Tests of Conditional Independence

We’ll talk in general terms of testing whether row \((X)\) and column \((Y)\) classifications are independent conditioning on levels of a third variable \((Z)\).

There are 3 kinds of tests:

1. Likelihood ratio tests (“LR” for short).
   (a) Comparing conditional independence model to homogeneous association model.
   (b) Comparing conditional independence model to saturated model.
2. Wald tests.
3. Efficient score tests, i.e. Generalized CMH.

The LR and Wald tests require the estimation of (model) parameters, while the Efficient score tests do not.

We also need to consider whether categories of variables as ordinal or nominal. We’ll consider 3 cases:

1. Nominal-Nominal
2. Ordinal-Ordinal
3. Nominal-Ordinal
So the possibilities are:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Likelihood</td>
</tr>
<tr>
<td>Row Column</td>
<td>Ratio</td>
</tr>
<tr>
<td>Nominal Nominal</td>
<td></td>
</tr>
<tr>
<td>Nominal Ordinal</td>
<td></td>
</tr>
<tr>
<td>Ordinal Ordinal</td>
<td></td>
</tr>
</tbody>
</table>

As an example for these tests, we’ll use some High School & Beyond data, i.e., the crossclassification of gender (G), SES (S) and high school program type (P).

<table>
<thead>
<tr>
<th>Females</th>
<th>SES</th>
<th>High School Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>VoTech</td>
<td>General</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>middle</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>71</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Males</th>
<th>SES</th>
<th>High School Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>low</td>
<td>VoTech</td>
<td>General</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>31</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>middle</td>
<td>38</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>76</td>
<td>85</td>
</tr>
</tbody>
</table>
Model Based Tests of Conditional Independence

The likelihood ratio test. We compare the fit of the conditional independence model and comparing it to the homogeneous association model.

For example to test whether \( X \) and \( Y \) are conditionally independent given \( Z \), i.e.,

\[
H_0: \quad \text{all } \lambda_{ij}^{XY} = 0
\]

the likelihood ratio test statistic is

\[
G^2 [(XZ, YZ) | (XY, XZ, YZ)] = G^2(XZ, YZ) - G^2(XY, XZ, YZ)
\]

with \( df = df(XZ, YZ) - df(XY, XZ, YZ) \).

Example: \( G = \text{Gender} \), \( S = \text{SES} \), and \( P = \text{Program type} \). Testing whether SES and program type are independent given gender,

\[
H_0: \quad \text{all } \lambda_{ij}^{SP} = 0
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>((GS, GP, SP))</td>
<td>df = 4</td>
<td>1.970</td>
</tr>
<tr>
<td>((GS, GP))</td>
<td>df = 8</td>
<td>55.519</td>
</tr>
</tbody>
</table>
Notes regarding this test:

- This test assumes that \((XY, XZ, YZ)\) holds.
- This single test is preferrable to conducting \((I - 1)(J - 1)\) Wald tests, one for each of the non-redundant \(\lambda_{ij}^{XY}\)'s. For our example, the result is pretty unambiguous; that is,

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>ASE</th>
<th>Wald</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_{iv}^{SP})</td>
<td>1.8133</td>
<td>.3233</td>
<td>31.450</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(\lambda_{ig}^{SP})</td>
<td>1.6600</td>
<td>.3033</td>
<td>29.952</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(\lambda_{mv}^{SP})</td>
<td>1.1848</td>
<td>.2786</td>
<td>18.079</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(\lambda_{mg}^{SP})</td>
<td>.8004</td>
<td>.2639</td>
<td>9.198</td>
<td>.0024</td>
</tr>
</tbody>
</table>

- For binary \(Y\), this is the same as performing the likelihood ratio test of whether \(H_O: all \beta_i^X = 0\) in the logit model

  \[
  \logit(\pi_{Ik}) = \alpha + \beta_i^X + \beta_k^Z
  \]

  which corresponds to the \((XY,XZ,YZ)\) loglinear model.
- For \(2 \times 2 \times K\) tables, this likelihood ratio test of conditional independence has the same purpose as the Cochran–Mantel–Haenszel (CMH) test. For the CMH test,
  - It works the best when the partial odds ratios are similar in each of the partial tables.
  - It’s natural alternative (implicit) hypothesis is that of homogeneous association.
  - CMH is the efficient score tests of \(H_O: \lambda_{ij}^{XY} = 0\) in the loglinear model.
**Direct Goodness-of-Fit Test.** We compare the fit of the conditional independence model to the saturated model; that is,

\[
G^2[(XZ,YZ)|(XYZ)] = G^2(XZ,YZ) - G^2(XYZ)
\]

The null hypothesis for this test statistic is

\[
H_O: \text{ all } \lambda_{ij}^{XY} = 0 \text{ and all } \lambda_{ijk}^{XYZ} = 0
\]

Example: \(G=\) Gender , \(S=\) SES, and \(P=\) Program type. Testing whether SES and program type are independent given gender,

\[
H_O: \text{ all } \lambda_{ij}^{SP} = 0 \text{ and all } \lambda_{ijk}^{GSP} = 0
\]

<table>
<thead>
<tr>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>df</td>
</tr>
<tr>
<td>((GS,GP,SP))</td>
<td>4</td>
</tr>
<tr>
<td>((GS,GP))</td>
<td>8</td>
</tr>
</tbody>
</table>

A direct goodness-of-fit test does not assume that \((YX, XZ, YZ)\) holds, while using \(G^2[(XZ,YZ)|\(XY, XZ, YZ)\]) does assume that the model of homogeneous association holds.

