patients.
  In the progabide trial, 59 epileptics were randomly assigned to receive the anti-epileptic treatment (progabide) or placebo. The number of seizure counts was recorded in 4 consecutive 2-week intervals. Age and baseline seizure counts (in an eight week period prior to the treatment assignment) were also recorded.

Study objectives:
1. Does the treatment work?
2. What is the treatment effect adjusting for available covariates?

Features of this data set:
1. Outcome is count data, implying a Poisson regression.
2. Baseline seizure counts were for 8 weeks, as opposed to 2 weeks for other seizure counts.
3. Randomization may be taken into account in the data analysis.
A glimpse of the seizure data

Print the first 20 observations:

<table>
<thead>
<tr>
<th>Obs</th>
<th>id</th>
<th>seize</th>
<th>trt</th>
<th>visit</th>
<th>interval</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101</td>
<td>76</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>101</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>102</td>
<td>38</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>102</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>102</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>102</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>11</td>
<td>103</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>103</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>13</td>
<td>103</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>103</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>15</td>
<td>103</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>16</td>
<td>104</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>104</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>18</td>
<td>104</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>19</td>
<td>104</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>20</td>
<td>104</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>31</td>
</tr>
</tbody>
</table>
Epileptic seizure counts from the progabide trial

Seizure counts for progabide arm

Seizure counts for control arm

Order of visit

Seizure counts

Seizure counts
• Data:
  ⋆ 59 patients, 28 in control group, 31 in treatment (progabide) group.
  ⋆ 5 seizure counts (including baseline) were obtained.
  ⋆ Covariates: treatment (covariate of interest), age.

• GEE Poisson model: $y_{ij} =$seizure counts obtained at the $j$th
  ($j = 1, \ldots, 5$) time point for patient $i$, $y_{ij} \sim$ over-dispersed
  Poisson$(\mu_{ij})$, $\mu_{ij} = E(y_{ij}) = t_{ij} \lambda_{ij}$, where $t_{ij}$ is the length of time
  from which the seizure count $y_{ij}$ was observed, $\lambda_{ij}$ is hence the rate to
  have a seizure. First consider model

$$
\log(\lambda_{ij}) = \beta_0 + \beta_1 I(j > 1) + \beta_2 \text{trt}_i + \beta_3 \text{trt}_i I(j > 1)
$$

$$
\log(\mu_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 I(j > 1) + \beta_2 \text{trt}_i + \beta_3 \text{trt}_i I(j > 1)
$$

Note that $\log(t_{ij})$ is an offset.
• Interpretation of $\beta$’s:

<table>
<thead>
<tr>
<th>Group</th>
<th>log of seizure rate $\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (trt=0)</td>
<td>Before randomization</td>
</tr>
<tr>
<td></td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Treatment (trt=1)</td>
<td>$\beta_0 + \beta_2$</td>
</tr>
</tbody>
</table>

Therefore, $\beta_1 =$ time + placebo effect, $\beta_2 =$ difference in seizure rates at baseline between two groups, $\beta_3 =$ treatment effect of interest after taking into account of time + placebo effect.

If randomization is taken into account ($\beta_2 = 0$), we can consider the following model

$$
\log(\mu_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 I(j > 1) + \beta_2 trt_i I(j > 1)
$$
data seize;
    infile "seize.dat";
    input id seize visit trt age;
    nobs=_n_;  
    interval = 2;  
    if visit=0 then interval=8;  
    logtime = log(interval);
    assign = (visit>0);
run;

proc genmod data=seizure;
    class id;
    model seize = assign trt assign*trt
        / dist=poisson link=log offset=logtime;
    repeated subject=id / type=exch corrw;
run;

Working Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
<th>Col5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
</tr>
<tr>
<td>Row2</td>
<td>0.7716</td>
<td>1.0000</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
</tr>
<tr>
<td>Row3</td>
<td>0.7716</td>
<td>0.7716</td>
<td>1.0000</td>
<td>0.7716</td>
<td>0.7716</td>
</tr>
<tr>
<td>Row4</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
<td>1.0000</td>
<td>0.7716</td>
</tr>
<tr>
<td>Row5</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
<td>0.7716</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter  | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z|  |
|------------|----------|----------------|-----------------------|--------|---|
| Intercept  | 1.3476   | 0.1574         | 1.0392 1.6560         | 8.56   | <.0001 |
| assign     | 0.1108   | 0.1161         | -0.1168 0.3383         | 0.95   | 0.3399 |
| trt        | 0.0265   | 0.2219         | -0.4083 0.4613         | 0.12   | 0.9049 |
| assign*trt | -0.1037  | 0.2136         | -0.5223 0.3150         | -0.49  | 0.6274 |
title "Model 2: take randomization into account";
proc genmod data=seizure;
   class id;
   model seize = assign assign*trt
       / dist=poisson link=log offset=logtime scale=pearson aggregate=nobs;
   repeated subject=id / type=exch corrw;
run;

<table>
<thead>
<tr>
<th>Row</th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
<th>Col5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
</tr>
<tr>
<td>Row2</td>
<td>0.7750</td>
<td>1.0000</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
</tr>
<tr>
<td>Row3</td>
<td>0.7750</td>
<td>0.7750</td>
<td>1.0000</td>
<td>0.7750</td>
<td>0.7750</td>
</tr>
<tr>
<td>Row4</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
<td>1.0000</td>
<td>0.7750</td>
</tr>
<tr>
<td>Row5</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
<td>0.7750</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter    | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z| |
|--------------|----------|----------------|-----------------------|--------|---|
| Intercept    | 1.3616   | 0.1111         | 1.1438                | 1.5794 | 12.25 | <.0001 |
| assign       | 0.1173   | 0.1283         | -0.1341               | 0.3688 | 0.91  | 0.3604 |
| assign*trt   | -0.1170  | 0.2076         | -0.5240               | 0.2900 | -0.56 | 0.5731 |
V  GEE Analysis of Longitudinal Ordinal Data

- Data from Insomnia Clinical Trial (Table 9.6 on page 285)

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Initial</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 20</td>
<td>20 – 30</td>
</tr>
<tr>
<td>Active</td>
<td>&lt; 20</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>20 – 30</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>30 – 60</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>&gt; 60</td>
<td>9</td>
</tr>
<tr>
<td>Placebo</td>
<td>&lt; 20</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>20 – 30</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>30 – 60</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>&gt; 60</td>
<td>4</td>
</tr>
</tbody>
</table>
• Consider the cumulative logit model for $Y$ at each occasion:

$$\text{logit}\{P[Y_{ij} \leq k]\} = \alpha_k + \beta_1 I(j = 2) + \beta_2 \text{trt}_i + \beta_3 I(j = 2) \times \text{trt}_i,$$

$$i = 1, 2, ..., 239, \quad j = 1, 2, \quad k = 1, 2, 3.$$

• Interpretation of $\beta_1, \beta_2, \beta_3$:

