ST3241 Categorical Data Analysis I
Loglinear Models

2 × 2 Models For Contingency Tables
Two-way Tables

- Consider an $I \times J$ contingency table that crossclassifies a sample of $n$ subjects on two categorical responses.
- Let $Y_{ij}$ be the observed cell frequency and $\mu_{ij}$ be the expected cell frequency of the $(i, j)$-th cell.
- Then we assume that the cell counts $Y_{ij}$ are independent having Poisson($\mu_{ij}$) distribution.
- Note that, if $\pi_{ij}$ is the cell probability, then $\mu_{ij} = n\pi_{ij}$. 
**Independence Model**

- Under statistical independence of the row and column classifications, \( \pi_{ij} = \pi_i + \pi_j \) and hence \( \mu_{ij} = n \pi_i + \pi_j \).

- Denote the row variable by \( X \) and the column variable by \( Y \).

- The formula expressing independence is multiplicative. Thus, \( \log \mu_{ij} \) is additive

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

for a row effect \( \lambda_i^X \) and a column effect \( \lambda_j^Y \).

- This is the loglinear model of independence.

- The null hypothesis of independence between two categorical variables is simply the hypothesis that this model holds.
**Example: Belief in Afterlife**

<table>
<thead>
<tr>
<th>Observed Frequency</th>
<th>Fitted Value</th>
<th>Value</th>
<th>Log Fitted Value</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>435 147</td>
<td>432.10</td>
<td>149.90</td>
<td>6.069</td>
<td>5.010</td>
</tr>
<tr>
<td>375 134</td>
<td>377.90</td>
<td>131.10</td>
<td>5.935</td>
<td>4.876</td>
</tr>
</tbody>
</table>
### Example: Belief in Afterlife

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>4.876</td>
<td>6.069</td>
<td>5.472</td>
</tr>
<tr>
<td>$\lambda_1^X$</td>
<td>0.134</td>
<td>0</td>
<td>0.0067</td>
</tr>
<tr>
<td>$\lambda_2^X$</td>
<td>0</td>
<td>-0.134</td>
<td>-0.067</td>
</tr>
<tr>
<td>$\lambda_1^Y$</td>
<td>1.059</td>
<td>0</td>
<td>0.529</td>
</tr>
<tr>
<td>$\lambda_2^Y$</td>
<td>0</td>
<td>-1.059</td>
<td>-0.529</td>
</tr>
</tbody>
</table>
Some SAS Codes

data after;
  input female $ belief $ count;
datalines;
  Female Yes 435
  Female No 147
  Male Yes 375
  Male No 134
;
run;
SAS Codes

proc genmod data=after order=data;
   class belief female;
   model count= female belief/ dist=poisson;
   output out=temp p=predict;
run;
proc print data=temp;
   var female belief count predict;
run;
## Partial Output

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>0.1620</td>
<td>0.1620</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>1</td>
<td>0.1620</td>
<td>0.1620</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>1</td>
<td>0.1621</td>
<td>0.1621</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>1</td>
<td>0.1621</td>
<td>0.1621</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>5164.1959</td>
<td></td>
</tr>
</tbody>
</table>

Algorithm converged.
## Partial Output

### Analysis Of Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
<th>Square</th>
<th>Pr&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.8760</td>
<td>0.0679</td>
<td>4.7429 5.0090</td>
<td>5160.87</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>female Female</td>
<td>1</td>
<td>0.1340</td>
<td>0.0607</td>
<td>0.0151 0.2530</td>
<td>4.88</td>
<td>0.0272</td>
</tr>
<tr>
<td>female Male</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
<td>4.88</td>
<td>0.0272</td>
</tr>
<tr>
<td>belief Yes</td>
<td>1</td>
<td>1.0587</td>
<td>0.0692</td>
<td>0.9230 1.1944</td>
<td>233.83</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>belief No</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000 0.0000</td>
<td>4.88</td>
<td>0.0272</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td>4.88</td>
<td>0.0272</td>
</tr>
</tbody>
</table>
### Partial Output

<table>
<thead>
<tr>
<th>Obs</th>
<th>female</th>
<th>belief</th>
<th>count</th>
<th>predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>Yes</td>
<td>435</td>
<td>432.099</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>No</td>
<td>147</td>
<td>149.901</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>Yes</td>
<td>375</td>
<td>377.901</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>No</td>
<td>134</td>
<td>131.099</td>
</tr>
</tbody>
</table>
Interpretations of Parameters

• For $I \times J$ tables, loglinear models treat the $N = IJ$ cell counts as $N$ independent observations of a Poisson random component.

• The data are the $N$ cell counts rather than the individual classifications of the $n$ subjects.

• The model does not distinguish between response and explanatory variables.

• Differences between two parameters for a given variable relate to the log odds of making response, relative to the other, on that variable.

• If the response is binary, one can use logit models directly but loglinear models are useful for modeling relationships among two or more categorical response variables.
In $I \times 2$ Table

- Response $Y$ has only 2 levels.
- In row $i$, the logit for the probability $\pi$ that $Y = 1$ is:
  \[
  \log\left( \frac{\pi_i}{1 - \pi_i} \right) = \log\left( \frac{\mu_{i1}}{\mu_{i2}} \right) = \log \mu_{i1} - \log \mu_{i2}
  \]
  \[
  = (\lambda + \lambda_i^X + \lambda_1^Y) - (\lambda + \lambda_i^X + \lambda_2^Y)
  \]
  \[
  = \lambda_1^Y - \lambda_2^Y
  \]
- logit for $Y$ does not depend on the levels of $X$.  

**Saturated Model**

- A more complex model

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

- Here \( \lambda_{ij}^{XY} \) are the association terms reflecting deviation from the independence or interaction effect between \( i \)-th category of \( X \) and \( j \)-th category of \( Y \).

- This is the saturated loglinear model in a two-way table.
**Interpretation of Interaction**

- There is a direct relationship between log odds ratios and \( \{ \lambda_{ij}^{XY} \} \) association parameters.

- In a 2 x 2 table,

\[
\log \theta = \log \left( \frac{\mu_{11} \mu_{22}}{\mu_{12} \mu_{21}} \right) \\
= \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21} \\
= (\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}) + (\lambda + \lambda_2^X + \lambda_2^Y + \lambda_{22}^{XY}) \\
- (\lambda + \lambda_1^X + \lambda_2^Y + \lambda_{12}^{XY}) - (\lambda + \lambda_2^X + \lambda_1^Y + \lambda_{21}^{XY}) \\
= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}
\]
**Example: Belief in Afterlife**

<table>
<thead>
<tr>
<th>Association</th>
<th>Parameter</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{11}^{XY}$</td>
<td>0.056</td>
<td>0.014</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{12}^{XY}$</td>
<td>0.0</td>
<td>-0.014</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{21}^{XY}$</td>
<td>0.0</td>
<td>-0.014</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>$\lambda_{22}^{XY}$</td>
<td>0.0</td>
<td>0.014</td>
<td>0.056</td>
<td></td>
</tr>
</tbody>
</table>
Example: Belief in Afterlife

- From the data, the odds ratio is 
  \[(435 \times 134)/(147 \times 375) = 1.057.\]

- Therefore, the log odds ratio = 0.056

- From the association parameters

  \[\hat{\lambda}^{XY}_{11} + \hat{\lambda}^{XY}_{22} - \hat{\lambda}^{XY}_{12} - \hat{\lambda}^{XY}_{21} = 0.056\]
Notes

• In $I \times J$ tables, only $(I - 1)(J - 1)$ parameters are non-redundant.

• These *interaction* parameters in the saturated model are coefficients of cross products of $(I - 1)$ dummy variables for $X$ with $(J - 1)$ dummy variables for $Y$.

• Tests of independence analyze whether these $(I - 1)(J - 1)$ parameters equal 0, so they have residual $d.f. = (I - 1)(J - 1)$. 
Notes

• The saturated model has as many parameters as it has Poisson observations.

• Thus, it gives a perfect fit.