Disadvantages of the goodness-of-fit test as a test of conditional independence

1. It has lower power.

2. It has more \(df\) than the Wald test, the CMH, and the LR test (i.e., \(G^2[(XZ,YZ)|\(XY, XZ, YZ)\)].

42
**Ordinal Conditional Association**

If the categories of one or both variables are ordered, then there are more powerful ways of testing for conditional independence.

With respect to models, we can use a generalized linear by linear model, more specifically a “homogeneous linear by linear association” model.

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

where \(u_i\) are scores for the levels of variable \(X\), and \(v_j\) are scores for the levels of variable \(Y\).

Note:

- The model of conditional independence is a special case of this model; that is, \(\beta = 0\)

- This model is a special case of the homogeneous association model.

Example: Using as equally spaced scores for SES (i.e., \(u_1 = 1\), \(u_2 = 2\), and \(u_3 = 3\)), and unequally spaced scores for program type (i.e., \(v_1 = 1\), \(v_2 = 2\), and \(v_3 = 4\)), we fit the model

\[
\log(\mu_{ijk}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_k^G + \beta u_i v_j + \lambda_{ik}^{SG} + \lambda_{jk}^{PG}
\]
\[
\log(\mu_{ijk}) = \lambda + \lambda^S_i + \lambda^P_j + \lambda^G_k + \beta u_i v_j + \lambda^{SG}_{ik} + \lambda^{PG}_{jk}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>( G^2 )</td>
</tr>
<tr>
<td>((GS, GP, SP))</td>
<td>4</td>
<td>1.970</td>
</tr>
<tr>
<td>((GS, GP, SP))–L×L</td>
<td>7</td>
<td>7.476</td>
</tr>
<tr>
<td>((GS, GP))</td>
<td>8</td>
<td>55.519</td>
</tr>
</tbody>
</table>

- The null hypothesis for the likelihood ratio test statistic (in the last row of table) is now

\[H_O : \beta = 0, \quad \text{with} \quad df = 1\]

whereas before it was

\[H_O : \text{all } \lambda^{SP}_{ij} = 0 \quad \text{with} \quad df = 4\]

The likelihood ratio statistic when using \((GS, GP, SP)\) equals 53.548 (with \(df = 4\)).

Comparing \(G^2/df\) for the two tests,

\[53.548/4 = 13.387 \quad \text{versus} \quad 48.043/1 = 48.043\]

Conclusion: If data exhibit linear by linear partial association, then using scores gives you a stronger (more powerful) test of conditional independence.
• The Wald statistic for $\beta$ equals 43.939, $df = 1$, and $p < .0001$. This is comparable to the new likelihood ratio test statistic.

• $\hat{\beta} = .3234$. The estimated partial odds ratio equals

$$\hat{\theta}_{SP(k)} = \exp \left[ .3234(u_i - u_i')(v_j - v_j') \right]$$

For example, the smallest partial odds ratio is for low and middle SES and votech and general programs,

$$\hat{\theta}_{SP(k)} = \exp \left[ .3234(2 - 1)(2 - 1) \right] = \exp (.3234) = 1.38$$

and the largest partial odds ratio is for low and high SES and votech and academic programs equals

$$\hat{\theta}_{SP(k)} = \exp \left[ .3234(3 - 1)(4 - 1) \right] = \exp(1.9404) = 6.96$$

So far we’ve discussed,

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Likelihood</td>
</tr>
<tr>
<td>Column</td>
<td>Ratio</td>
</tr>
<tr>
<td>Nominal</td>
<td>Nominal</td>
</tr>
<tr>
<td>Nominal</td>
<td>Ordinal</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Ordinal</td>
</tr>
</tbody>
</table>

Let’s now discuss the model based nominal–ordinal case.

For the nominal–ordinal case, we only put in scores for the categories of the ordinal variable and estimate a $\beta$ for each category of the nominal variable.
For example, let’s suppose that we don’t have or don’t want to assume scores for the program types, but only for SES. We can perform likelihood ratio tests (and Wald tests if desired) by fitting the following model

$$\log(\mu_{ijk}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_k^G + \beta_j^P u_i + \lambda_{ik}^{SG} + \lambda_{jk}^{PG}$$

where $u_i$ are scores for SES (i.e., $u_1 = 1$, $u_2 = 2$, and $u_3 = 3$), and $\beta_j^P$ are estimated parameters.

A summary and comparison of the nominal–nominal, ordinal–ordinal, and nominal–ordinal likelihood ratio tests of conditional independence and the model goodness-of-fit tests.

<table>
<thead>
<tr>
<th>Model</th>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>$G^2$</td>
</tr>
<tr>
<td>$(GS, GP, SP)$</td>
<td>4</td>
<td>1.970</td>
</tr>
<tr>
<td>$(GS, GP)$</td>
<td>8</td>
<td>55.519</td>
</tr>
<tr>
<td>$(GS, GP)$ – L×L</td>
<td>7</td>
<td>7.476</td>
</tr>
<tr>
<td>$(GS, GP)$</td>
<td>8</td>
<td>55.519</td>
</tr>
<tr>
<td>$(GS, GP, SP)$ with $u_i$</td>
<td>6</td>
<td>4.076</td>
</tr>
<tr>
<td>$(GS, GP)$</td>
<td>8</td>
<td>55.519</td>
</tr>
</tbody>
</table>

Note that for the nominal–ordinal model $\hat{\beta}_{votech}^P = -.8784$ and $\hat{\beta}_{gen}^P = -.8614$, which suggests that the “best” scores for VoTech and General programs are much closer together than we had been assuming.
So we have now discussed,

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Likelihood</td>
</tr>
<tr>
<td>Column</td>
<td>Ratio</td>
</tr>
<tr>
<td>Nominal</td>
<td>X</td>
</tr>
<tr>
<td>Nominal</td>
<td>X</td>
</tr>
<tr>
<td>Ordinal</td>
<td>X</td>
</tr>
</tbody>
</table>

To complete our table, we need to talk about efficient score tests for testing conditional independence for each of the three cases.