1. $\beta_1$: Effect of time + placebo
2. $\beta_2$: Group difference at baseline (can be set to 0 by randomization)
3. $\beta_3$: Treatment effect after taking into account the time and placebo effects.
• SAS program and part of output:

```sas
data table9_6;
  input trt y0 y1-y4;
  cards;
  1 1 7 4 1 0
  1 2 11 5 2 2
  1 3 13 23 3 1
  1 4 9 17 13 8
  0 1 7 4 2 1
  0 2 14 5 1 0
  0 3 6 9 18 2
  0 4 4 11 14 22
; title "Recover individual data";
data table9_6; set table9_6;
  array temp {4} y1-y4;
  retain id;
  if _n_=1 then id=0;
  do k=1 to 4;
    do i=1 to temp(k);
      id = id + 1;
      do time=0 to 1;
        if time=0 then y=y0;
        else y=k;
        if y=1 then ttfa=10;
        else if y=2 then ttfa=25;
        else if y=3 then ttfa=45;
        else ttfa=75;
        output;
      end;
    end;
  end;
run;
```
title "GEE cumulative logit model for insomnia longitudinal data";
proc GenMod data=table9_6;
  class id;
  model y = time trt time*trt / dist=multinomial link=clogit;
  repeated subject=id / type=ind;
run;

***********************************************************************
Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter  | Estimate | Standard Error | 95% Confidence 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|
| time       | 1.0381   | 0.1676         | 0.7096 1.3665         | 6.19  | <.0001|
| trt        | 0.0336   | 0.2384         | -0.4337 0.5009        | 0.14  | 0.8879|
| time*trt   | 0.7078   | 0.2435         | 0.2305 1.1850         | 2.91  | 0.0037|

- **Note**: We can only specify *independence* working correlation matrix for ordinal longitudinal data. However, the SE’s for $\hat{\beta}$’s are correct even if this working correlation is (likely) wrong.
What we see from the output:

1. There is a strong time + placebo effect: $\hat{\beta}_1 = 1.038 (SE = 0.17)$. The odds of having shorter time to falling asleep for placebo patients 2 weeks later is $e^{\hat{\beta}_1} = e^{1.038} = 2.8$ times their odds at baseline.

2. There is not much group difference at baseline (p-value = 0.88), which is expected.

3. Strong evidence of treatment effect: $\hat{\beta}_3 = 0.71 (SE = 0.24)$. $e^{\hat{\beta}_1 + \hat{\beta}_3} = e^{1.746} = 5.7$: the odds that treated patients have shorter time to falling asleep 2 weeks later is 5.7 times their odds at baseline.
• Assign scores (midpoints) 10, 25, 45, 75 for the 4 categories of $Y$, representing the actual *time to falling asleep*. Denote it by $Y^*$ and consider the model:

$$E\{Y_{ij}^*\} = \alpha + \beta_1 I(j = 2) + \beta_2 \text{trt}_i + \beta_3 I(j = 2) \times \text{trt}_i,$$

$$i = 1, 2, \ldots, 239, \quad j = 1, 2, \quad k = 1, 2, 3.$$

• Interpretation of $\beta_1, \beta_2, \beta_3$:

1. $\beta_1$: Effect of time + placebo
2. $\beta_2$: Group difference at baseline (can be set to 0 by randomization)
3. $\beta_3$: Treatment effect after taking into account the time and placebo effects.
title "GEE model using scores for time to falling asleep";
proc GenMod data=table9_6;
    class id;
    model ttfa = time trt time*trt / dist=normal;
    repeated subject=id / type=un;
run;

***********************************************************************

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z | Pr > |Z|
|-----------|----------|----------------|-----------------------|---|------|
| Intercept |  50.3333 |     2.1673      | 46.0856 - 54.5811     | 23.22 | <.0001 |
| time      | -12.9583 |     2.0535      | -16.9832 - -8.9335    | -6.31 | <.0001 |
| trt       |  -0.3754 |     3.0134      | -6.2815 - 5.5308      | -0.12 |  0.9009 |
| time*trt  |  -9.2265 |     3.0275      | -15.1604 - -3.2927    | -3.05 |  0.0023 |
What we see from the output:

1. There is a strong time + placebo effect: \( \hat{\beta}_1 = -13 (SE = 2.05) \). The average time to falling asleep for patients receiving placebo 2 weeks later is about 13 minutes shorter than baseline.

2. There is not much difference in time to falling asleep between 2 groups at baseline (p-value = 0.9), which is expected.

3. Strong evidence of treatment effect: \( \hat{\beta}_3 = 9.2 (SE = 3.0) \). The average reduced time to falling asleep for treated patients is 9.2 minutes shorter than untreated patients (so the actual reduction compared to baseline for treated patients is about: 13 + 9.2 = 22.2 minutes).
VI Transitional Models

VI.1 Use previous responses as covariates

- In a longitudinal study with time \( t = 1, 2, \cdots \), for each individual, we have response variables \( \{y_1, y_2, \cdots, y_t, \cdots\} \).

- We may model \( Y_t \) given the past \( \{y_1, y_2, \cdots, y_{t-1}\} \) and covariates \( x_1, x_2, \cdots, x_k \). Usually, the correlation in \( \{Y_t\} \)'s can be totally explained by the past \( \Rightarrow \{Y_t\} \)'s are conditionally independent given the past \( \Rightarrow \text{Markov chain} \).

- In the above Markov chain model, we may assume that \( Y_t \) only depends on \( y_{t-1} \), this is the Markov chain with order = 1.

- When \( Y \) is binary, the above Markov model with order 1 may be

\[
\logit\{P[Y_t = 1]\} = \alpha + \beta y_{t-1} + \beta_1 x_1 + \cdots + \beta_k x_k.
\]

- Transitional models are good for prediction.
Example: Child's respiratory illness and maternal smoking (Table 9.8)

<table>
<thead>
<tr>
<th>Child's Respiratory Illness</th>
<th>No Maternal Smoking</th>
<th>Maternal Smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age 10</td>
<td>Age 10</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>237</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Age 9</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Age 8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Age 7</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Source: Thanks to Dr. James Ware for these data.
• Let $Y_t$ be respiratory illness (1/0) at age $t$ and consider transitional model

$$\text{logit}\{P[Y_t = 1]\} = \alpha + \beta y_{t-1} + \beta_1 \text{smoke} + \beta_2 t, \quad t = 8, 9, 10.$$  

• Since $t = 8, 9, 10$, baseline data ($t = 7$) is deleted!

• If deleting baseline data results in deleting subjects, this analysis may be invalid and less efficient!

