Model Building for Logit and Log-Linear Models

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

©Board of Trustees, University of Illinois

Fall 2019
Outline

- Association Graphs.
  1. Introduction.
  2. Collapsibility.
  3. Representing models.
- Modeling ordinal association.
  1. linear by linear association, (and RC(M) association model & correspondence analysis)
  2. ordinal tests of independence.
- Testing conditional independence.
- Effects of sparse data.
- Model fitting details.
- A hybrid model (log-linear with numerical predictors)
Graphical Models

- **Statistical Physics** (Gibbs, 1902). In large systems of particles, each particle occupies a site and can be in different states. The total energy of the system is composed of an external potential and a potential due to *interactions* of groups of particles. It is assumed that particles that are close to each other (i.e., they are “neighbors”) interact while those that are not close to each other do not interact.

- **Genetics & Path Analysis.** (Wright, 1921, 1923, 1934). In studying the heritability of properties of natural species, graphs were used to represent *directed relations*. Arrows point from a “parent” to a “child”. These ideas were taken up by Wold (1954) and Blalock (1971) in economics and social sciences and lead to what we know as path analysis.

Usefulness of Graphical Models

Graphical models are useful and are widely applicable because

1. Graphs visually represent scientific content of models and thus facilitate communication.

2. Graphs break down complex problems/models into smaller and simpler pieces that can be studied separately.

3. Graphs are natural data structures for digital computers.


Graphical Models & Contingency Tables

We’ll be using graphs to

1. Help determine when marginal and partial associations are the same such that we can collapse a multi-way table into a smaller table (or tables) to study certain associations.

2. Represent substantive theories and hypotheses, which correspond to certain loglinear/logit models.

Some terminology & definitions (common to all graphical models)…
Terminology & Definitions

- **Vertices** (or “nodes”) are points that represent variables.
- **Edges** are lines that connect two vertices.

The presence of an edge between two vertices indicates that an association exists between the two variables.

![Edge Diagram]

X \quad Y

The absence of an edge between two vertices indicates that the two variables are independent.

![No Edge Diagram]

X \quad Y

We will be (mostly) restricting our attention to undirected relationships, so our lines won’t have arrows on them (lines with arrows represent directed relationships).
More Terminology & Definitions

- A **Graph** consists of a set of vertices and edges.
- A **Path** is a sequence of edges that go from one variable to another.

![Graph example](image)

- **Separated.** Two variables are said to be separated if all paths between the two variables intersect a third variable (or set of variables).

![Separated example](image)
Even More Definitions

A *Clique* is a set of vertices (variables) where each variable is connected to every other variable in the set.

\[
\begin{array}{ccc}
W & \text{ } & X \\
\text{ } & \text{ } & \\
Y & \text{ } & Z
\end{array}
\]

This is also known as a “complete graph” and if this is part of a larger graph, a “complete subgraph”.

**Fundamental Result (cornerstone of graphical modeling)**

Two variables are conditionally independent given any subset of variables that separates them.
Graphs for Log-linear Models of...

The graph for the **Complete Independence**, \((X, Y, Z)\)

\[ X \]

\[ Y \]

\[ Z \]

The graphs for **Joint Independence**,

\((XY, Z)\)

\((XZ, Y)\)

\((X, YZ)\)

\[ X \]

\[ Y \]

\[ Z \]

\[ X \]

\[ Y \]

\[ Z \]

\[ X \]

\[ Y \]

\[ Z \]
Graphs for Conditional Independence

The graphs for Conditional Independence:

- \((XY, XZ)\)
- \((XY, YZ)\)
- \((XZ, YZ)\)

\[
\begin{align*}
(X, Y, Z) & \quad \text{Graph for } (XY, XZ) \\
(Y, Z) & \quad \text{Graph for } (XY, YZ) \\
(Y, Z) & \quad \text{Graph for } (XZ, YZ)
\end{align*}
\]
Graphs for 3–Way Association model

The graph for 3–Way Association model, $(XYZ)$:

This is also a graph for Homogeneous Association, $(XY, XZ, YX)$, which is also a model of dependence.
Association Graphs & Log-linear Models

- All Log-linear models have graphical representations.
- All independence log-linear models imply a unique graph, but not all dependence log-linear models have unique graphical representations.
- Each graph implies at least one log-linear model. Unless otherwise specified, the model “read” from a graph will be the most complex one.

What is the log-linear model for this graph?
More Association Graphs & Log-linear Models

What is the log-linear model for this graph?

\[
\begin{align*}
A & \quad X & \quad Y \\
S & \quad E & \quad C \\
& \quad T
\end{align*}
\]
More Association Graphs & Log-linear Models

What is the graph for this log-linear model?

\((WY, YZ, ZX)\)

Are there other log-linear models with this graphical representation?

What is the graph for this log-linear model?

\((WXY, WXZ)\)
Collapsibility in 3–Way Tables

Under certain conditions, marginal associations and partial associations are the same (i.e., the partial odds ratios equal the marginal odds ratios).

The collapsibility condition for 3–way tables is

For 3–way tables, X–Y marginal and partial odds ratios are identical if either

- Z and X are conditionally independent, or
- Z and Y are conditionally independent.