The efficient score test of conditional independence of $X$ and $Y$ given $Z$ for an $I \times J \times K$ cross-classification is a generalization of the Cochran-Mantel-Haenszel statistic, which we discussed as a way to test conditional independence in $2 \times 2 \times K$ tables.

For each of three cases, the test statistic is a

**Generalized CMH Statistic**

The formulas are relatively complex (see Agresti, 1990; or SAS/FREQ manual if you want them).

The generalized CMH statistic is appropriate when the partial associations between $X$ and $Y$ are comparable for each level of $Z$ (the same is true for the LR test $G^2 [(XZ, YZ) | XY, XZ, YZ]$).
Generalized Cochran-Mantel-Haenszel Tests

**Ordinal–Ordinal.** The generalized CMH uses a generalized correlation and tests for a linear trend in the $X$–$Y$ partial association.

The null hypothesis is $H_O : \rho_{XY(k)} = 0$, and the alternative is $H_A : \rho_{XY(k)} \neq 0$.

The statistic gets large

- as the correlation increases.
- as the sample size per (partial) table increases.

When $H_O$ is true, the test statistic has an approximate chi-square distribution with $df = 1$.

It is the efficient score test for testing $H_O : \beta = 0$ in the homogeneous linear by linear association model

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

where $u_i$ are scores for the levels of variable $X$ and $v_j$ are scores for the levels of variable $Y$.

**HSB Example:** Generalized CMH statistic = 46.546, $p < .0001$.

On the handout, it is labeled “Nonzero Correlation”, which is the alternative hypothesis.
Nominal–Ordinal. Let’s suppose that $X$ (row) is nominal and $Y$ (column) is ordinal.

The responses on each row can be summarized by the mean score over the columns.

The generalized CMH test statistic for conditional independence compares the $I$ row means and is designed to detect whether the means are difference across the rows.

If the null hypothesis is true (i.e., $H_0 : \mu_{Y_j} = \mu_{Y_t}$, that is, the row means are all equal, or equivalently conditional independence between $X$ and $Y$ given $Z$), then the statistic is approximately chi-squared distributed with $df = (I - 1)$.

When the scores for $Y$ are normally distributed, a 1–way ANOVA would be the appropriate test of whether the rows have differ mean scores; that is, the nominal–ordinal generalized CMH statistic is analogous to a 1–way ANOVA.

When midranks are used as scores in the generalized CMH statistic, this is equivalent to the Kruskal–Wallis (non-parametric) test for comparing mean ranks.
Example: On the handout, the second Cochran–Mantel–Haenszel statistic labeled “Row Mean Scores Differ” corresponds to the test for conditional independence between nominal SES and ordinal program type.

In our example, it make more sense to let program type be nominal, which yields

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Alternative Hypothesis</th>
<th>df</th>
<th>Value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nonzero correlation</td>
<td>1</td>
<td>46.546</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>2</td>
<td>49.800</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>4</td>
<td>51.639</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

We can compute the mean SES scores for each program type for each gender, e.g.,

\[
\frac{[1(15) + 2(44) + 3(12)]}{(15 + 44 + 12)} = 139/71 = 1.96
\]

<table>
<thead>
<tr>
<th>Gender</th>
<th>School</th>
<th>Program</th>
<th>SES</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>females</td>
<td>VoTech</td>
<td></td>
<td>15</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td></td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td></td>
<td>16</td>
<td>70</td>
</tr>
<tr>
<td>males</td>
<td>VoTech</td>
<td></td>
<td>30</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td></td>
<td>31</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td></td>
<td>28</td>
<td>77</td>
</tr>
</tbody>
</table>

50
**Nominal–Nominal**, which is the third (and final) generalized CMH test that we need to complete of our table of tests for conditional independence for 3-way tables in the three different cases.

This CMH test statistic is a test of “general association”.

It is designed to detect any pattern or type of association that is similar across tables.

Both $X$ and $Y$ are treated as nominal variables.

The CMH test of general association is the efficient score test of $H_O: \lambda_{ij}^{XY} = 0$ in the $(XY, XZ, YZ)$ loglinear model.

If the null is true, then the statistic is approximately chi–squared distributed with $df = (I - 1)(J - 1)$.

High School & Beyond example (all CMH tests):

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Alternative Hypothesis</th>
<th>$df$</th>
<th>Value</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nonzero correlation</td>
<td>1</td>
<td>46.546</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>2</td>
<td>Row Mean Scores Differ</td>
<td>2</td>
<td>49.800</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>3</td>
<td>General Association</td>
<td>4</td>
<td>51.639</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
So we have now covered all cases and test statistics:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of Test</th>
<th>Likelihood Ratio</th>
<th>Wald</th>
<th>CMH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal Nominal</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Nominal Ordinal</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ordinal Ordinal</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

and for the curious...