• SAS program and part of output:

```sas
data table9_8;
  input y7 y8 y9 count1-count4;
cards;
  0  0  0  237 10 118 6
  0  0  1  15  4  8  2
  0  1  0  16  2 11  1
  0  1  1  7  3  6  4
  1  0  0  24  3  7  3
  1  0  1  3  2  3  1
  1  1  0  6  2  4  2
  1  1  1  5 11  4  7;
```

title "Recover individual data"

data table9_8; set table9_8;
  array smk0 {2} count1-count2;
  array smk1 {2} count3-count4;
  array y7_9 {3} y7-y9;

retain id;
if _n_=1 then id=0;

do j=1 to 2;
  do i=1 to smk0[j];
    id = id+1;
    smoke = 0;
    do k=1 to 4;
      age=k+6;
      if k<4 then y=y7_9[k];
      if k=4 then y=j-1;
      output;
    end;
  end;
end;

do j=1 to 2;
  do i=1 to smk1[j];
    id = id+1;
    smoke = 1;
    do k=1 to 4;
      age=k+6;
      if k<4 then y=y7_9[k];
      if k=4 then y=j-1;
      output;
    end;
  end;
end;
run;
data lagdata; set table9_8;
    by id age;
    lagy=lag(y);

    retain basey;
    if first.id then do;
        lagy = .;
        basey = y;
    end;
run;

proc print data=lagdata (firstobs=2001 obs=2020);
    var id y lagy basey age smoke;
run;

*****************************************************************
Obs  id  y  lagy  basey  age  smoke
2001 501   1   .    1     7     0
2002 501   1   1    1     8     0
2003 501   0   0    1     9     0
2004 501   0   0    1    10     0
2005 502   1   .    1     7     0
2006 502   1   1    1     8     0
2007 502   0   1    1     9     0
2008 502   0   0    1    10     0
2009 503   1   .    1     7     0
2010 503   1   1    1     8     0
2011 503   0   1    1     9     0
2012 503   1   0    1    10     0
2013 504   1   .    1     7     0
2014 504   1   1    1     8     0
2015 504   0   1    1     9     0
2016 504   1   0    1    10     0
2017 505   1   .    1     7     1
2018 505   1   1    1     8     1
2019 505   0   1    1     9     1
2020 505   0   0    1    10     1

title "Transitional model for respiratory illness";
proc genmod data=lagdata descending;
   class id;
   model y = lagy smoke age / dist=bin link=logit;
run;
******************************************************************************

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Wald Chi-Square</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.2926</td>
<td>0.8460</td>
<td>-1.9508 - 1.3656</td>
<td>0.12</td>
<td>0.7295</td>
</tr>
<tr>
<td>lagy</td>
<td>1</td>
<td>2.2111</td>
<td>0.1582</td>
<td>1.9010 - 2.5211</td>
<td>195.36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>smoke</td>
<td>1</td>
<td>0.2960</td>
<td>0.1563</td>
<td>-0.0105 - 0.6024</td>
<td>3.58</td>
<td>0.0583</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>-0.2428</td>
<td>0.0947</td>
<td>-0.4283 - 0.0573</td>
<td>6.58</td>
<td>0.0103</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 - 1.0000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Obviously, previous year's respiratory illness status is a very strong predictor for current year's respiratory illness. The odds-ratio of having a respiratory illness at any year is $e^{2.21} = 9.1$ between children with or without a respiratory illness at the previous year.

- Maternal smoking has a marginally significant effect. Age has a significant negative effect.
• **Note:** If we model 4 longitudinal data points for each child, we have to take into account the correlation using, say, GEE:

```plaintext
proc genmod data=table9_8 descending;
  class id;
  model y = smoke age / dist=bin link=logit;
  repeated subject=id / type=exch corrw;
run;
```

Working Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
<th>Col4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.3541</td>
<td>0.3541</td>
<td>0.3541</td>
</tr>
<tr>
<td>Row2</td>
<td>0.3541</td>
<td>1.0000</td>
<td>0.3541</td>
<td>0.3541</td>
</tr>
<tr>
<td>Row3</td>
<td>0.3541</td>
<td>0.3541</td>
<td>1.0000</td>
<td>0.3541</td>
</tr>
<tr>
<td>Row4</td>
<td>0.3541</td>
<td>0.3541</td>
<td>0.3541</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Analysis Of GEE Parameter Estimates

| Parameter | Estimate | Standard Error | 95% Confidence Limits | Z Pr > |Z| |
|-----------|----------|----------------|-----------------------|------|-----|
| Intercept | -0.8600  | 0.3805         | -1.6057 -0.1142      | -2.26 | 0.0238 |
| smoke     | 0.2651   | 0.1777         | -0.0833 0.6135       | 1.49 | 0.1359 |
| age       | -0.1134  | 0.0439         | -0.1993 -0.0274      | -2.59 | 0.0097 |

• The estimated correlation is \( \hat{\rho} = 0.354 \).
VI.2 Use baseline response as a covariate

- We may use the baseline response variable as a covariate. However, we have to delete the baseline data for each individual.
- For example, for the respiratory illness data, we may consider
  \[
  \logit\{P[Y_t = 1]\} = \alpha + \beta y_7 + \beta_1 \text{smoke} + \beta_2 t, \quad t = 8, 9, 10.
  \]
- In this case, we need to account for the correlation in $Y$'s using, say, GEE.
- If deleting baseline data results in deleting subjects, this analysis may be invalid and less efficient!
data lagdata; set lagdata;
by id age;
if first.id then delete;
run;

title "Use baseline response as a covariate";
proc genmod data=lagdata descending;
   class id;
   model y = basey smoke age / dist=bin link=logit;
   repeated subject=id / type=exch corrw;
run;

********************************************************************
Working Correlation Matrix
<table>
<thead>
<tr>
<th>Col1</th>
<th>Col2</th>
<th>Col3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>1.0000</td>
<td>0.2755</td>
</tr>
<tr>
<td>Row2</td>
<td>0.2755</td>
<td>1.0000</td>
</tr>
<tr>
<td>Row3</td>
<td>0.2755</td>
<td>0.2755</td>
</tr>
</tbody>
</table>

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

| Parameter | Estimate | Error | 95% Confidence Limits | Z | Pr > |Z|
|-----------|----------|-------|-----------------------|---|------|
| Intercept | -0.2867  | 0.7046 | -1.6677 1.0942       | -0.41 | 0.6840 |
| basey     | 1.9012   | 0.2042 | 1.5009 2.3014        | 9.31 | <.0001 |
| smoke     | 0.3851   | 0.1921 | 0.0086 0.7616        | 2.00 | 0.0450 |
| age       | -0.2340  | 0.0784 | -0.3877 -0.0802      | -2.98 | 0.0029 |

- Similar results as those from *Markov* model.
10 Random Effects: Generalized Linear Mixed Models (GLMMs)

I GLMMs for Binary/ Binomial Clustered/ Longitudinal Data

I.1 GLMMs for binary matched data from a prospective study

- Table 8.1 revisited:

<table>
<thead>
<tr>
<th>Pay higher taxes ($Y_1$)</th>
<th>Yes (1)</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes (1)</td>
<td>227</td>
<td>132</td>
</tr>
<tr>
<td>No (0)</td>
<td>107</td>
<td>678</td>
</tr>
</tbody>
</table>