• This model is also a hierarchical model as it includes all lower order terms composed from variables contained in a higher order term in the model.
Three-way Tables

• Denote the cell expected frequencies in the contingency table by \( \{ \mu_{ijk} \} \).

• Single factor terms in loglinear models for \( \{ \mu_{ijk} \} \) represent marginal distributions.

• E.g. including \( \hat{\lambda}_i^X \) in the model forces the fitted values to have the same totals at the various levels of \( X \) as do the observed data.

• Partial associations between variables correspond to two factor terms.
Partial Association Models

- Consider the loglinear model

\[
\log \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XZ}_{ik} + \lambda^{YZ}_{jk}
\]

- Since it contains an $X - Z$ two factor term, it permits association between $X$ and $Z$ controlling for $Y$.

- This model also permits a $Y - Z$ association, controlling for $X$.

- It does not contain a two-factor term for $X - Y$ association.

- It specifies conditional independence between $X$ and $Y$, controlling for $Z$. 
Partial Association Models

- We symbolize the previous model as $(XZ, YZ)$.
- For $2 \times 2 \times k$ tables, this model corresponds to the hypothesis tested using the Cochran–Mantel–Haenszel statistic.
- The model that contains only single factor terms, denoted by $(X, Y, Z)$, is called the mutual independence model.
- It treats each pair of variables as independent.
Partial Association Models

• The model which permits all three pairs of variables to be conditionally dependent is

\[
\log \mu_{ijk} = \lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XY}_{ij} + \lambda^{XZ}_{ik} + \lambda^{YZ}_{jk}
\]

• For this model, the conditional odds ratios between any two variables are identical at each level of the third variable.

• We refer to this model as the homogeneous association model and symbolize it by \((XY, YZ, XZ)\).
Saturated Model

• The model:

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

• The three factor term pertains to three-factor interaction.

• Denoted by \((XYZ)\), this model permits the odds ratio between any two variables to vary across levels of the third variable.

• It provides a perfect fit in a three-way table.
Interpreting Model Parameters

- Interpretation of loglinear model parameters refer to their highest order terms.

- E.g. interpretations for the homogeneous association model use the two factor terms to describe associations.

- The two-factor parameters relate directly to conditional odds ratios.

- $X - Y$ conditional odds ratio $\theta_{XY(k)}$ in the homogeneous association model can be shown to be

$$
\log \theta_{XY(k)} = \log \left( \frac{\mu_{11k} \mu_{22k}}{\mu_{12k} \mu_{21k}} \right) = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}
$$

which does not depend on $k$. 
Example: Alcohol, Cigarette and Marijuana

Use For High School Seniors

<table>
<thead>
<tr>
<th>Alcohol Use</th>
<th>Cigarette Use</th>
<th>Marijuana Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>911</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>44</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>2</td>
</tr>
</tbody>
</table>
Example: SAS Codes

data drug;
   input a $ c $ m $ count;
datalines;
Yes Yes Yes 911
Yes Yes No 538
Yes No Yes 44
Yes No No 456
No Yes Yes 3
No Yes No 43
No No Yes 2
No No No 279
;
runc;
Example: SAS Codes

```sas
proc catmod data=drug order=data;
  weight count;
  model a*c*m =_response_/ 
    pred=freq;
  loglin a|m c|m;
run;
quit;
```
The CATMOD Procedure

Data Summary

Response a*c*m Response Levels 8
Weight Variable count Populations 1
Data Set DRUG Total Frequency 2276
Frequency Missing 0 Observations 8

Population Profiles
Sample Sample Size
+++++++++++++++++++++
1 2276
### Partial Output

**Maximum Likelihood Analysis**

Maximum likelihood computations converged.

**Maximum Likelihood 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>a</td>
<td>1</td>
<td>198.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>m</td>
<td>1</td>
<td>155.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>a*m</td>
<td>1</td>
<td>83.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>292.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c*m</td>
<td>1</td>
<td>401.17</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>2</td>
<td>187.75</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
## Partial Output

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Chi-Square</th>
<th>Pr&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.5949</td>
<td>0.1132</td>
<td>198.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>m</td>
<td>-1.4731</td>
<td>0.1182</td>
<td>155.20</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>a*m</td>
<td>1.0313</td>
<td>0.1132</td>
<td>83.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c</td>
<td>0.6885</td>
<td>0.0402</td>
<td>292.68</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>c*m</td>
<td>0.8061</td>
<td>0.0402</td>
<td>401.17</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
### Partial Output

**Maximum Likelihood Predicted Values for Frequencies**

<table>
<thead>
<tr>
<th>a</th>
<th>c</th>
<th>m</th>
<th>Frequency</th>
<th>Error Frequency</th>
<th>Error</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes Yes Yes</td>
<td>911</td>
<td>23.37434</td>
<td>909.2396</td>
<td>23.36195</td>
<td>1.760417</td>
<td></td>
</tr>
<tr>
<td>Yes Yes No</td>
<td>538</td>
<td>20.26889</td>
<td>438.8404</td>
<td>17.15391</td>
<td>99.15957</td>
<td></td>
</tr>
<tr>
<td>Yes No Yes</td>
<td>44</td>
<td>6.568819</td>
<td>45.76042</td>
<td>6.679323</td>
<td>-1.76042</td>
<td></td>
</tr>
<tr>
<td>Yes No No</td>
<td>456</td>
<td>19.09554</td>
<td>555.1596</td>
<td>18.96775</td>
<td>-99.1596</td>
<td></td>
</tr>
<tr>
<td>No Yes Yes</td>
<td>3</td>
<td>1.730909</td>
<td>4.760418</td>
<td>2.126048</td>
<td>-1.76042</td>
<td></td>
</tr>
<tr>
<td>No Yes No</td>
<td>43</td>
<td>6.495199</td>
<td>142.1596</td>
<td>8.562114</td>
<td>-99.1596</td>
<td></td>
</tr>
<tr>
<td>No No Yes</td>
<td>2</td>
<td>1.413592</td>
<td>0.239583</td>
<td>0.112401</td>
<td>1.760417</td>
<td></td>
</tr>
<tr>
<td>No No No</td>
<td>279</td>
<td>15.64606</td>
<td>179.8404</td>
<td>10.27909</td>
<td>99.15957</td>
<td></td>
</tr>
</tbody>
</table>
### Fitted Values for Loglinear Models

<table>
<thead>
<tr>
<th>Alcohol</th>
<th>Cigarette</th>
<th>Marijuana</th>
<th>(A,C,M)</th>
<th>(AC,M)</th>
<th>(AM,CM)</th>
<th>(AC,AM,CM)</th>
<th>(ACM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>540.0</td>
<td>611.2</td>
<td>909.24</td>
<td>910.4</td>
<td>911</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>740.2</td>
<td>837.8</td>
<td>438.84</td>
<td>538.6</td>
<td>538</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
<td>282.1</td>
<td>210.9</td>
<td>45.76</td>
<td>44.6</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>386.7</td>
<td>289.1</td>
<td>555.16</td>
<td>455.4</td>
<td>456</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>90.6</td>
<td>19.4</td>
<td>4.76</td>
<td>3.6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>124.2</td>
<td>26.6</td>
<td>142.16</td>
<td>42.4</td>
<td>43</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td></td>
<td>47.3</td>
<td>118.5</td>
<td>0.24</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>64.9</td>
<td>162.5</td>
<td>179.84</td>
<td>279.6</td>
<td>279</td>
</tr>
</tbody>
</table>
Estimated Odds Ratios

<table>
<thead>
<tr>
<th>Model</th>
<th>Conditional Association</th>
<th>Marginal Association</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A-C</td>
<td>A-M</td>
</tr>
<tr>
<td>(A,C,M)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>(AC,M)</td>
<td>17.7</td>
<td>1.0</td>
</tr>
<tr>
<td>(AM,CM)</td>
<td>1.0</td>
<td>61.9</td>
</tr>
<tr>
<td>(AC,AM,CM)</td>
<td>7.8</td>
<td>19.8</td>
</tr>
<tr>
<td>(ACM) Level 1</td>
<td>13.8</td>
<td>24.3</td>
</tr>
<tr>
<td>(ACM) Level 2</td>
<td>7.7</td>
<td>13.5</td>
</tr>
</tbody>
</table>
Example

• The fit for model \((AC, AM, CM)\) is close to the observed data.