Models for the SES × Program Type × Gender data

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>$G^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(GS,GP,SP)</td>
<td>4</td>
<td>1.970</td>
<td>.741</td>
</tr>
<tr>
<td>(GS,GP)</td>
<td>8</td>
<td>55.519</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(GP,SP)</td>
<td>6</td>
<td>8.532</td>
<td>.202</td>
</tr>
<tr>
<td>(SG,SP)*</td>
<td>6</td>
<td>3.312</td>
<td>.769</td>
</tr>
<tr>
<td>(GP,S)</td>
<td>10</td>
<td>62.247</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(GS,P)</td>
<td>10</td>
<td>57.027</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>(G,SP)*</td>
<td>8</td>
<td>10.040</td>
<td>.262</td>
</tr>
<tr>
<td>(G,SP)–L × L*</td>
<td>11</td>
<td>16.221</td>
<td>.133</td>
</tr>
<tr>
<td>(G,P,S)</td>
<td>12</td>
<td>63.754</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

The simplest model the appears to fit the data:

$$\log(\mu_{ijk}) = \lambda + \lambda^S_i + \lambda^P_j + \lambda^G_k + \beta u_i v_j$$
Effects of Sparse Data and Incomplete Tables Methodology

1. Types of empty cells (sampling and structural zeros).
2. Effects of sampling zeros and strategies for dealing with them.
3. Fitting models to tables with structural zeros.

A “Sparse” table is one where there are “many” cells with “small” counts.

How many is “many” and how small is “small” are relative. We need to consider both

- The sample size \( n \) (i.e., the total number of observations).
- The size of the table \( N \) (i.e., how many cells there are).
Types of Empty Cells:

1. **Sampling Zeros** are ones where you just do not have an observation for the cell; that is, $n_{ij} = 0$.

   In principle if you increase your sample size $n$, you might get $n_{ij} > 0$.

   $$ P(\text{getting an observation in a cell}) > 0 $$

2. **Structural Zeros** are cells that are theoretically impossible to observe a value.

   $$ P(\text{getting an observation in a cell}) = 0 $$

Tables with structural zeros are “*structurally incomplete***”.

This is different from a “*partial classification***” where an incomplete table results from not being able to completely cross-classify all individuals.
Example of a **partial classification**: Data from a study conducted by the College of Pharmacy at the Univ of Florida (Agresti, 1990) where elderly individuals were asked whether they took tranquilizers. Some of the subjects were interviewed in 1979, some were interviewed in 1985, and some were interviewed in both 1979 and 1985.

<table>
<thead>
<tr>
<th>1975</th>
<th>1985 yes</th>
<th>1985 no</th>
<th>sampled</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes 175</td>
<td>190</td>
<td>230</td>
<td>595</td>
<td></td>
</tr>
<tr>
<td>no 139</td>
<td>1518</td>
<td>982</td>
<td>2639</td>
<td></td>
</tr>
<tr>
<td>not sampled 64</td>
<td>595</td>
<td>—</td>
<td>659</td>
<td></td>
</tr>
<tr>
<td>total 378</td>
<td>2303</td>
<td>1212</td>
<td>3893</td>
<td></td>
</tr>
</tbody>
</table>

Example of a **structurally incomplete table**: Survey of teenagers regarding their health concerns (Fienberg):

<table>
<thead>
<tr>
<th>Health Concern</th>
<th>Gender</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex/Reproduction</td>
<td>6 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Menstrual problems</td>
<td>— 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How healthy am I?</td>
<td>49 29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>77 102</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The probability of a male with menstrual problems = 0.
It is important to recognize that a table is incomplete and why it is incomplete, because this has implications for how you deal with the incompleteness. If you have structural zeros or an incomplete classifications you should **not**

1. Fill in cells with zeros

2. Collapse the tables until the structurally empty cells “disappear”.

3. Abandon the analysis.
Effects of Sparse Data
(sampling zeros)

We’ll talk about

1. Problems that can be encountered when modeling sparse tables.
2. The effect of spareseness on hypothesis testing.

Problems in modeling Sparse Tables.
There are two major ones

1. Maximum likelihood estimates of loglinear/logit models may not exist.
2. If MLE estimates exist, they could be very biased.

Non-existence of MLE estimates. Depending on what effects are included in a model and the pattern of the sampling zeros determines whether non-zero and finite estimates of odds ratios exist.

When \( n_{ij} > 0 \) for all cells, MLE estimates of parameters are finite.

When a table has a 0 marginal frequency and there is a term in the model corresponding to that margin, MLE estimates of the parameter are infinite.
Hypothetical example (from Wickens, 1989):

\[ Z = 1 \quad Z = 2 \quad Z = 3 \]
\[
\begin{array}{c|cccc}
Y = & 1 & 2 & 3 & 4 \\
\hline
X = 1 & 5 & 0 & 7 & 8 \\
X = 2 & 10 & 0 & 6 & 7 \\
\end{array}
\]

\[
\begin{array}{c|cccc}
Y = & 1 & 2 & 3 & 4 \\
\hline
X = 1 & 9 & 8 & 3 & 12 \\
X = 2 & 8 & 3 & 0 & 9 \\
\end{array}
\]

\[
\begin{array}{c|cccc}
Y = & 1 & 2 & 3 & 4 \\
\hline
X = 1 & 6 & 3 & 5 & 11 \\
X = 2 & 0 & 2 & 8 & 11 \\
\end{array}
\]

The 1–way margins of this 3–way table:

\[
\begin{array}{c|c|c|c}
X & Y & Z & \\
\hline
1 & 77 & 64 & \\
2 & 38 & 16 & 29 & 58 \\
\end{array}
\]

\[
\begin{array}{c|c|c|c}
X & Y & Z & \\
\hline
1 & 43 & 52 & 46 \\
\end{array}
\]

The 2–way margins:

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
Y & & & & & & & & & & & & \\
\hline
X = 1 & 1 & 2 & 3 & 4 & & & & & & & & \\
X = 2 & & & & & 18 & 5 & 14 & 27 & & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
Z & & & & & & & & & & & & \\
\hline
X = 1 & 1 & 2 & 3 & & & & & & & & & \\
X = 2 & & & & 23 & 20 & 21 & & & & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
Y & & & & & & & & & & & & \\
\hline
X = 1 & 20 & 11 & 15 & 31 & & & & & & & & \\
X = 2 & 18 & 5 & 14 & 27 & & & & & & & & \\
\end{array}
\]

Since \( n_{+21} = 0 \), any \( YZ \) partial odds ratios involving this cell equal 0 or \(+\infty\).