Cut living standard ($Y_2$)

<table>
<thead>
<tr>
<th>Yes (1)</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>334</td>
<td>810</td>
</tr>
</tbody>
</table>
• Data for individual $i$

<table>
<thead>
<tr>
<th>$X$</th>
<th>Yes (1)</th>
<th>No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pay higher taxes (1)</td>
<td>$y_{i1}$</td>
<td>$1 - y_{i1}$</td>
</tr>
<tr>
<td>Cut living standard (0)</td>
<td>$y_{i2}$</td>
<td>$1 - y_{i2}$</td>
</tr>
</tbody>
</table>

• Let $\pi_i(x) = P[Y_{ij} = 1|x, \alpha_i]$ the individual probability of responding “Yes” to question $j$ and consider the logit model:

$$\text{logit}\{\pi_i(x)\} = \alpha_i + \beta x,$$

where $\alpha_i$ is specific to subject $i$. Since subject $i$ is a random subject drawn from the population, it is natural to assume $\alpha_i \sim N(\alpha, \sigma^2)$.

• Let $u_i = \alpha_i - \alpha$. Then $u_i \sim N(0, \sigma^2)$ and the model becomes

$$\text{logit}\{\pi_i(x)\} = \alpha + u_i + \beta x.$$

This is a special case of GLMM – logistic-normal model.
• In the above model, $\alpha, \beta$ are called *fixed effects*, $u_i$’s are called *random effects*. The fixed effects are the parameters of major interest.

• Interpretation of $\beta$: $e^{\beta} = $ odds ratio of responding “Yes” between question 1 and question 2 for any subject $i$. The comparison is on *subject level, not population level*!

• However, approximately on population level, we have:

$$\logit\{P[Y = 1]\} \approx (1 + 0.346\sigma^2)^{-1/2} \times (\alpha + \beta x).$$

That is, approximately, $e^{(1+0.346\sigma^2)^{-1/2}\beta}$ is the population odds-ratio of responding “Yes” between question 1 and question 2.
• **Note 1:** In the above model, we usually assume that $Y_{i1}, Y_{i2}$ are conditionally independent given random effects $u_i$. However, marginally $Y_{i1}, Y_{i2}$ are correlated. The correlation is induced by the shared random effect $u_i$. The variance $\sigma^2$ of $u_i$ characterizes the magnitude of between-subject variance, and hence the correlation. Greater $\sigma^2$ corresponds to greater marginal correlation between $Y_{i1}$ and $Y_{i2}$.

• **Note 2:** We could also estimate random effects $u_i$ by borrowing information from other subjects (taking into account $u_i \sim N(0, \sigma^2)$). This method is different from *treating $u_i$ as parameters*. The only model parameters are $\alpha, \beta$ and $\sigma^2$. 
• SAS program and part of output:

```sas
data table8_1;
  input payht y1 y2;
  cards;
  1 227 132
  0 107 678
;

data table8_1; set table8_1;
  array temp [2] y1-y2;
  do j=1 to 2;
    count=temp(j);
    cutls = 2-j;
    output;
  end;
run;

title "Recover individual data";
data newdata; set table8_1;
  retain id;
  if _n_=1 then id=0;
  do i=1 to count;
    id = id+1;
    do question=1 to 2;
      x = 2-question;
      if question=1 then
        y=payht;
      else
        y=cutls;
      output;
    end;
  end;
run;
```
title "Use mixed model for matched opinion data";
proc glimmix data=newdata method=quad;
    class id;
    model y = x / dist=bin link=logit s;
    random int / subject=id type=vc;
run;

Use mixed model for matched opinion data

The GLIMMIX Procedure

Model Information

Data Set WORK.NEWDATA
Response Variable y
Response Distribution Binomial
Link Function Logit
Variance Function Default
Variance Matrix Blocked By id
Estimation Technique Maximum Likelihood
Likelihood Approximation Gauss-Hermite Quadrature
Degrees of Freedom Method Containment

Iteration History

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Restarts</th>
<th>Evaluations</th>
<th>Objective Function</th>
<th>Change</th>
<th>Max Gradient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2585.9233051</td>
<td></td>
<td>150.1262</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2555.3944038</td>
<td>30.52890133</td>
<td>58.06731</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2545.5849822</td>
<td>9.80942165</td>
<td>28.41184</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2534.5126265</td>
<td>11.07235569</td>
<td>15.44879</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>2521.9729972</td>
<td>12.53962923</td>
<td>12.94123</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>4</td>
<td>2520.5584416</td>
<td>1.41455560</td>
<td>1.495088</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>3</td>
<td>2520.5440308</td>
<td>0.01441087</td>
<td>0.114691</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>3</td>
<td>2520.5439581</td>
<td>0.00007268</td>
<td>0.005691</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>2520.5439579</td>
<td>0.00000222</td>
<td>0.002225</td>
</tr>
</tbody>
</table>

Slide 485
Convergence criterion (GCONV=1E-8) satisfied.

**Fit Statistics**

-2 Log Likelihood 2520.54  
AIC (smaller is better) 2526.54  
AICC (smaller is better) 2526.55  
BIC (smaller is better) 2541.67  
CAIC (smaller is better) 2544.67  
HQIC (smaller is better) 2532.26

**Fit Statistics for Conditional Distribution**

-2 log L(y | r. effects) 1041.77  
Pearson Chi-Square 702.92  
Pearson Chi-Square / DF 0.31

**Covariance Parameter Estimates**

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>id</td>
<td>8.1120</td>
<td>1.2028</td>
</tr>
</tbody>
</table>

The GLIMMIX Procedure

**Solutions for Fixed Effects**

| Effect | Estimate | Standard Error | DF | t Value | Pr > |t| |
|--------|----------|----------------|----|---------|-------|
| Intercept | -1.8361 | 0.1614         | 1143 | -11.38  | <.0001 |
| x       | 0.2094   | 0.1299         | 1143 | 1.61    | 0.1072 |
• For this special example, $\hat{\beta} = \log(n_{12}/n_{21}) = \log(132/107) = 0.21$
  with SE=$\sqrt{1/n_{12} + 1/n_{21}} = \sqrt{1/132 + 1/107} = 0.13$. Identical
  results to those from conditional logistic regression.

• $\hat{\sigma^2} = 8.11$, $\hat{\sigma} = 2.45 \Rightarrow$ A lot of between-subject variation.