• The other models seem to fit poorly.

• The model \((AM, CM)\) implies conditional independence between alcohol use and cigarette use, controlling for marijuana use, and yields odds ratios of 1.0 for the \(A \rightarrow C\) conditional association.

• The entry 2.7 for the \(A \rightarrow C\) marginal association for this model is the fitted odds ratio for the marginal \(A \rightarrow C\) fitted table.

• Model \((AC, AM, CM)\) permits all pairwise associations but maintains homogeneous odds ratios between two variables at each level of the third variable.

• The estimated conditional odds ratios equal 1.0 for each pairwise term not appearing in a model.

• The estimates of conditional and marginal odds ratios are highly dependent on the model.
Fitting Loglinear Models

• Some loglinear models have explicit formulas for the fitted values \( \{\hat{\mu}_{ijk}\} \) in terms of \( \{n_{ijk}\} \).

• For example, the model \((XZ, YZ)\) of \(X - Y\) conditional independence has

\[
\hat{\mu}_{ijk} = \frac{n_{i+k}n_{+jk}}{n_{++k}}
\]

• Many loglinear models do not have any direct estimates.

• ML estimation then uses iterative methods.
Chi-Square Goodness-of-Fit Tests

• Consider the null hypothesis that the expected frequencies for a three-way table satisfy a given loglinear model.

• The LR and Pearson Chi-square statistics are:

\[
G^2 = 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} \log\left( \frac{n_{ijk}}{\hat{\mu}_{ijk}} \right),
\]

\[
\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}
\]

• The degrees of freedom equals the number of cell counts minus the number of non-redundant parameters in the model.

• The saturated model has \( d.f. = 0 \).
**Example: Drug Use Data**

<table>
<thead>
<tr>
<th>Model</th>
<th>$G^2$</th>
<th>$\chi^2$</th>
<th>df</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A,C,M)</td>
<td>1286.0</td>
<td>1411.4</td>
<td>4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(A,CM)</td>
<td>534.2</td>
<td>505.6</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(C,AM)</td>
<td>939.6</td>
<td>824.2</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(M,AC)</td>
<td>843.8</td>
<td>704.9</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,AM)</td>
<td>497.4</td>
<td>443.8</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,CM)</td>
<td>92.0</td>
<td>80.8</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AM,CM)</td>
<td>187.8</td>
<td>177.6</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,AM,CM)</td>
<td>0.4</td>
<td>0.4</td>
<td>1</td>
<td>0.54</td>
</tr>
<tr>
<td>(ACM)</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Residuals

- One can study the quality of the fit more closely by studying cell residuals.
- They may indicate why a particular model does not fit well or highlight cells that display lack of fit.
- We may use adjusted residuals or Pearson residuals.
- When the model holds, the adjusted residuals have approximately standard normal distribution.
- So, absolute values of adjusted residuals larger than 2 when there are few cells and larger than 3 when there are many cells indicate lack of fit.
Example: SAS Codes (AM,CM)

```sas
proc genmod data=drug;
   class a c m;
   model count= a c m a*m c*m / dist=p;
   output out=temp reslik=reslik p=pred;
run;

data temp;
   set temp;
   id = _n_;
   rename reslik=adjres1;
   rename pred=fitted1;
run;
```
Example: SAS Codes (AC,AM,CM)

```sas
proc genmod data=drug;
  class a c m;
  model count= a c m a*m c*m a*c/ dist=p;
  output out=temp1 reslik=reslik p=pred;
run;

data temp1;
  set temp1;
  id = _n_
  rename reslik=adjres2;
  rename pred=fitted2; run;
```
Example: SAS Codes (Print Output)

data combo;
   merge temp temp1;
   by id;
run;
proc print data=combo noobs;
   var a c m count fitted1 adjres1 fitted2 adjres2;
run;
Partial Output

<table>
<thead>
<tr>
<th>a</th>
<th>c</th>
<th>m</th>
<th>count</th>
<th>fitted1</th>
<th>adjres1</th>
<th>fitted2</th>
<th>adjres2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>911</td>
<td>909.240</td>
<td>3.6955</td>
<td>910.383</td>
<td>0.63332</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>538</td>
<td>438.840</td>
<td>12.7451</td>
<td>538.617</td>
<td>-0.63333</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>44</td>
<td>45.760</td>
<td>-3.6956</td>
<td>44.617</td>
<td>-0.63336</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>456</td>
<td>555.160</td>
<td>-12.8498</td>
<td>455.383</td>
<td>0.63332</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>3</td>
<td>4.761</td>
<td>-3.7093</td>
<td>3.617</td>
<td>-0.63848</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>43</td>
<td>142.160</td>
<td>-13.7941</td>
<td>42.383</td>
<td>0.63329</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>2</td>
<td>0.240</td>
<td>2.3852</td>
<td>1.383</td>
<td>0.60617</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
<td>279</td>
<td>179.840</td>
<td>12.4902</td>
<td>279.617</td>
<td>-0.63333</td>
</tr>
</tbody>
</table>
Tests About Partial Association

- Test about partial association by comparing different loglinear models.

- For model \((AC, AM, CM)\), the null hypothesis of no partial association between alcohol use and cigarette smoking states that the \(\lambda^{AC}\) term equals zero.

- To test whether the simpler model \((AM, CM)\) of \(A - C\) conditional independence holds against the alternative that the model \((AC, AM, CM)\) holds
Likelihood Ratio Tests

• The likelihood ratio statistic $-2(L_0 - L_1)$ is identical to the goodness-of-fit $G^2$ statistics (deviance) for the model without that term and the model containing that term.

• The $d.f.$ equals the difference between the corresponding $d.f.$ values.

• The test statistic for testing $\lambda^{AC} = 0$ in model $(AC, AM, CM)$ is the difference

$$G^2[(AM, CM)|(AC, AM, CM)] = G^2(AM, CM) - G^2(AC, AM, CM)$$

between $G^2 = 187.8(d.f. = 2)$ for model $(AM, CM)$ and $G^2 = 0.4(d.f. = 1)$ for model $(AC, AM, CM)$. 
**Example: Drug Use Data**

<table>
<thead>
<tr>
<th>Model</th>
<th>$G^2$</th>
<th>$\chi^2$</th>
<th>d.f.</th>
<th>$P$ – value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A,C,M)</td>
<td>1286.0</td>
<td>1411.4</td>
<td>4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(A,CM)</td>
<td>534.2</td>
<td>505.6</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(C,AM)</td>
<td>939.6</td>
<td>824.2</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(M,AC)</td>
<td>843.8</td>
<td>704.9</td>
<td>3</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,AM)</td>
<td>497.4</td>
<td>443.8</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,CM)</td>
<td>92.0</td>
<td>80.8</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AM,CM)</td>
<td>187.8</td>
<td>177.6</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(AC,AM,CM)</td>
<td>0.4</td>
<td>0.4</td>
<td>1</td>
<td>0.54</td>
</tr>
<tr>
<td>(ACM)</td>
<td>0.0</td>
<td>0.0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Example

• The difference of 187.4 is based on $d.f. = 2 - 1 = 1$.

• The small $P$-value provides strong evidence against the null hypothesis and in favor of an $A - C$ partial association.

• The statistics comparing models $(AC, CM)$ and $(AC, AM)$ with model $(AC, AM, CM)$ also provides strong evidence of $A - M$ and $C - M$ partial associations.

• We should use the model $(AC, AM, CM)$ rather than any simpler models.
Notes

• The sample size can strongly influence results of any inferential procedure.

• One is more likely to detect an effect of given size as the sample size increases.

• For small sizes, reality may be much more complex than indicated by the simplest model that passes the goodness-of-fit test.

• For large sample sizes, statistically significant effects can be weak and unimportant.