Since the \( YZ \) margin has a zero, there is no MLE estimate of \( \lambda_{21}^{YZ} \).

You cannot fit any model with \( \lambda_{jk}^{YZ} \) terms because \( n_{+21} = 0 \).

Suppose that \( n_{121} > 0 \), could you fit \((XY, YZ)\)?

Could you fit the saturated model \((XYZ)\)?
Signs of a problem:

- The iterative algorithm that the computer used to compute MLE of a model do not converge.

In SAS/GENMOD, in the log file you find the following WARNING

The negative of the Hessian is not positive definite. The convergence is questionable.

The procedure is continuing but the validity of the model fit is questionable.

The specified model did not converge

- The estimated standard errors of parameters and fitted counts are huge relative to the rest. They “blow up”.

For example, when the \((X, YZ)\) joint independence model is fit to the hypothetical table using SAS/GENMOD,

\[ \hat{\lambda}_{21}^{YZ} = -23.9833, \quad \text{ASE} = 87,417.4434 \]

while all other ASE’s are less than .70.

\[ \hat{\mu}_{121} = 7.15 \times 10^{-11}, \quad \log(\hat{\mu}_{121}) = -23.3519, \quad \text{std err} = 87,417 \]

\[ \hat{\mu}_{221} = 5.94 \times 10^{-11}, \quad \log(\hat{\mu}_{221}) = -23.3468, \quad \text{std err} = 87,417 \]

etc
**Severe bias in odds ratio estimation.** Sparseness can cause odds ratio estimates to be severely biased and the sampling distribution of fit statistics will be poorly approximated by the chi–squared distribution.

Solution: add .5 to each cell in the table.

Why?

Adding .5 shrinks the estimated odds ratios that are \( \infty \) to finite values and increases estimates that are 0.

Qualifications: For unstaturated models, adding .5 will over smooth the data.

Remedies/Strategies/Comments:

- An infinite estimate of a model parameter maybe OK, but an infinite estimate of a true odds ratio is “unsatisfactory”.

- When a model does not converge, try adding a tiny number (e.g., \( 1 \times 10^{-8} \)) to all cells in the table.

- Do a sensitivity analysis by adding different numbers of varying sizes to the cells (e.g., \( 1 \times 10^{-8}, 1 \times 10^{-5}, .01, .1 \)). Examine fit statistics and parameter estimates to see if they change very much.
Example using hypothetical data and the \((X, Y Z)\) loglinear model:

<table>
<thead>
<tr>
<th>Number added</th>
<th>(G^2)</th>
<th>(X^2)</th>
<th>Converge?</th>
<th>(\text{ASE for } \hat{\lambda}_{21}^{YZ})</th>
</tr>
</thead>
<tbody>
<tr>
<td>—</td>
<td>16.86</td>
<td>13.38</td>
<td>no</td>
<td>87,417.44</td>
</tr>
<tr>
<td>0.000000001</td>
<td>15.43</td>
<td>17.92</td>
<td>yes</td>
<td>7,071.07</td>
</tr>
<tr>
<td>0.000001</td>
<td>16.83</td>
<td>13.37</td>
<td>yes</td>
<td>223.61</td>
</tr>
<tr>
<td>0.0001</td>
<td>16.87</td>
<td>13.38</td>
<td>yes</td>
<td>22.37</td>
</tr>
<tr>
<td>0.1</td>
<td>18.86</td>
<td>13.78</td>
<td>yes</td>
<td>2.30</td>
</tr>
</tbody>
</table>

- Use an alternative estimation procedure (i.e., Bayesian).

**Effect of Sparseness on \(X^2\) and \(G^2\).**

Guidelines

1. When \(df > 1\), it is “permissible” to have the \(\hat{\mu}\) as small as 1 so long as less than 20% of the cells have \(\hat{\mu} < 5\).

2. The permissible size of \(\hat{\mu}\) decreases as the size of the table \(N\) increases.

3. The chi–squared distribution of \(X^2\) and \(G^2\) can be poor for sparse tables with both very small and very large \(\hat{\mu}\)’s (relative to \(n/N\)).

4. No single rule covers all situations.

5. \(X^2\) tends to be valid with smaller \(n\) and sparser tables than \(G^2\).
6. $G^2$ usually is poorly approximated by the chi-squared distribution when $n/N < 5$.

The $p$–values for $G^2$ may be too large or too small (it depends on $n/N$).

7. For fixed $n$ and $N$, chi-squared approximations are better for smaller $df$ than for larger $df$.

So while the $G^2$ for model fit may not be well approximated by the chi-squared distribution, the difference between $G^2$’s for two nested models will be closer to chi–squared than the $G^2$ for fit of either model.

Chi-squared comparison tests depend more on the size of marginals than on cell sizes in the joint table.

So if margins have cells $> 5$, the chi-squared approximation of $G^2(M_O) - G^2(M_1)$ should be reasonable.

8. Exact tests and exact analyses for models.
9. An alternative test statistic: the **Cressie-Read statistic**


They proposed a family of statistics of the form

\[
RC^2 = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^{N} n_i \left[ \left( \frac{n_i}{\hat{\mu}_i} \right)^\lambda - 1 \right]
\]

where \(-\infty < \lambda < \infty\).