• In general, the results from a GLMM will be different from those from
  a conditional logistic regression. There are several differences:
  1. GLMM allows making inference for the covariates that are fixed at
     subject level, while conditional logistic regression cannot.
  2. GLMM allows us to investigate the random effects variation among
     individuals.
  3. GLMM will be more efficient if the model is correct.
  4. However, we have to assume a distribution (usually normal) for the
     random effects.
I.2 GLMMs for binary repeated responses on similar items

- Example: Table 10.4 on legalization abortion in 3 situations

<table>
<thead>
<tr>
<th>Gender</th>
<th>(1,1,1)</th>
<th>(1,1,2)</th>
<th>(2,1,1)</th>
<th>(2,1,2)</th>
<th>(1,2,1)</th>
<th>(1,2,2)</th>
<th>(2,2,1)</th>
<th>(2,2,2)</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>

Source: Data from 1994 General Social Survey. Items 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.
Let \( y_{it} = 1/0 \) be the response (1=yes, 0=no) for subject \( i \) on item \( t (t = 1, 2, 3) \) and consider

\[
\text{logit}\{P[Y_{it} = 1|u_i]\} = u_i + \beta_t + \gamma x_i, \quad t = 1, 2, 3,
\]

where \( x_i = 1/0 \) for females/males, \( u_i \sim \text{N}(0, \sigma^2) \), \( \beta_t \)'s characterizes the response difference on items, \( \gamma \) characterizes the gender effect, \( \sigma^2 \) characterizes the between-subject variation after adjusting for gender effect and the item difference.

Note: We can use conditional logistic approach to fit the above model. But we will not be able to assess gender effect.
• SAS program and output:

```sas
data table10_4;
  input gender$ y1-y8;
  female=(gender="Female");
cards;
  Male  342  26  6  21  11  32  19  356
  Female  440  25  14  18  14  47  22  457;

  title "Recover individual data";
  data table10_4; set table10_4;
  array temp [8] y1-y8;
  retain id;
  if _n_=1 then id=0;
  do k=1 to 8;
    do i=1 to temp(k);
      id = id + 1;
      do item=1 to 3;
        if k=1 then y = 1;
        if k=2 then y = (item ne 3);
        if k=3 then y = (item ne 1);
        if k=4 then y = (item = 2);
        if k=5 then y = (item ne 2);
        if k=6 then y = (item = 1);
        if k=7 then y = (item = 3);
        if k=8 then y = 0;
        item1 = (item=1);  item2 = (item=2);  item3 = (item=3);
        output;
      end;
    end;
  end;
run;
```

title "Use GLMM for opinion on abortion: dummies for items 1, 2";
proc glimmix method=quad(qpoints=19);
class id;
model y = item1 item2 female / dist=bin link=logit s;
random int / subject=id type=vc;
run;

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>id</td>
<td>77.4375</td>
<td>8.0860</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect  | Estimate | Standard Error | DF  | t Value | Pr > |t| |
|---------|----------|----------------|-----|---------|-------|
| Intercept | -0.6108  | 0.3757         | 1848| -1.63   | 0.1042|
| item1    | 0.8222   | 0.1585         | 3698| 5.19    | <.0001|
| item2    | 0.2878   | 0.1554         | 3698| 1.85    | 0.0641|
| female   | 0.01316  | 0.4868         | 3698| 0.03    | 0.9784|

- $\hat{\sigma}^2 = 77.44$, $\hat{\beta}_1 - \hat{\beta}_3 = 0.82 (SE = 0.16)$, $\hat{\beta}_2 - \hat{\beta}_3 = 0.29 (SE = 0.16)$, $\hat{\gamma} = 0.013 (SE = 0.49)$.

- The gender effect is not significant. Drop it from the model. The resulting model is called an item response model - the Rasch model.
title "Use GLMM for opinion on abortion: dummies for items 1, 3";
proc glimmix method=quad(qpoints=19);
  class id;
  model y = item1 item3 female / dist=bin link=logit s;
  random int / subject=id type=vc;
run;

************************************************************************

Solutions for Fixed Effects

| Effect    | Estimate | Standard Error | DF   | t Value | Pr > |t| |
|-----------|----------|----------------|------|---------|------|---|
| Intercept | -0.3224  | 0.3754         | 1848 | -0.86   | 0.3905 |
| item1     | 0.5344   | 0.1558         | 3698 | 3.43    | 0.0006 |
| item3     | -0.2878  | 0.1554         | 3698 | -1.85   | 0.0641 |
| female    | 0.01258  | 0.4868         | 3698 | 0.03    | 0.9794 |

- $\hat{\beta}_1 - \hat{\beta}_2 = 0.53 (SE = 0.16)$.
- There is no gender effect on the response.
- There is an ordering of responding “yes” to items 1, 2, 3. For example, the odds of an individual saying “yes” for abortion at situation 1 is $e^{0.53} = 1.7$ times the odds of the same individual saying “yes” for abortion at situation 2.
- There is a lot of between-subject variation ($\hat{\sigma}^2 = 77.44$, $\hat{\sigma} = 8.8$).
• Note that we can also use GEE to fit a marginal model:

\[
\logit\{P[Y_{it} = 1]\} = \beta_t + \gamma x_i, \quad t = 1, 2, 3.
\]

title "Using GEE for abortion data";
proc genmod descending;
    class id;
    model y = item1 item2 female / dist=bin link=logit;
    repeated subject=id / type=exch corrw;
run;

************************************************************************

Exchangeable Working
Correlation

Correlation 0.8173308153

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.1253</td>
<td>0.0676</td>
<td>-0.2578 0.0071</td>
<td>-1.85</td>
<td>0.0637</td>
<td></td>
</tr>
<tr>
<td>item1</td>
<td>0.1493</td>
<td>0.0297</td>
<td>0.0911 0.2076</td>
<td>5.02</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>item2</td>
<td>0.0520</td>
<td>0.0270</td>
<td>-0.0010 0.1050</td>
<td>1.92</td>
<td>0.0544</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>0.0034</td>
<td>0.0878</td>
<td>-0.1687 0.1756</td>
<td>0.04</td>
<td>0.9688</td>
<td></td>
</tr>
</tbody>
</table>
proc genmod descending;
  class id;
  model y = item1 item3 female / dist=bin link=logit;
  repeated subject=id / type=exch corrw;
run;

*************************************************************************

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.0733</td>
<td>0.0676</td>
<td>-0.2058</td>
<td>-1.08</td>
<td>0.2780</td>
<td></td>
</tr>
<tr>
<td>item1</td>
<td>0.0973</td>
<td>0.0275</td>
<td>0.0434</td>
<td>3.54</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>item3</td>
<td>-0.0520</td>
<td>0.0270</td>
<td>-0.1050</td>
<td>-1.92</td>
<td>0.0544</td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>0.0034</td>
<td>0.0878</td>
<td>-0.1687</td>
<td>0.04</td>
<td>0.9688</td>
<td></td>
</tr>
</tbody>
</table>

- Because of very large $\hat{\sigma}^2$, the parameters $\beta_t$’s and $\gamma$ from this model will be much smaller than those in the mixed model. For example, $\hat{\beta}_1 - \hat{\beta}_2 = 0.1 (SE = 0.028)$. 

Slide 494
I.3 Small-area estimation for binomial probabilities

- Suppose $Y_i \sim \text{Bin}(n_i, \pi_i)$, $i = 1, 2, ..., m$. The best estimate for $\pi_i$ is its sample proportion $p_i = y_i/n_i$.