• This is a limitation of hypothesis testing.
Confidence Intervals For Odds Ratios

- ML estimators of parameters have large sample normal distributions.
- For models in which the highest order terms are twofactor associations, the estimates refer to the conditional log odds ratios.
- One can use the estimates along with their standard errors to construct confidence intervals for true log odds ratios and then exponentiate them to form intervals for odds ratios.
**Example**

- Assume the model \((AC, AM, CM)\) holds.
- We estimate the conditional odds ratio between alcohol use and cigarette use.
- The software reports \(\hat{\lambda}_{11}^{AC} = 2.054\) with \(ASE = 0.174\)
- The lone nonzero term equals the estimated conditional log odds ratio.
- A 95\% C.I. for the true conditional log odds ratio is
  \[
  2.054 \pm 1.96 \times (0.174) = (1.71, 2.39)
  \]
- Thus, a 95\% C.I. for the true conditional odds ratio is
  \[
  (e^{1.71}, e^{2.39}) = (5.5, 11.0)
  \]
Example

- For model \((AC, AM, CM)\), the 95% confidence intervals are \((8.0, 49.2)\) for the \(A - M\) conditional odds ratio and \((12.5, 23.8)\) for the \(C - M\) conditional odds ratio.

- The intervals are wide, but these associations also are strong.

- In summary, there is a strong tendency for users of one drug to be users of a second drug, and this is true both for users and nonusers of the third drug.
## Automobile Accident Example

<table>
<thead>
<tr>
<th>Gender</th>
<th>Location</th>
<th>Seat Belt</th>
<th>Injury No</th>
<th>Injury Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Urban</td>
<td>No</td>
<td>7287</td>
<td>996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>11587</td>
<td>759</td>
</tr>
<tr>
<td>Rural</td>
<td>No</td>
<td>3246</td>
<td>973</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>6134</td>
<td>757</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>No</td>
<td>10381</td>
<td>812</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>10969</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>No</td>
<td>6123</td>
<td>1084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>6693</td>
<td>513</td>
<td></td>
</tr>
</tbody>
</table>
Example

• The table refers to observations of 68,694 passengers in autos and light trucks involved in accidents in the state of Maine in 1991.

• The table classifies passengers by *gender* (*G*), *location* of accident (*L*), *seat belt* use (*S*), and *injury* (*I*). 
Four-way Tables

• Basic concepts of three-way tables extend readily to multi-way tables.

• We consider a four-way table with variables $W, X, Y, \text{ and } Z$.

• Interpretations are simplest when there are no three-factor interaction terms.

• The homogeneous association model is $(WX, WY, WZ, XY, XZ, YZ)$.

• Here each pair of variables is conditionally dependent, with the same odds ratios at each combination of levels of the other two variables.
Four-way Tables

- An absence of a two factor term implies conditional independence for those variables.
- Model \((W X, W Y, W Z, X Z, Y Z)\) does not contain an \(X - Y\) term, so it treats \(X\) and \(Y\) as conditionally independent at each combination of levels of \(W\) and \(Z\).
- A model could contain any of the four possible three factor interaction terms: \(W X Y, W X Z, W Y Z, X Y Z\).
- The saturated model contains all these terms plus a four factor interaction term.
Example: SAS Codes

data injury;
   input G L S I count;
datalines;
0 0 0 0 7287
0 0 0 1 996
... 
1 1 1 1 5
;
run;
SAS Codes

ods listing close;
proc catmod data=injury;
    weight count;
    model G*I*L*S= _response_/ pred=freq;
    loglin g|i g|l g|s i|l i|s l|s;
    ods output predictedfreqs=temp1;
run;
quit;
data temp1 (keep=p1 functionnum);
    set temp1;
    rename predfunction=p1;
run;
SAS Codes

proc catmod data=injury;
  weight count;
  model G*I*L*S=_response_ / pred=freq;
  loglin g|l|s g|i i|l i|s;
  ods output predictedfrequs=temp2 anova=temp3;
run;
quit;
ods output close;
ods listing;
data temp2;
  set temp2;
  rename predfunction=p2;
run;
SAS Codes

data combo;
    merge temp1 temp2;
    by functionnum;
    Male=G+0;
    Location=L+0;
    Seat=S+0;
    Injury=I+0;
    rename obsfunction = observed;
run;
proc format;
    value male 0='Female' 1='Male';
    value location 0='Urban' 1='Rural';
    value Yesno 0='No' 1='Yes';
run;
SAS Codes

proc sort data=combo;
  by male location seat injury;
run;
proc print data= combo noobs;
  format male male. location location. seat yesno. injury yesno.;
  var male location seat injury observed p1 p2;
run;
## Partial Output

<table>
<thead>
<tr>
<th>Male</th>
<th>Location</th>
<th>Seat</th>
<th>Injury</th>
<th>observed</th>
<th>p1</th>
<th>p2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Urban</td>
<td>No</td>
<td>No</td>
<td>7287</td>
<td>7166.369</td>
<td>7273.214</td>
</tr>
<tr>
<td>Female</td>
<td>Urban</td>
<td>No</td>
<td>Yes</td>
<td>996</td>
<td>993.0169</td>
<td>1009.786</td>
</tr>
<tr>
<td>Female</td>
<td>Urban</td>
<td>Yes</td>
<td>No</td>
<td>11587</td>
<td>11748.31</td>
<td>11632.62</td>
</tr>
<tr>
<td>Female</td>
<td>Urban</td>
<td>Yes</td>
<td>Yes</td>
<td>759</td>
<td>721.3055</td>
<td>713.3779</td>
</tr>
<tr>
<td>Female</td>
<td>Rural</td>
<td>No</td>
<td>No</td>
<td>3246</td>
<td>3353.829</td>
<td>3254.662</td>
</tr>
<tr>
<td>Female</td>
<td>Rural</td>
<td>No</td>
<td>Yes</td>
<td>973</td>
<td>988.7848</td>
<td>964.3382</td>
</tr>
<tr>
<td>Female</td>
<td>Rural</td>
<td>Yes</td>
<td>No</td>
<td>6134</td>
<td>5985.493</td>
<td>6093.502</td>
</tr>
<tr>
<td>Female</td>
<td>Rural</td>
<td>Yes</td>
<td>Yes</td>
<td>757</td>
<td>781.8927</td>
<td>797.4979</td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>No</td>
<td>No</td>
<td>10381</td>
<td>10471.5</td>
<td>10358.93</td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>No</td>
<td>Yes</td>
<td>812</td>
<td>845.1187</td>
<td>834.0683</td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>Yes</td>
<td>No</td>
<td>10969</td>
<td>10837.83</td>
<td>10959.23</td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>Yes</td>
<td>Yes</td>
<td>380</td>
<td>387.5588</td>
<td>389.7677</td>
</tr>
<tr>
<td>Male</td>
<td>Rural</td>
<td>No</td>
<td>No</td>
<td>6123</td>
<td>6045.306</td>
<td>6150.192</td>
</tr>
<tr>
<td>Male</td>
<td>Rural</td>
<td>No</td>
<td>Yes</td>
<td>1084</td>
<td>1038.08</td>
<td>1056.808</td>
</tr>
<tr>
<td>Male</td>
<td>Rural</td>
<td>Yes</td>
<td>No</td>
<td>6693</td>
<td>6811.371</td>
<td>6697.644</td>
</tr>
<tr>
<td>Male</td>
<td>Rural</td>
<td>Yes</td>
<td>Yes</td>
<td>513</td>
<td>518.2429</td>
<td>508.3564</td>
</tr>
</tbody>
</table>
## Example: Automobile Accidents