The value of \(\lambda\) defines a specific statistic (note: \(\lambda\) here is not a parameter of the loglinear model).

For

\(\lambda = 1, \ RC^2 = X^2.\)
\(\lambda \to 0, \ RC^2 = G^2.\)
\(\lambda = 2/3\) works pretty well for sparse data. The sampling distribution of \(RC^2\) is approximately chi-square.
Modeling Incomplete Tables

While structural zeros (and partial cross-classifications) are not as common as sampling zeros, there are a number of uses of the methodology for structurally incomplete tables:

1. Dealing with anomalous cells.
2. Excluding “problem” sampling zeros from an analysis.
3. Check collapsibility across categories of a variable.
8. (Guttman) scaling of response patterns.
9. Estimate missing cells.
10. Estimation of population size.
11. Other.

We’ve discuss 1, 2, and 3 now, and later 4, 5 and 6. (For the others, check Fienberg text and/or Wickens texts.)
The Methodology

We remove the cell(s) from the model building and analysis by only fitting models to cells with non-structural zeros.

We can arbitrarily fill in any number for a structural zero, generally we just put in 0.

To remove the \((i, j)\) cell from the modeling, an indicator variable is created for it,

\[
I(i, j) = \begin{cases} 
1 & \text{if cell is the structural zero} \\
0 & \text{for all other cells} 
\end{cases}
\]

When this indicator is included in a loglinear model as a (numerical) explanatory variable, a single parameter is estimated for the structural zero, which used up 1 \(df\), and the cell is fit perfectly. Since structural zeros are fit perfectly, they have 0 weight in the fit statistics \(X^2\) and \(G^2\).

For example, consider the teens and health concerns data.

<table>
<thead>
<tr>
<th>Health Concern</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td>Sex/Reproduction</td>
<td>6</td>
</tr>
<tr>
<td>Menstrual problems</td>
<td>—</td>
</tr>
<tr>
<td>How healthy am I?</td>
<td>49</td>
</tr>
<tr>
<td>None</td>
<td>77</td>
</tr>
</tbody>
</table>

We can express the saturated loglinear model as
\[
\log(\mu_{ij}) = 0 \quad \text{for the (2,1) cell}
\]
\[
= \lambda + \lambda_i^H + \lambda_j^G + \lambda_{ij}^{HG} \quad \text{for the rest}
\]

or equivalently we define an indicator variable

\[
I(2,1) = 1 \quad \text{for the (2,1) cell}
\]
\[
= 0 \quad \text{otherwise}
\]

A single equation for the saturated loglinear model is

\[
\log(\mu_{ij}) = \lambda + \lambda_i^H + \lambda_j^G + \lambda_{ij}^{HG} + \delta_{21} I(2,1)
\]

The \(\delta_{21}\) is a parameter that will equal whatever it needs to equal such that the (2,1) cell is fit perfectly (i.e., the fitted value will be exactly equal to whatever arbitrary constant you filled in for it).

For the independence model, we just delete the \(\lambda_{ij}^{HG}\) term from the model, but we still include the indicator variable for the (2,1) cell.

What happens to degrees of freedom?

\[
\text{df} = (# \text{ of cells}) - (# \text{ nonredundant parameters})
\]

or

\[
\text{df} = (\text{usual df for the model}) - (# \text{ cells fit perfectly})
\]
Example: Independence between teen health concerns and gender:

\[
df = (I - 1)(J - 1) - 1 \\
= (4 - 1)(2 - 1) - 1 \\
= 3 - 1 \\
= 2
\]

\[G^2 = 12.60, \text{ and } X^2 = 12.39, \text{ which provide evidence that health concerns and gender are not independent.}\]

When \(n_{21}\) is set equal to 0, the estimated parameters for the independence model are

\[\hat{\lambda} = 4.5466\]

\[\hat{\lambda}_1^H = -2.0963, \quad \hat{\lambda}_1^G = -1.1076\]

\[\hat{\lambda}_2^H = -2.0671, \quad \hat{\lambda}_1^G = 0.0000\]

\[\hat{\lambda}_3^H = -0.8307\]

\[\hat{\lambda}_4^H = 0.0000, \quad \hat{\delta}_{21} = -22.9986\]

(2, 1) cell,

\[\hat{\mu}_{21} = \exp(4.5466 - 2.0671 - 1.1076 - 22.9986) \sim 0\]
Anomalous Cells.

Suppose that a model fits a table well, except for one or just a few cells. The methodology for incomplete tables can be used to show that except for these cells, the model fits.

Of course, you would then need to talk about the anomalous cells (e.g., speculate why they’re not being fit well).

Example (from Fienberg, original source Duncan, 1975): Mothers of children under the age of 19 were asked whether boys, girls, or both should be required to shovel snow off sidewalks. The responses were cross-classified according to the year in which the question was asked (1953, 1971) and the religion of the mother.

Since none of the mothers said just girls, there are only 2 responses (boys, both girls and boys).