- When $n_i$’s are small, the sample proportion $p_i$ as an estimate of $\pi_i$ is not very good, e.g. $p_i$ has a large variation.

- We could assume $\pi_i$ is random and satisfies the model:

$$\text{logit}(\pi_i) = \alpha + u_i,$$

where $u_i \sim N(0, \sigma^2)$.

- After we fit this GLMM, we can get the estimates $\hat{\alpha}$ and $\hat{u}_i$, and then get the new estimate of $\pi_i$:

$$\hat{\pi}_i = \frac{e^{\hat{\alpha}+\hat{u}_i}}{1 + e^{\hat{\alpha}+\hat{u}_i}} = \text{logit}^{-1}(\hat{\alpha} + \hat{u}_i),$$

which can be obtained using “output out=randeff pred(ilink)=pihat;” in proc glimmix.
Example: estimating basketball free throw success (Table 10.2)

Table 10.2. Estimates of Probability of Centers Making a Free Throw, Based on Data from First Week of 2005–2006 NBA Season

<table>
<thead>
<tr>
<th>Player</th>
<th>$n_i$</th>
<th>$p_i$</th>
<th>$\hat{\pi}_i$</th>
<th>Player</th>
<th>$n_i$</th>
<th>$p_i$</th>
<th>$\hat{\pi}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yao</td>
<td>13</td>
<td>0.769</td>
<td>0.730</td>
<td>Curry</td>
<td>11</td>
<td>0.545</td>
<td>0.663</td>
</tr>
<tr>
<td>Frye</td>
<td>10</td>
<td>0.900</td>
<td>0.761</td>
<td>Miller</td>
<td>10</td>
<td>0.900</td>
<td>0.761</td>
</tr>
<tr>
<td>Camby</td>
<td>15</td>
<td>0.667</td>
<td>0.696</td>
<td>Haywood</td>
<td>8</td>
<td>0.500</td>
<td>0.663</td>
</tr>
<tr>
<td>Okur</td>
<td>14</td>
<td>0.643</td>
<td>0.689</td>
<td>Olowokandi</td>
<td>9</td>
<td>0.889</td>
<td>0.754</td>
</tr>
<tr>
<td>Blount</td>
<td>6</td>
<td>0.667</td>
<td>0.704</td>
<td>Mourning</td>
<td>9</td>
<td>0.778</td>
<td>0.728</td>
</tr>
<tr>
<td>Mihm</td>
<td>10</td>
<td>0.900</td>
<td>0.761</td>
<td>Wallace</td>
<td>8</td>
<td>0.625</td>
<td>0.692</td>
</tr>
<tr>
<td>Ilgauskas</td>
<td>10</td>
<td>0.600</td>
<td>0.682</td>
<td>Ostertag</td>
<td>6</td>
<td>0.167</td>
<td>0.608</td>
</tr>
<tr>
<td>Brown</td>
<td>4</td>
<td>1.000</td>
<td>0.748</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $p_i =$ sample, $\hat{\pi}_i =$ estimate using random effects model.

Source: nba.com.
• SAS program and part of output:

```sas
data table10_4;
  input player$ n p;
  y = round(n*p);
  cards;
  Yao     13  0.769
  Curry   11  0.545
  Frye    10  0.900
  Miller  10  0.900
  Camby   15  0.667
  Haywood 8  0.500
  Okur    14  0.643
  Olowokandi 9  0.889
  Blount  6  0.667
  Mourning 9  0.778
  Mihm    10  0.900
  Wallace 8  0.625
  Ilgauskas 10  0.600
  Ostertag 6  0.167
  Brown   4  1.000
;

proc glimmix method=quad(qpoints=19);
  class player;
  model y/n = / dist=bin link=logit s;
  random int / subject=player type=vc s;
  output out=randeff pred(ilink)=pihat;
run;
```

### Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>player</td>
<td>0.1779</td>
<td>0.3312</td>
</tr>
</tbody>
</table>

### Solutions for Fixed Effects

| Effect    | Estimate | Standard Error | DF | t Value | Pr > |t| |
|-----------|----------|----------------|----|---------|-------|
| Intercept | 0.9076   | 0.2244         | 14 | 4.04    | 0.0012|

### Solution for Random Effects

| Effect            | Subject     | Estimate   | Std Err Pred | DF | t Value | Pr > |t| |
|-------------------|-------------|------------|--------------|----|---------|-------|
| Intercept         | player Blount | -0.04008   | 0.3899       | 0  | -0.10   | .     |
|                    | player Brown | 0.1794     | 0.4906       | 0  | 0.37    | .     |
|                    | player Camby | -0.07862   | 0.3640       | 0  | -0.22   | .     |
|                    | player Curry | -0.2303    | 0.4762       | 0  | -0.48   | .     |
|                    | player Frye  | 0.2481     | 0.5003       | 0  | 0.50    | .     |
|                    | player Haywood | -0.2317   | 0.5031       | 0  | -0.46   | .     |
|                    | player Ilgauska | -0.1455  | 0.4196       | 0  | -0.35   | .     |
|                    | player Mihm   | 0.2481     | 0.5003       | 0  | 0.50    | .     |
|                    | player Miller | 0.2481     | 0.5003       | 0  | 0.50    | .     |
|                    | player Mourning | 0.07902   | 0.3843       | 0  | 0.21    | .     |
|                    | player Okur   | -0.1139    | 0.3823       | 0  | -0.30   | .     |
|                    | player Olowokan | 0.2151    | 0.4775       | 0  | 0.45    | .     |
|                    | player Ostertag | -0.4705  | 0.8039       | 0  | -0.59   | .     |
|                    | player Wallace | -0.09598  | 0.4016       | 0  | -0.24   | .     |
|                    | player Yao    | 0.08956    | 0.3696       | 0  | 0.24    | .     |
• We see that compared to the sample proportion $p_i$’s, $\hat{\pi}_i$’s are closer to overall sample proportion $101/143 = 0.706$. That is, $p_i$’s that are larger than 0.706 are shrunk and $p_i$’s that are smaller than 0.706 are inflated.
The estimate $\hat{\alpha}$ and $\hat{\sigma}^2 = 0.18$ allow us to make a probability statement for a randomly selected player (from the population to which the studied players belong):

\[
u_i \sim \text{N}(0, \sigma^2)
\]

\[
P[-1.96\sigma \leq u_i \leq 1.96\sigma] = 0.95
\]

\[
P[\alpha - 1.96\sigma \leq \alpha + u_i \leq \alpha + 1.96\sigma] = 0.95
\]

\[
P[\logit^{-1}(\alpha - 1.96\sigma) \leq \logit^{-1}(\alpha + u_i) \leq \logit^{-1}(\alpha + 1.96\sigma)] = 0.95
\]

\[
P[\logit^{-1}(\alpha - 1.96\sigma) \leq \pi_i \leq \logit^{-1}(\alpha + 1.96\sigma)] = 0.95
\]

\[
\logit^{-1}(\alpha - 1.96\sigma) = \frac{e^{\alpha-1.96\sigma}}{1 + e^{\alpha-1.96\sigma}} = \frac{e^{0.9076-1.96 \times 0.424}}{1 + e^{0.9076-1.96 \times 0.424}} = 0.52
\]

\[
\logit^{-1}(\alpha + 1.96\sigma) = \frac{e^{\alpha+1.96\sigma}}{1 + e^{\alpha+1.96\sigma}} = \frac{e^{0.9076+1.96 \times 0.424}}{1 + e^{0.9076+1.96 \times 0.424}} = 0.85
\]

\[
P[0.52 \leq \pi_i \leq 0.85] = 0.95,
\]

that is, the prob that this player’s success prob is between 0.52 and 0.85 is 0.95.
I.4 GLMM for clustered binomial data