<table>
<thead>
<tr>
<th>Gender</th>
<th>Location</th>
<th>Belt</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Urban</td>
<td>No</td>
<td>7287</td>
<td>996</td>
<td>7166.4</td>
<td>993.0</td>
<td>7273.2</td>
<td>1009.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>11587</td>
<td>759</td>
<td>11748.3</td>
<td>721.3</td>
<td>11632.6</td>
<td>713.4</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>No</td>
<td>3246</td>
<td>973</td>
<td>3353.8</td>
<td>988.8</td>
<td>3254.7</td>
<td>964.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>6134</td>
<td>757</td>
<td>5985.5</td>
<td>781.9</td>
<td>6093.5</td>
<td>797.5</td>
</tr>
<tr>
<td>Male</td>
<td>Urban</td>
<td>No</td>
<td>10381</td>
<td>812</td>
<td>10471.5</td>
<td>845.1</td>
<td>10358.9</td>
<td>834.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>10969</td>
<td>380</td>
<td>10837.8</td>
<td>387.6</td>
<td>10959.2</td>
<td>389.8</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>No</td>
<td>6123</td>
<td>1084</td>
<td>6045.3</td>
<td>1038.1</td>
<td>6150.2</td>
<td>1056.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yes</td>
<td>6693</td>
<td>513</td>
<td>6811.4</td>
<td>518.2</td>
<td>6697.6</td>
<td>508.4</td>
</tr>
</tbody>
</table>
## Example: Goodness of Fit Tests

<table>
<thead>
<tr>
<th>Model</th>
<th>$G^2$</th>
<th>d.f.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G,I,L,S)</td>
<td>2792.8</td>
<td>11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>(GI,GL,GS,IL,IS,LS)</td>
<td>23.4</td>
<td>5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(GIL,GIS,GLS,ILS)</td>
<td>1.3</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>(GIS,GL,IL,LS)</td>
<td>22.8</td>
<td>4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>(GLS,GI,IL,IS)</td>
<td>7.5</td>
<td>4</td>
<td>0.11</td>
</tr>
<tr>
<td>(ILS,GI,GL,GS)</td>
<td>20.6</td>
<td>4</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Discussions

• Model \((G, I, L, S)\), which implies mutual independence of the four variables, fits very poorly.

• Model \((GI, GL, GS, IL, IS, LS)\) fits much better but still has lack of fit.

• Model \((GIL, GIS, GLS, ILS)\) contains all three factor interactions seems to fit well, but is quite complex and difficult to interpret.

• This suggests studying models that are more complex than \((GI, GL, GS, IL, IS, LS)\) but simpler than \((GIL, GIS, GLS, ILS)\).
Three-Factor Interaction

- Interpretations are more complicated when a model contains three-factor interaction terms.
- Of the four possible models, \((GLS, GI, IL, IS)\) appears to fit best.
- For model \((GLS, GI, IL, IS)\), each pair of variables is conditionally dependent.
- At each level of \(I\) the association between \(G\) and \(L\) or between \(G\) and \(S\) or between \(L\) and \(S\) varies across the levels of the remaining variables.
- For this model, it is inappropriate to interpret the \(G − L\), \(G − S\) and \(L − S\) two factor terms on their own.
Three-Factor Interactions

- One would not convert $\hat{\lambda}^{GS}$ to a fitted $G - S$ odds ratio, because the presence of the $GLS$ three-factor interaction term implies that the $G - S$ odds ratio varies across the levels of $L$.

- Since $I$ does not occur in a three factor interaction, the conditional odds ratio between $I$ and each variable is the same at each combination of levels of the other two variables.

- When a model has a three factor interaction term but no term of higher order than that, one can study the interaction by calculating fitted odds ratios between two variables at each level of the third.
# Estimated Conditional Odds Ratios

<table>
<thead>
<tr>
<th>Odds Ratio</th>
<th>Loglinear Models (GI, GL, GS, IL, IS, LS)</th>
<th>Loglinear Models (GLS, GI, IL, IS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G - I$</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>$I - L$</td>
<td>2.13</td>
<td>2.13</td>
</tr>
<tr>
<td>$I - S$</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>$G - L(S = no)$</td>
<td>1.23</td>
<td>1.33</td>
</tr>
<tr>
<td>$G - L(S = yes)$</td>
<td>1.23</td>
<td>1.17</td>
</tr>
<tr>
<td>$G - S(L = urban)$</td>
<td>0.63</td>
<td>0.66</td>
</tr>
<tr>
<td>$G - S(L = rural)$</td>
<td>0.63</td>
<td>0.58</td>
</tr>
<tr>
<td>$L - S(G = female)$</td>
<td>1.09</td>
<td>1.17</td>
</tr>
<tr>
<td>$L - S(G = male)$</td>
<td>1.09</td>
<td>1.03</td>
</tr>
</tbody>
</table>
Statistical vs. Practical Significance

• Model \((GLS, GI, IL, IS)\) seems to fit much better than \((GI, GL, GS, IL, IS, LS)\).

• The difference in \(G^2\) values of \(23.4 - 7.5 = 15.9\) being based on \(d.f. = 5 - 4 = 1\).

• The fitted odds ratios, however, show that the degree of three-factor interaction is weak.

• The fitted odds ratio between any two of \(G, L,\) and \(S\) is similar at both levels of the third variable.

• The significantly better fit of model \((GLS, GI, IL, IS)\) mainly reflects the enormous sample size.
Effect of Large Sample Sizes

- Large sample provides small standard errors.
- A statistically significant effect need not be important in a practical sense.
- With huge samples, it is crucial to focus on estimation rather than hypothesis testing.
- e.g. the model \((GI, GL, GS, IL, IS, LS)\) is adequate for practical purposes. .. Simpler models are easier to summarize.
- One should not use goodness-of-fit tests alone to select a final model.
Dissimilarity Index

- For a table of arbitrary dimension with cell counts \( \{n_i = np_i\} \) and fitted values \( \{\hat{\mu}_i = n\hat{\pi}_i\} \) one can summarize the closeness of the model fit to the sample data by the dissimilarity index

\[
D = \sum |n_i - \hat{\mu}_i|/(2n) = \sum |p_i - \hat{\pi}_i|/2
\]

- This index takes values between 0 and 1, with smaller values representing a better fit.

- It represents the proportion of sample cases that must move to different cells in order for the model to achieve a perfect fit.
Dissimilarity Index

- The dissimilarity index $D$ estimates a corresponding index $\Delta$ that describes model lack-of-fit in the population sampled.
- The value $\Delta = 0$ occurs when the model holds perfectly.
- In that case $D$ overestimates $\Delta$, substantially so for small samples, because of sampling variation.
- When the model does not hold, for sufficiently large $n$, the goodness-of-fit statistics $G^2$ and $\chi^2$ will be large, showing lack-of-fit.
- The estimator $D$ then reveals whether the lack of fit suggested by those statistics is important in practical sense.
- $D < 0.03$ suggests that the sample data follow the model quite closely, even though the model is not *perfect*. 
Some SAS Codes

- As a continuation of the previous SAS codes, we can add the following statements to compute the dissimilarity indices for those models.

```sas
proc sql;
    select sum(abs(observed-p1))/
        (2*sum(observed)) as d1,
    sum( abs(observed-p2) ) / (2*sum(observed) )
    as d2
    from combo;
quit;
```
Some SAS Codes

<table>
<thead>
<tr>
<th>d1</th>
<th>d2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.008219</td>
<td>0.002507</td>
</tr>
</tbody>
</table>

- For either model, moving less than 1% of the data yields a perfect fit.
- The small value of $D$ for the model $(GI, GL, GS, IL, IS, LS)$ suggests that in practical terms, this model provides a decent fit.
Loglinear-Logit Connection

- Consider the loglinear model of homogeneous association in three-way tables

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

- Suppose \( Y \) is binary, and we treat it as a response and \( X \) and \( Z \) as explanatory.