<table>
<thead>
<tr>
<th>Religion</th>
<th>1953 Boys</th>
<th>1953 Both</th>
<th>1971 Boys</th>
<th>1971 Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>104</td>
<td>42</td>
<td>165</td>
<td>142</td>
</tr>
<tr>
<td>Catholic</td>
<td>65</td>
<td>44</td>
<td>100</td>
<td>130</td>
</tr>
<tr>
<td>Jewish</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>13</td>
<td>6</td>
<td>32</td>
<td>23</td>
</tr>
</tbody>
</table>
Since gender \((G)\) is the response/outcome variable and year \((Y)\) and religion \((R)\) are explanatory variables, all models should include \(\lambda_{ij}^{RY}\) terms in the loglinear model.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>(G^2)</th>
<th>(p)</th>
<th>(X^2)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((RY,G))</td>
<td>7</td>
<td>31.67</td>
<td>&lt; .001</td>
<td>31.06</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>((RY,GY))</td>
<td>6</td>
<td>11.25</td>
<td>.08</td>
<td>11.25</td>
<td>.08</td>
</tr>
<tr>
<td>((RY,GR))</td>
<td>4</td>
<td>21.49</td>
<td>&lt; .001</td>
<td>21.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>((RY,GY,GR))</td>
<td>3</td>
<td>0.36</td>
<td>.95</td>
<td>.36</td>
<td>.95</td>
</tr>
</tbody>
</table>

- The homogeneous association model fits well.
- The \((RY,GY)\) model fits much better than independence, but fits significantly worse than \((RY,GY,GR)\):

\[
G^2[\text{(RY, GY)} | \text{(RY, GY, GR)}] = 11.25 - .36 = 10.89
\]

with \(df = 3\) and \(p = .01\).

However, let’s take a closer look at \((RY,GY)\).

The Pearson residuals from the \((RY,GY)\) loglinear model

<table>
<thead>
<tr>
<th>Religion</th>
<th>1953 Boys</th>
<th>1953 Both</th>
<th>1971 Boys</th>
<th>1971 Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>.75</td>
<td>-1.05</td>
<td>.91</td>
<td>-.91</td>
</tr>
<tr>
<td>Catholic</td>
<td>-.84</td>
<td>1.18</td>
<td>-1.42</td>
<td>1.42</td>
</tr>
<tr>
<td>Jewish</td>
<td>-.29</td>
<td>.41</td>
<td>-.22</td>
<td>.22</td>
</tr>
<tr>
<td>Other</td>
<td>.12</td>
<td>-.17</td>
<td>.85</td>
<td>-.85</td>
</tr>
</tbody>
</table>

The 3 largest residuals correspond to mothers who are Catholic. The model under predicts “both” and overpredicts “boys”.

69
• Question: If we do not include Catholic mothers, would the model \((RY,GY)\) or the logit model with just a main effect of year fit the data?

Try the model that removes the 3 largest residuals (the 2nd row of the table)

\[
\log(\mu_{ijk}) = \lambda + \lambda_i^R + \lambda_j^Y + \lambda_k^G + \lambda_{ij}^{RY} + \lambda_{jk}^{GY} + \delta_{212} I(2, 1, 2) + \delta_{221} I(2, 2, 1) + \delta_{222} I(2, 2, 2)
\]

Where the indicator variables are defined as

\[
I(2, j, k) = 1 \quad \text{if Catholic and } j \neq 1 \text{ and } k \neq 1
\]
\[
= 0 \quad \text{otherwise}
\]

We only need 3 indicators to “remove” the row for Catholic mothers.

Why?

This model has \(df = 4\), \(G^2 = 1.35\), and \(X^2 = 1.39\). So the \((RY,GY)\) model fits well when we disregard the second row of the table.
Example using logit models.
Data from Farmer, Rotella, Anderson & Wardrop (1998).

Individuals from a longitudinal study who had chosen a career in science was cross-classified according to their gender and the primary Holland code describing the type of career in science that they had chosen.

Since interest was in testing whether women and men differed and if so describing the differences, will treat gender as a response variable.

<table>
<thead>
<tr>
<th>Holland Code</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
</tr>
<tr>
<td>Realistic</td>
<td>13</td>
</tr>
<tr>
<td>Investigative</td>
<td>31</td>
</tr>
<tr>
<td>Artistic</td>
<td>2</td>
</tr>
<tr>
<td>Social</td>
<td>1</td>
</tr>
<tr>
<td>Enterprising</td>
<td>2</td>
</tr>
<tr>
<td>Conventional</td>
<td>3</td>
</tr>
</tbody>
</table>

The logit model corresponding to the (H,G) loglinear model,

\[
\text{logit}(\pi_w) = \log(\pi_{women}/\pi_{men}) = \alpha
\]

has \( df = 5, G^2 = 42.12, \) and \( p < .001. \)

Based on previous research, it was expected that men would tend to choose jobs with primary code realistic and women primary code being social, and this is what was found in the residuals,
Adjusted residuals

<table>
<thead>
<tr>
<th>Holland Code</th>
<th>Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realistic</td>
<td>-3.76</td>
</tr>
<tr>
<td>Investigative</td>
<td>-2.15</td>
</tr>
<tr>
<td>Artistic</td>
<td>-0.16</td>
</tr>
<tr>
<td>Social</td>
<td>4.78</td>
</tr>
<tr>
<td>Enterprising</td>
<td>-0.73</td>
</tr>
<tr>
<td>Conventional</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Since the 2 largest residuals correspond to Realistic and Social, we’ll fit them perfectly but allow independence in the rest of the table,

\[
\text{logit}(\pi_w) = \alpha + \delta^R I_R(i) + \delta^S I_S(i)
\]

where

\[
I_R(i) = \begin{cases} 
1 & \text{if code is Realistic} \\
0 & \text{otherwise} 
\end{cases}
\]

and

\[
I_S(i) = \begin{cases} 
1 & \text{if code is Social} \\
0 & \text{otherwise} 
\end{cases}
\]

This model has \(df = 3\), \(G^2 = 4.32\), \(p = .23\), and fits pretty good.

We can do better.