- Example (Table 9.4): Low-iron rat study where iron-deficient female rats were assigned to 4 groups:
  Group 1: untreated (control)
  Group 2: injection of iron supplement on days 7, 10
  Group 3: injection on days 0, 7
  Group 4: injection weekly

- Data: $y_i$ = # of dead baby rats out of $n_i$ baby rats in litter $i = 1, 2, \ldots, m$.
  For the $i$th litter, the $n_i$ binary data are correlated since they all share the same death probability $\pi_i$.

- Consider logit model for $\pi_{ig}$:

$$\text{logit}(\pi_i) = u_i + \alpha + \beta_2gp_2 + \beta_3gp_3 + \beta_4gp_4, \quad u_i \sim \text{N}(0, \sigma^2),$$

where $gp_1, gp_2, gp_3, gp_4$ are dummy variables for groups 1, 2, 3, 4.

We may use $(1 + 0.346\sigma^2)^{-1/2}\beta_j$ to compare group $j$ to group 1.
<table>
<thead>
<tr>
<th>Group 1: untreated (low iron)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1) (11, 4) (12, 9) (4, 4) (10, 10) (11, 9) (9, 9) (11, 11) (10, 10) (10, 7) (12, 12)</td>
</tr>
<tr>
<td>(10, 9) (8, 8) (11, 9) (6, 4) (9, 7) (14, 14) (12, 7) (11, 9) (13, 8) (14, 5) (10, 10)</td>
</tr>
<tr>
<td>(12, 10) (13, 8) (10, 10) (14, 3) (13, 13) (4, 3) (8, 8) (13, 5) (12, 12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2: injections days 7 and 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 1) (3, 1) (13, 1) (12, 0) (14, 4) (9, 2) (13, 2) (16, 1) (11, 0) (4, 0) (1, 0) (12, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 3: injections days 0 and 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>(8, 0) (11, 1) (14, 0) (14, 1) (11, 0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 4: injections weekly</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 0) (13, 0) (9, 2) (17, 2) (15, 0) (2, 0) (14, 1) (8, 0) (6, 0) (17, 0)</td>
</tr>
</tbody>
</table>

data rat;
  input litter group n y;
  gp1 = (group=1); gp2 = (group=2); gp3 = (group=3); gp4 = (group=4);
  datalines;
  1  1  10  1
  2  1  11  4
  3  1  12  9
  4  1  1  4
  5  1  10 10
  6  1  11  9
  7  1  9  9
  8  1  11 11
  9  1  10 10
 10  1  10  7
 11  1  12 12
 12  1  10  9
 13  1  8  8
 14  1  11  9
 15  1  6  4
 16  1  9  7
 17  1  14 14
 18  1  12  7
 19  1  11  9
 20  1  13  8
 21  1  14  5
 22  1  1010
 23  1  12 10
 24  1  13  8
 25  1  1010
 26  1  14  3
 27  1  13 13
 28  1  4  3
 29  1  8  8
 30  1  13  5
 31  1  12 12
 32  2  10  1
 33  2  3  1
 34  2  13  1
 35  2  12  0
title "Glimmix to rat’s data";
Proc Glimmix method=quad data=rat;
  class litter group;
  model y/n = gp2 gp3 gp4 / dist=bin link=logit s;
  random int / subject=litter type=vc;
run;

*******************************************************************************

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>litter</td>
<td>2.3582</td>
<td>0.8873</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect | Estimate | Standard Error | DF | t Value | Pr > |t| |
|--------|----------|----------------|----|---------|------|
| Intercept | 1.8040   | 0.3630         | 54 | 4.97    | <.0001 |
| gp2     | -4.5178  | 0.7374         | 0  | -6.13   | .     |
| gp3     | -5.8576  | 1.1904         | 0  | -4.92   | .     |
| gp4     | -5.5975  | 0.9201         | 0  | -6.08   | .     |

- Ignore the DF=0 and compare t Value to N(0,1).
- \((1 + 0.346\hat{\sigma}^2)^{-1/2}\hat{\beta}_2 = (1 + 0.346 \times 2.3582)^{-1/2}(-4.5178) = -3.35, \ e^{-3.35} = 0.035, \Rightarrow \) the odds of death for group 2 is only about 0.035 times the odds of death of group 1. See slide 455 for GEE analysis.
II GLMM for Longitudinal Count Data

- Use seizure data as an example. Assume seizure counts

\[ y_{ij} | b_i \sim \text{Overdispersed } - \text{Poisson}(\mu_{ij}^b), \]

where

\[ \mu_{ij}^b = \mathbb{E}(y_{ij} | b_i) = t_{ij} \lambda_{ij}^b, \quad \text{var}(y_{ij} | b_i) = \phi \mu_{ij}^b, \]

\( \lambda_{ij}^b \) is the rate to have a seizure for subject \( i \). Consider model

\[ \log(\lambda_{ij}^b) = \beta_0 + \beta_1 I(j > 1) + \beta_2 \text{trt}_i I(j > 1) + b_i \]

\[ \log(\mu_{ij}^b) = \log(t_{ij}) + \beta_0 + \beta_1 I(j > 1) + \beta_2 \text{trt}_i I(j > 1) + b_i, \]

where \( b_i \sim N(0, \sigma^2) \) is a random intercept describing the between-subject variation.
• Interpretation of $\beta$’s:

<table>
<thead>
<tr>
<th>Group</th>
<th>log($\lambda^b$) for random subject $i$</th>
<th>Before randomization</th>
<th>After randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (trt=0)</td>
<td>$\beta_0 + b_i$</td>
<td>$\beta_0 + \beta_1 + b_i$</td>
<td></td>
</tr>
<tr>
<td>Treatment (trt=1)</td>
<td>$\beta_0 + b_i$</td>
<td>$\beta_0 + \beta_1 + \beta_2 + b_i$</td>
<td></td>
</tr>
</tbody>
</table>

$\beta_1$: difference in log of seizure rates comparing after randomization and before randomization for a random subject in the control group (time & placebo effect).

$\beta_2$: difference in log of seizure rates for a treated subject compared to if he/she received a placebo (treatment effect).