- Let \( \pi \) denote the probability that \( Y = 1 \), which depends on the levels of \( X \) and \( Z \).
Loglinear-Logit Connection

- The logit for $Y$ is

$$\text{logit}(\pi_{ik}) = \log\left(\frac{\pi_{ik}}{1 - \pi_{ik}}\right) = \log\left(\frac{P(Y = 1|X = i, Z = k)}{P(Y = 0|X = i, Z = k)}\right)$$

$$= \log\left(\frac{\mu_{i1k}}{\mu_{i2k}}\right) = \log(\mu_{i1k}) - \log(\mu_{i2k})$$

$$= (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ})$$

$$- (\lambda + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ})$$

$$= (\lambda_1^Y - \lambda_2^Y) + (\lambda_{i1}^{XY} - \lambda_{i2}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{2k}^{YZ})$$
Loglinear-Logit Connection

- The first term is a constant which does not depend on $i$ or $k$.
- The second term depends on the level $i$ of $X$.
- The third term depends on the level $k$ of $Z$.
- The logit has the additive form

$$\text{logit}(\pi_{ik}) = \alpha + \beta_i^X + \beta_k^Z$$
Accident Data Revisited

• The model \((GLS, GI, IL, IS)\) has the form

\[
\begin{align*}
\log \mu_{gils} &= \lambda + \lambda^G_g + \lambda^I_i + \lambda^L_l + \lambda^S_s + \lambda^{GI}_{gi} + \lambda^{GL}_{gl} \\
&\quad + \lambda^{GS}_{gs} + \lambda^{IL}_{il} + \lambda^{IS}_{is} + \lambda^{LS}_{ls} + \lambda^{GLS}_{glsl}
\end{align*}
\]

• One can treat injury \((I)\) as a response variable and gender \((G)\), location \((L)\) and seat belt use \((S)\) as explanatory variables.

• Let \(\pi\) denote the probability of injury.

• Forming \(\text{logit}(\pi_{gls})\) at each combination of levels of \(G, L,\) and \(S\), one can show that the above loglinear model is equivalent to logit model

\[
\text{logit}(\pi_{gls}) = \alpha + \beta^G_g + \beta^L_l + \beta^S_s
\]
Interpretations

- Here, $G, L$, and $S$ all affect $I$, but without interacting.

- The parameters in the two models are related by

$$
\alpha = \lambda_1^I - \lambda_2^I, \beta^G_g = \lambda_{g1}^{GI} - \lambda_{g2}^{GI}, \beta^L_l = \lambda_{1l}^{IL} - \lambda_{2l}^{IL}, \beta^S_s = \lambda_{1s}^{IS} - \lambda_{2s}^{IS}
$$

- In the logit calculation, all terms in the loglinear model not having the injury index $i$ in the subscript are cancelled.
## Example: Birth Control Data

<table>
<thead>
<tr>
<th>Premarital Sex</th>
<th>Teenage Birth Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly Disagree</td>
</tr>
<tr>
<td>Always wrong</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>(42.4)</td>
</tr>
<tr>
<td></td>
<td>7.6</td>
</tr>
<tr>
<td>Almost always wrong</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>(16.0)</td>
</tr>
<tr>
<td></td>
<td>2.3</td>
</tr>
<tr>
<td>Wrong only sometimes</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>(30.1)</td>
</tr>
<tr>
<td></td>
<td>-2.7</td>
</tr>
<tr>
<td>Not wrong at all</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>(70.6)</td>
</tr>
<tr>
<td></td>
<td>-6.1</td>
</tr>
</tbody>
</table>
Example: Birth Control Data

- The loglinear model of independence has goodness-of-fit statistic $G^2 = 127.6$ based on $d.f. = 9$.
- The model fits poorly, providing strong evidence of dependence.
- But, adding the interaction term makes the model saturated and of little use.
Example: Birth Control Data

- The table also contains fitted values and adjusted residuals for the independence model.
- The residuals in the corners of the table are very large.
- Observed counts are much larger than the independence model predicts in the corners where both responses are the most negative possible or the most positive possible.
- Cross-classifications of ordinal variables often exhibit their greatest deviations from independence in the corner cells.
- This pattern indicates lack of fit in the form of a positive trend.
- Subjects who feel more favorable to making birth control available to teenagers also tend to feel more tolerant about premarital sex.
Linear-by-Linear Association

- Assign scores $u_i$ to the $I$ rows and $v_j$ to the $J$ columns.
- We must have $u_1 \leq u_2 \leq \cdots \leq u_I$ and $v_1 \leq v_2 \leq \cdots \leq v_J$ to reflect the category ordering.
- The model is

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j$$

- The independence model is the special case $\beta = 0$. The final term represents the deviation from independence.
Notes

• The deviation is linear in the $Y$ scores at a fixed level of $X$ and linear in the $X$ scores at a fixed level of $Y$.

• In column $j$, the deviation is a linear function of $X$, having form $(\text{slope}) \times (\text{score for } X)$, with slope $\beta v_j$

• This model is called the linear-by-linear association model.
Interpretations

• The parameter $\beta$ refers to the direction and strength of association.

• When $\beta > 0$, there is a tendency for $Y$ to increase as $X$ increases.

• When $\beta < 0$, there is a tendency for $Y$ to decrease as $X$ increases.

• When the data display a positive or negative trend, this model usually fits much better than the independence model.
Describing Associations

• For the $2 \times 2$ table using the cells intersecting rows $a$ and $c$ with columns $b$ and $d$, the model has odds ratio equal to

$$\frac{\mu_{ab}\mu_{cd}}{\mu_{ad}\mu_{cb}} = \exp[\beta(u_c - u_a)(v_d - v_b)]$$

• The association is stronger as $|\beta|$ increases.

• For given $\beta$ pairs of categories that are farther apart have greater differences between their scores and odds ratios farther from 1.
Further Comments

- In practice, the most common choice of scores is \( u_i = i \) and \( v_j = j \), simply the row and column numbers.

- The odds ratios formed using adjacent rows and adjacent columns are called *local odds ratios*.

- For these unit spaced scores, the local odds ratios simplifies so that \( e^\beta \) is the common value of all the local odds ratios.

- Any set of equally-spaced row and column scores has the property of uniform local odds ratios.

- This special case of the model is called *uniform association*. 
Example: SAS Codes (Inputting the Data)

data sex;
   input premar birth count @@;
   assoc = premar*birth;
datalines;
  1 4 38 1 3 60 1 2 68 1 1 81
  2 4 14 2 3 29 2 2 26 2 1 24
  3 4 42 3 3 74 3 2 41 3 1 18
  4 4 157 4 3 161 4 2 57 4 1 36
;
run;
SAS Codes: Fitting Independence Model

```sas
proc genmod data = sex order=data;
  class premar birth;
  model count = premar birth /dist = poi link = log ;
  output out = table7_3 pred=pred STDRESCHI = r;
run;
proc sort data=table7_3;
  by birth;
run;
```
SAS Codes: Printing Some Output

```sas
proc format;
    value premar 1='Always wrong' 2='Usually Wrong'
      3='Sometimes Wrong' 4='Never wrong';
    value birth 1='Strongly Disagree' 2='Disagree'
      3='Agree' 4='Strongly Agree';
run;
proc print data = table7 noobs;
    format premar premar. birth birth.;
run;
```
## Output

<table>
<thead>
<tr>
<th>premar</th>
<th>birth</th>
<th>count</th>
<th>assoc</th>
<th>pred</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always wrong</td>
<td>Strongly Disagree</td>
<td>81</td>
<td>1</td>
<td>42.411</td>
<td>7.60318</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Strongly Disagree</td>
<td>24</td>
<td>2</td>
<td>15.969</td>
<td>2.32832</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Strongly Disagree</td>
<td>18</td>
<td>3</td>
<td>30.049</td>
<td>-2.68175</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Strongly Disagree</td>
<td>36</td>
<td>4</td>
<td>70.571</td>
<td>-6.06333</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Disagree</td>
<td>68</td>
<td>2</td>
<td>51.214</td>
<td>3.07671</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Disagree</td>
<td>26</td>
<td>4</td>
<td>19.283</td>
<td>1.81148</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Disagree</td>
<td>41</td>
<td>6</td>
<td>36.285</td>
<td>0.97623</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Disagree</td>
<td>57</td>
<td>8</td>
<td>85.218</td>
<td>-4.60386</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Agree</td>
<td>60</td>
<td>3</td>
<td>86.423</td>
<td>-4.11671</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Agree</td>
<td>29</td>
<td>6</td>
<td>32.540</td>
<td>-0.81148</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Agree</td>
<td>74</td>
<td>9</td>
<td>61.231</td>
<td>2.24730</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Agree</td>
<td>161</td>
<td>12</td>
<td>143.806</td>
<td>2.38456</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Strongly Agree</td>
<td>38</td>
<td>4</td>
<td>66.951</td>
<td>-4.83965</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Strongly Agree</td>
<td>14</td>
<td>8</td>
<td>25.208</td>
<td>-2.75682</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Strongly Agree</td>
<td>42</td>
<td>12</td>
<td>47.435</td>
<td>-1.02637</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Strongly Agree</td>
<td>157</td>
<td>16</td>
<td>111.405</td>
<td>6.78455</td>
</tr>
</tbody>
</table>
SAS Codes: PROC CATMOD

proc catmod data=sex;
  weight count;
  model premar*birth= _response_
      / predict=freq noprofile noiter
         noresponse ;
  loglin premar birth;
run;
quit;
### Partial Output