Note that the residuals from the independence models for realistic and social are both quite large but opposite signs.
Let’s define a new variable to capture the suspected association structure,

\[ I(i) = \begin{array}{c}
-1 & \text{if code is Realistic} \\
1 & \text{if code is Social} \\
0 & \text{otherwise}
\end{array} \]

and fit the model

\[ \logit(\pi_w) = \alpha + \beta I(i) \]

This model has \( df = 4, G^2 = 4.54, p = .24 \). This fits almost as good as the model in which the odds for realistic and social are fit perfectly:

\[ \Delta G^2 = 4.54 - 4.32 = .22 \]

with \( \Delta df = 4 - 3 = 1 \), which is the likelihood ratio test of \( H_0 : \beta = 0 \). Furthermore, the residuals look pretty good

### Adjusted residuals

<table>
<thead>
<tr>
<th>Holland Code</th>
<th>Independence</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Realistic</td>
<td>-3.76</td>
<td>.37</td>
</tr>
<tr>
<td>Investigative</td>
<td>-2.15</td>
<td>-.86</td>
</tr>
<tr>
<td>Artistics</td>
<td>-.16</td>
<td>.02</td>
</tr>
<tr>
<td>Social</td>
<td>4.78</td>
<td>.27</td>
</tr>
<tr>
<td>Enterprising</td>
<td>-.73</td>
<td>-.56</td>
</tr>
<tr>
<td>Conventional</td>
<td>1.55</td>
<td>1.77</td>
</tr>
</tbody>
</table>
Interpretation: \( \hat{\beta} = 2.9240 \) with ASE= .7290.

- Gender and codes are independent, except for the codes other than Realistic and Social.

- The odds that a woman (versus a man) with a science career has a primary code of Social is

\[
\exp \left( \hat{\beta}(1 - (-1)) \right) = \exp(2(2.9240)) = e^{5.848} = 346.54
\]
times the odds that the career has a primary code of Realistic.

- The odds ratio of Social versus Other than Realistic equals

\[
\exp \left( \hat{\beta}(1 - 0) \right) = \exp(2.9240) = 18.62
\]

- The odds ratio of Realistic versus Other than Social equals

\[
\exp \left( \hat{\beta}(0 - 1) \right) = \exp(-2.9240) = 1/18.62 = .05
\]
Collapsing Over Categories

Returning to the snow shoveling data, rather than deleting Catholics, perhaps the effect of religion on the response can be accounted for by a single religious category. If so, then we can collapse the religion variable and get a more parsimonious and compact summary of the data.

To investigate this, we replace religion by a series of 4 binary variables

\[ P = \text{Protestant} \ (i.e., P = 1 \text{ if Protestant, 0 otherwise}). \]

\[ C = \text{Catholic} \ (i.e., C = 1 \text{ if Catholic, 0 otherwise}). \]

\[ J = \text{Jewish} \ (i.e., J = 1 \text{ if Jewish, 0 otherwise}). \]

\[ O = \text{Other} \ (i.e., O = 1 \text{ if not } P, C, \text{ or } J, \text{ and 0 otherwise}). \]

Using all 4 variables (instead of just 3), we introduce redundancy in the data. This allows us to treat the 4 categories of religion symmetrically.
New display of the data in the form of a 6-way, incomplete table

<table>
<thead>
<tr>
<th>Protestant</th>
<th>Catholic</th>
<th>Jewish</th>
<th>Other</th>
<th>1953</th>
<th>Both</th>
<th>1971</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>104</td>
<td>42</td>
<td>165</td>
<td>142</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>44</td>
<td>100</td>
<td>130</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>6</td>
<td>32</td>
<td>23</td>
</tr>
</tbody>
</table>

Since G (gender) is considered the response and the 4 religion variables and year are all considered explanatory variables, all loglinear models must include $\lambda_{YPJCJO}$ terms (and lower order ones). This reduces the set models that we need to consider.

Note:

- $P =$ Protestant (i.e., $P = 1$ if Protestant, 0 otherwise).
- $C =$ Catholic (i.e., $C = 1$ if Catholic, 0 otherwise).
- $J =$ Jewish (i.e., $J = 1$ if Jewish, 0 otherwise).
- $O =$ Other (i.e., $O = 1$ if not $P$, $C$, or $J$, and 0 otherwise).
\( P \) = Protestant (i.e., \( P = 1 \) if Protestant, 0 otherwise).

\( C \) = Catholic (i.e., \( C = 1 \) if Catholic, 0 otherwise).

\( J \) = Jewish (i.e., \( J = 1 \) if Jewish, 0 otherwise).

\( O \) = Other (i.e., \( O = 1 \) if not \( P \), \( C \), or \( J \), and 0 otherwise).

Here are some of the fit to the data models.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>( G^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit previously</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (YPCJO,GY) )</td>
<td>6</td>
<td>11.2</td>
</tr>
<tr>
<td>( (YPCJO,GY,GPCJO) )</td>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>New ones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (YPCJO,GY,GO) )</td>
<td>5</td>
<td>9.8</td>
</tr>
<tr>
<td>( (YPCJO,GY,GJ) )</td>
<td>5</td>
<td>10.9</td>
</tr>
<tr>
<td>( (YPCJO,GY,GC) )</td>
<td>5</td>
<td>1.4</td>
</tr>
<tr>
<td>( (YPCJO,GY,GP) )</td>
<td>5</td>
<td>4.8</td>
</tr>
</tbody>
</table>

The \( (YPCJO,GY,GC) \) model which has a main effect for year (GY) and an effect of being Catholic fits well.

In other words, the interaction between religion and response is due primarily to Catholic mothers.

In this example, we can collapse religion into a single dichotomous variable (Catholic, Not Catholic).