• It can be shown that

$$\lambda_{ij} = \frac{\mu_{ij}}{t_{ij}} = \frac{E(\mu_{ij}^b)}{t_{ij}} = e^{\beta_0 + \sigma^2/2 + \beta_1 I(j>1) + \beta_2 \text{trt}_i I(j>1)},$$

so $\beta_1$ and $\beta_2$ also have population average interpretation.
- SAS program and output:

```sas
/*------------------------------------------------------*/
/* */
/* Proc Glimmix to fit random intercept model to the */
/* epileptic seizure count data */
/* */
/* */
/*------------------------------------------------------*/

data seizure;
   infile "seize.dat";
   input id seize visit trt age;
   nobs=_n_;    
   interval = 2;
   if visit=0 then interval=8;
   logtime = log(interval);
   assign = (visit>0);
   agn_trt = assign*trt;
run;

title "Random intercept model for seizure data with conditional overdispersion";
proc glimmix data=seizure;
   class id;
   model seize = assign agn_trt / dist=poisson link=log offset=logtime s;
   random int / subject=id type=vc;
   random _residual_; *for conditional overdispersion;
run;
```
Random intercept model for seizure data with conditional overdispersion

The GLIMMIX Procedure

Fit Statistics

-2 Res Log Pseudo-Likelihood 675.86
Generalized Chi-Square 822.08
Gener. Chi-Square / DF 2.82

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept id</td>
<td></td>
<td>0.5704</td>
<td>0.1169</td>
</tr>
<tr>
<td>Residual (VC)</td>
<td></td>
<td>2.8154</td>
<td>0.2591</td>
</tr>
</tbody>
</table>

Solutions for Fixed Effects

| Effect    | Estimate | Standard Error | DF | t Value | Pr > |t| |
|-----------|----------|----------------|----|---------|-------|
| Intercept | 1.0655   | 0.1079         | 58 | 9.88    | <.0001|
| assign    | 0.1122   | 0.07723        | 234| 1.45    | 0.1477|
| agn_trt   | -0.1063  | 0.1054         | 234| -1.01   | 0.3144|
• **Remark:** There is considerable amount of over-dispersion for $y_{ij|b_i}$. It is estimated that

\[
\text{var}(y_{ij|b_i}) = 2.82\text{E}(y_{ij|b_i}).
\]

• There is considerable between-patient variance in log-seizure rate. That variation $\sigma^2$ of $b_i$ is estimated to be 0.57.

• The regression coefficient estimates (except the intercept) have population-average interpretation, and they are almost the same as those from the GEE model.

For example, $\hat{\beta}_2 = -0.1063$ with SE = 0.1054. Then if a subject switches from control to treatment, the rate of having seizure will decrease by 10\% (since $e^{-0.1063} = 0.9$). The same rate reduction can also be used to compare treatment and control groups (i.e., population interpretation).
III GLMM for Ordinal Longitudinal Data

- Consider the cumulative logit mixed model for the insomnia data

\[
\text{logit}\{P[Y_{ij} \leq k | b_i]\} = \alpha_k + b_i + \beta_1 I(j = 2) + \beta_2 \text{trt}_i + \beta_3 I(j = 2) \times \text{trt}_i,
\]

where \(b_i \sim \mathcal{N}(0, \sigma^2)\) models the between-subject variation in the subject-specific cumulative logits.

- Interpretation of \(\beta_1, \beta_2, \beta_3\):
  1. \(\beta_1\): Effect of time + placebo
  2. \(\beta_2\): Group difference at baseline (can be set to 0 by randomization)
  3. \(\beta_3\): Treatment effect after taking into account the time and placebo effects.

- The interpretation of \(\beta_1\) and \(\beta_3\) are all in subject level. Even though we cannot directly use \(\beta_2\) to compare those 2 groups at baseline, \(\beta_2 = 0 \iff \text{no group difference at baseline}\).
• SAS program and output:

```
title "Cumulative logit mixed model for insomnia longitudinal data";
proc Glimmix method=quad data=table9_6;
  class id;
  model y = time trt time*trt / s dist=multinomial link=clogit;
  random int / subject=id type=vc;
run;
```

***********************************************************************

Cumulative logit mixed model for insomnia longitudinal data

Response Profile

<table>
<thead>
<tr>
<th>Ordered Value</th>
<th>y</th>
<th>Total Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>118</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>129</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>134</td>
</tr>
</tbody>
</table>

The GLIMMIX procedure is modeling the probabilities of levels of y having lower Ordered Values in the Response Profile table.

Convergence criterion (GCONV=1E-8) satisfied.

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>id</td>
<td>3.6162</td>
<td>0.8768</td>
</tr>
</tbody>
</table>
### Solutions for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>y</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept 1</td>
<td>1</td>
<td>-3.4874</td>
<td>0.3584</td>
<td>237</td>
<td>-9.73</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 2</td>
<td>2</td>
<td>-1.4836</td>
<td>0.2901</td>
<td>237</td>
<td>-5.11</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept 3</td>
<td>3</td>
<td>0.5610</td>
<td>0.2699</td>
<td>237</td>
<td>2.08</td>
<td>0.0387</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>4</td>
<td>1.6010</td>
<td>0.2834</td>
<td>235</td>
<td>5.65</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>trt</td>
<td>5</td>
<td>0.05776</td>
<td>0.3659</td>
<td>235</td>
<td>0.16</td>
<td>0.8747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time*trt</td>
<td>6</td>
<td>1.0801</td>
<td>0.3803</td>
<td>235</td>
<td>2.84</td>
<td>0.0049</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- $\hat{\beta}_1 = 1.6$, $e^{\hat{\beta}_1} = 5$: for a placebo patient, his/her odds of having shorter *time to falling asleep* 2 weeks later is 5 times his/her odds at baseline.
- P-value for $H_0 : \beta_2 = 0$ is 0.87, no group difference at baseline.
- $e^{\hat{\beta}_1 + \hat{\beta}_3} = 15$: for a treated patient, his/her odds of having shorter *time to falling asleep* 2 weeks later is 15 times the odds at baseline.
• **Note 1**: Here the interpretation is on subject level. The interpretation presented on slide 467 is on the population level.

• $\hat{\sigma}^2 = 3.6162$ – variability of subject-specific cumulative logits in the population.

• **Note 2**: We can also get approximate population level interpretation:

  1. $\hat{\beta}_1^* \approx (1+0.346 \times \hat{\sigma}^2)^{-1/2}\hat{\beta}_1 = (1+0.346 \times 3.6162)^{-1/2} \times 1.6 = 1.07$, very close to the estimate of $\beta_1$ (1.04) on slides 467.

  2. $\hat{\beta}_1^* + \hat{\beta}_3^* \approx (1 + 0.346 \times 3.6162)^{-1/2} \times 2.68 = 1.79$, very close to the estimate of $\beta_1 + \beta_3$ (1.75) from slide 467.