**Maximum Likelihood Predicted Values for Frequencies**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Frequency</th>
<th>Error</th>
<th>Frequency</th>
<th>Error</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Standard</td>
<td></td>
<td>Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>81</td>
<td>8.597365</td>
<td>42.41145</td>
<td>3.835381</td>
<td>38.58855</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>68</td>
<td>7.937662</td>
<td>51.21382</td>
<td>4.314471</td>
<td>16.78618</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>60</td>
<td>7.490815</td>
<td>86.42333</td>
<td>6.095867</td>
<td>-26.4233</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>38</td>
<td>6.036605</td>
<td>66.9514</td>
<td>5.130779</td>
<td>-28.9514</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>24</td>
<td>4.835077</td>
<td>15.96868</td>
<td>1.94804</td>
<td>8.031317</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>26</td>
<td>5.026925</td>
<td>19.28294</td>
<td>2.265306</td>
<td>6.717063</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>29</td>
<td>5.300169</td>
<td>32.53996</td>
<td>3.516572</td>
<td>-3.53996</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>14</td>
<td>3.713265</td>
<td>25.20842</td>
<td>2.827013</td>
<td>-11.2084</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>18</td>
<td>4.201203</td>
<td>30.0486</td>
<td>2.981299</td>
<td>-12.0486</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>41</td>
<td>6.259766</td>
<td>36.2851</td>
<td>3.396646</td>
<td>4.714903</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>74</td>
<td>8.251448</td>
<td>61.2311</td>
<td>4.989831</td>
<td>12.7689</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>42</td>
<td>6.332064</td>
<td>47.43521</td>
<td>4.118554</td>
<td>-5.43521</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>36</td>
<td>5.882213</td>
<td>70.57127</td>
<td>5.716958</td>
<td>-34.5713</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>57</td>
<td>7.313779</td>
<td>85.21814</td>
<td>6.309364</td>
<td>-28.2181</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>161</td>
<td>11.53289</td>
<td>143.8056</td>
<td>8.335372</td>
<td>17.19438</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>157</td>
<td>11.41846</td>
<td>111.405</td>
<td>7.268973</td>
<td>45.59503</td>
</tr>
</tbody>
</table>
SAS Codes: Fitting L by L Model

\begin{verbatim}
proc genmod data=sex order=data;
   format premar premar. birth birth. ;
   class premar birth;
   model count = premar birth assoc/
                  dist=poi link=log ;
   output out=temp p=predict ;
run;
\end{verbatim}
Output

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>8</td>
<td>11.5337</td>
<td>1.4417</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>8</td>
<td>11.5337</td>
<td>1.4417</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>8</td>
<td>11.5085</td>
<td>1.4386</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>8</td>
<td>11.5085</td>
<td>1.4386</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>3041.7446</td>
<td></td>
</tr>
</tbody>
</table>
### Output

**Analysis Of Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Confidence</th>
<th>Limits</th>
<th>Square</th>
<th>Pr&gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2.3532</td>
<td>0.2258</td>
<td>1.9106</td>
<td>2.7957</td>
<td>108.59</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>premar Always wrong</td>
<td>1</td>
<td>1.7537</td>
<td>0.2343</td>
<td>1.2944</td>
<td>2.2129</td>
<td>56.01</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>premar Usually Wrong</td>
<td>1</td>
<td>0.1077</td>
<td>0.1988</td>
<td>-0.2820</td>
<td>0.4974</td>
<td>0.29</td>
<td>0.5880</td>
</tr>
<tr>
<td>premar Sometimes Wrong</td>
<td>1</td>
<td>-0.0163</td>
<td>0.1264</td>
<td>-0.2641</td>
<td>0.2314</td>
<td>0.02</td>
<td>0.8972</td>
</tr>
<tr>
<td>premar Never Wrong</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>birth Strongly Agree</td>
<td>1</td>
<td>-1.8797</td>
<td>0.2491</td>
<td>-2.3679</td>
<td>-1.3914</td>
<td>56.94</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>birth Agree</td>
<td>1</td>
<td>-0.7245</td>
<td>0.1620</td>
<td>-1.0420</td>
<td>-0.4070</td>
<td>20.00</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>birth Disagree</td>
<td>1</td>
<td>-0.4641</td>
<td>0.1195</td>
<td>-0.6984</td>
<td>-0.2298</td>
<td>15.08</td>
<td>0.0001</td>
</tr>
<tr>
<td>birth Strongly Disagree</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>assoc</td>
<td>1</td>
<td>0.2858</td>
<td>0.0282</td>
<td>0.2305</td>
<td>0.3412</td>
<td>102.46</td>
<td>&lt;.0001</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>
<td></td>
</tr>
</tbody>
</table>
SAS Codes: Printing Predicted Values

```sas
proc sort data = temp;
   by birth;
run;
proc print data = temp;
   format premar premar.   birth birth. ;
   var premar birth count predict;
run;
```
## Output

<table>
<thead>
<tr>
<th>premar</th>
<th>birth</th>
<th>count</th>
<th>predict</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always wrong</td>
<td>Strongly Disagree</td>
<td>81</td>
<td>80.857</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Strongly Disagree</td>
<td>24</td>
<td>20.750</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Strongly Disagree</td>
<td>18</td>
<td>24.394</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Strongly Disagree</td>
<td>36</td>
<td>33.000</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Disagree</td>
<td>68</td>
<td>67.654</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Disagree</td>
<td>26</td>
<td>23.107</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Disagree</td>
<td>41</td>
<td>36.152</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Disagree</td>
<td>57</td>
<td>65.088</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Agree</td>
<td>60</td>
<td>69.396</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Agree</td>
<td>29</td>
<td>31.543</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Agree</td>
<td>74</td>
<td>65.681</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Agree</td>
<td>161</td>
<td>157.379</td>
</tr>
<tr>
<td>Always wrong</td>
<td>Strongly Agree</td>
<td>38</td>
<td>29.094</td>
</tr>
<tr>
<td>Usually Wrong</td>
<td>Strongly Agree</td>
<td>14</td>
<td>17.600</td>
</tr>
<tr>
<td>Sometimes Wrong</td>
<td>Strongly Agree</td>
<td>42</td>
<td>48.773</td>
</tr>
<tr>
<td>Never wrong</td>
<td>Strongly Agree</td>
<td>157</td>
<td>155.533</td>
</tr>
</tbody>
</table>
## Fit of The Linear Association Model

<table>
<thead>
<tr>
<th>Premarital Sex</th>
<th>Strongly Disagree</th>
<th>Strongly Disagree</th>
<th>Strongly Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always wrong</td>
<td>81 (80.91)</td>
<td>68 (67.6)</td>
<td>60 (69.4)</td>
<td>38 (29.1)</td>
</tr>
<tr>
<td>Almost always wrong</td>
<td>24 (20.8)</td>
<td>26 (23.1)</td>
<td>29 (31.5)</td>
<td>14 (17.6)</td>
</tr>
<tr>
<td>Wrong only sometimes</td>
<td>18 (24.4)</td>
<td>41 (36.1)</td>
<td>74 (65.7)</td>
<td>42 (48.8)</td>
</tr>
<tr>
<td>Not wrong at all</td>
<td>36 (33)</td>
<td>57 (65.1)</td>
<td>161 (157.4)</td>
<td>157 (155.5)</td>
</tr>
</tbody>
</table>
Example: Continued

- ML estimate of $\beta = 0.286$, with $ASE = 0.028$.

- The positive estimate suggests that subjects having more favorable attitudes about availability of teen birth control also tend have more tolerant attitudes about premarital sex.

- Goodness-of-fit statistics are: $G^2 = 11.5$ and $\chi^2 = 11.5$ with $d.f. = 8$.

- P-value is $0.1749$ suggesting the model to be a good fit.
Example: Continued

• The estimated local odds ratio is

\[ \exp(\hat{\beta}) = \exp(0.286) = 1.33 \]

• The strength of association seems weak.

• Non-local odds ratios are stronger.

• The estimated odds ratio for the four corner cells equals

\[ \exp[\hat{\beta}(u_4 - u_1)(v_4 - v_1)] = \exp[0.286(4-1)(4-1)] = \exp(2.57) = 13.199 \]
Notes

• Two sets of scores having the same spacing yield the same estimate of $\beta$ and the same fit.

• Any other set of equally spaced scores yield the same fit but an appropriately rescaled estimate of $\beta$, so that the fitted odds ratios do not change.

• It is not necessary to use equally-spaced scores in the $L \times L$ model.
Ordinal Tests of Independence

- To test independence, $H_0 : \beta = 0$.

- The LR test equals the reduction in $G^2$ between the independence ($I$) model and $L \times L$ model.

- $G^2(I|L \times L) = G^2(I) - G^2(L \times L)$

- This statistic refers to a single parameter $\beta$, and is based on $d.f. = 1$.

- For the example, this statistic equals $127.6 - 11.5 = 116.1$ with P-value $< 0.0001$, which shows extremely strong association.

- The Wald’s statistic provides an alternative to test this hypothesis.
Tests of Conditional Independence

- We have seen how to test a partial association by comparing two loglinear models that contain or omit that association.

- The LR test compares the models by the difference of the $G^2$ statistics, which is identical to the difference of deviances.

- An important application of this test refers to the null hypothesis of X-Y conditional independence.

- One compares the model $(XZ, YZ)$ of $X \perp Y$ conditional independence to the more complex model $(XY, XZ, YZ)$ of homogeneous association.

- The test statistic is $G^2[(XZ, YZ) | (XY, XZ, YZ)] = G^2(XZ, YZ) - G^2(XY, XZ, YZ)$.

- This test assumes that the homogeneous association model holds.
Continued …

- For $2 \times 2 \times K$ tables, the test of conditional independence comparing two loglinear or logit models has the same purpose as the Cochran- Mantel-Haenszel (CMH) test.

- The CMH test works well when the $X - Y$ odds ratio is similar in each partial table.

- It is also naturally directed toward the alternative of homogeneous association.

- For large samples, the model based LR test usually gives similar results as the CMH test.
**Direct Goodness-of-Fit Test**

- A statistic of the form $G^2(\text{XY}, \text{XZ})$ (*deviance*) does not require an assumption about homogeneous association.

- The statistic could be large if there are three-factor interactions, or if there is no three-factor interaction but conditional dependence.

- A disadvantage in using this test is that it often has low power.
Detecting Ordinal Conditional Association

• A useful model

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

The model is called a homogeneous linear – by – linear association model.

• The conditional independence model \((XZ, YZ)\) is the special case of this model with \(\beta = 0\).

• Unless this models fits very poorly, the tests comparing this model are more powerful than tests that ignore the ordering.
Cochran-Mantel-Haenszel Test for $2 \times 2 \times K$ table

- To test: $X$ and $Y$ are conditionally independent given $Z$.
- So, $H_0 : \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(K)} = 1.0$.
- In the $k$-th partial table, the row totals are $n_{1+k}, n_{2+k}$ and column totals are $n_{+1k}, n_{+2k}$.
- Given both these totals, $n_{11k}$ has a hypergeometric distribution and that determines all other cell counts in the $k$-th partial table.
Cochran-Mantel-Haenszel Test (Continued)

- Under the null hypothesis of independence,
  \[ \mu_{11k} = E(n_{11k}) = \frac{n_{1+k}n_{+1k}}{n}, \]
  \[ Var(n_{11k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^2(n_{++k} - 1)}, k = 1, \ldots, K \]

- The test statistic is given by
  \[ CMH = \frac{\left[ \sum_{k=1}^{K} (n_{11k} - \mu_{11k}) \right]^2}{\sum_{k=1}^{K} Var(n_{11k})} \]

- This is called the Cochran–Mantel–Haenszel (CMH) statistic.

- It has a large sample chi-squared distribution with \( df = 1 \).
Generalized Cochran-Mantel-Haenszel Test

- We can generalize CMH test to $I \times J \times K$ tables.
- When $X$ and $Y$ are ordinal, the test statistic generalizes the correlation statistic for two-way tables.
- It is designed to detect a linear trend in the $X - Y$ association that has the same direction in each partial table.
- The generalized correlation statistic has approximately a chi-squared distribution with $df = 1$. 
Generalized CMH Test

- The generalized correlation statistic is:

\[
M^2 = \frac{\left[ \sum_{k=1}^{K} (\Delta_k - E\Delta_k) \right]^2}{\sum_{k=1}^{K} \text{Var}(\Delta_k)}, \quad \Delta_k = \sum_{i=1}^{I} \sum_{j=1}^{J} u_i v_j n_{ijk},
\]

\[
E(\Delta_k) = n_{++k} \mu_k(u) \mu_k(v),
\]

\[
\text{Var}(\Delta_k) = \frac{n_{++k}^2}{(n_{++k} - 1)} \sigma_k(u)^2 \sigma_k(v)^2,
\]

\[
\mu_k(u) = \sum_{i=1}^{I} u_i n_{i+k}/n_{++k},
\]

\[
\mu_k(v) = \sum_{j=1}^{J} v_j n_{+jk}/n_{++k},
\]

\[
\sigma_k^2(u) = \sum_{i=1}^{I} [u_i - \mu_k(u)]^2 n_{i+k}/n_{++k},
\]

\[
\sigma_k^2(v) = \sum_{j=1}^{J} [v_j - \mu_k(v)]^2 n_{+jk}/n_{++k}
\]
Generalized CMH Test

• $X$ is nominal and $Y$ is ordinal

• The test of conditional independence compares the $I$ rows using a statistic based on the variation in those $I$ averaged row mean responses that is designed to detect differences among their true values.

• It has a large sample chi-squared distribution with $df = 1$.

• The formula for this statistic is complex and for $K = 1$, it is

\[
M^2 = (n - 1) \frac{\left[ \sum_{i=1}^{I} \sum_{j=1}^{J} (u_i - \bar{u})(v_j - \bar{v})n_{ij} \right]^2}{\sum_{i=1}^{I} (u_i - \bar{u})^2 n_{i+} + \sum_{j=1}^{J} (v_j - \bar{v})^2 n_{+j}}
\]

where $\bar{u} = \sum_{i=1}^{I} u_i n_{i+} / n$ and $\bar{v} = \sum_{j=1}^{J} v_j n_{+j} / n$. 
Generalized CMH Test

- When $X$ and $Y$ both are nominal, another CMH-type statistic based on $df = (I - 1)(J - 1)$, provides a \textit{general association test}.
- It is designed to detect any type of association that is similar in each partial tables.
- In SAS, the \textit{general association} alternative treats both $X$ and $Y$ as nominal and has $df = (I - 1)(J - 1)$.
- The \textit{row mean scores differ} alternative treats the rows of $X$ as nominal and the columns of $Y$ as ordinal and has $df = I - 1$.
- The \textit{nonzero correlation} alternative treats both $X$ and $Y$ as ordinal and has $df = 1$. 