In other words,

The X–Y marginal and partial odds ratios are identical if either the

- Log-linear model \((XY, ZY)\) holds, or
- Log-linear model \((XY, XZ)\) holds.
Collapsibility in Graphical Terms

In terms of graphs,

The $X$-$Y$ marginal and partial odds ratios are identical if either of the following graphical models (or simpler ones) hold

- \( X \perp Y \mid Z \)
- \( X \perp Y \)

Demonstration: On the next page are the partial (conditional) odds ratios and the marginal odds ratios computed based on fitted values from various log-linear models that we fit to the blue collar worker data.
## Example of Collapsibility

Observed and fitted values from selected models:

<table>
<thead>
<tr>
<th>Manage</th>
<th>Super</th>
<th>Worker</th>
<th>$n_{ijk}$</th>
<th>$M, S, W$</th>
<th>$MS, W$</th>
<th>$MS, MW$</th>
<th>$MSW$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bad</td>
<td>low</td>
<td>low</td>
<td>103</td>
<td>50.15</td>
<td>71.78</td>
<td>97.16</td>
<td>102.26</td>
</tr>
<tr>
<td>bad</td>
<td>low</td>
<td>high</td>
<td>87</td>
<td>82.59</td>
<td>118.22</td>
<td>92.84</td>
<td>87.74</td>
</tr>
<tr>
<td>bad</td>
<td>high</td>
<td>low</td>
<td>32</td>
<td>49.59</td>
<td>27.96</td>
<td>37.84</td>
<td>32.74</td>
</tr>
<tr>
<td>bad</td>
<td>high</td>
<td>high</td>
<td>42</td>
<td>81.67</td>
<td>46.04</td>
<td>36.16</td>
<td>41.26</td>
</tr>
<tr>
<td>good</td>
<td>low</td>
<td>low</td>
<td>59</td>
<td>85.10</td>
<td>63.47</td>
<td>51.03</td>
<td>59.74</td>
</tr>
<tr>
<td>good</td>
<td>low</td>
<td>high</td>
<td>109</td>
<td>140.15</td>
<td>104.53</td>
<td>116.97</td>
<td>108.26</td>
</tr>
<tr>
<td>good</td>
<td>high</td>
<td>low</td>
<td>78</td>
<td>84.15</td>
<td>105.79</td>
<td>85.97</td>
<td>77.26</td>
</tr>
<tr>
<td>good</td>
<td>high</td>
<td>high</td>
<td>205</td>
<td>138.59</td>
<td>174.21</td>
<td>197.28</td>
<td>205.74</td>
</tr>
</tbody>
</table>

Partial and marginal odds ratios computed using fitted values.

<table>
<thead>
<tr>
<th>Model</th>
<th>Partial Odds Ratio</th>
<th>Marginal Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W-S$</td>
<td>$M-W$</td>
<td>$M-S$</td>
</tr>
<tr>
<td>$(M, S, W)$</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$(MS, W)$</td>
<td>1.00</td>
<td>2.40</td>
</tr>
<tr>
<td>$(MS, MW)$</td>
<td>1.00</td>
<td>2.11</td>
</tr>
<tr>
<td>$(MS, WS, MW)$</td>
<td>1.47</td>
<td>2.11</td>
</tr>
<tr>
<td>$(MSW)$ level 1</td>
<td>1.55</td>
<td>2.19</td>
</tr>
<tr>
<td>$(MSW)$ level 2</td>
<td>1.42</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Collapsibility & Logit Models

The collapsibility condition for log-linear models applies to logit models as well.

Example: Problem 5.14 (page 138). Data from NCAA study of graduation rates of college athletes:

<table>
<thead>
<tr>
<th>Race</th>
<th>Sex</th>
<th>Graduates</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>women</td>
<td>498</td>
<td>796</td>
</tr>
<tr>
<td>White</td>
<td>men</td>
<td>878</td>
<td>1625</td>
</tr>
<tr>
<td>Black</td>
<td>women</td>
<td>54</td>
<td>143</td>
</tr>
<tr>
<td>Black</td>
<td>men</td>
<td>197</td>
<td>660</td>
</tr>
</tbody>
</table>

The best logit model for these data is

$$\text{logit}(\pi_{ij}) = \alpha + \beta_i^R + \beta_j^S$$

Recall that $\exp(\beta_f^S - \beta_m^S)$ equals the odds ratio for graduation and gender of the athlete holding race fixed; that is,

$$\theta_{SG(i)} = \exp(\beta_f^S - \beta_m^S)$$
Collapsibility & Logit Models (continued)

The logit model \( \text{logit}(\pi_{ij}) = \alpha + \beta_i R + \beta_j S \) corresponds to the no 3–factor association log-linear models; that is, \((RS, RG, SG)\) where \(G\) = whether the student athlete graduated or not.

If the logit model

\[
\text{logit}(\pi_{ij}) = \alpha + \beta_j^S
\]

had fit, which corresponds to the \((RS, SG)\) log-linear model, then we could have studied the gender–graduation relationship by looking at the gender \(\times\) graduation marginal table.

According to the collapsibility condition, if the \((RS, SG)\) log-linear model fit, then the partial S-G odds ratio equals the marginal odds ratio; that is,

\[
\theta_{SG}(i) = \theta_{SG}
\]
Collapsibility for Multiway Tables

from Agresti

Suppose that variables in a model for a multiway table partition into three exclusive subsets, A, B, and C, such that B separates A and C; thus, the model does not contain parameters linking variables from A with variables from C. When one collapses the table over the variables in C, model parameters relating variables in A and model parameters relating variables in A with variables in B are unchanged.

Graphically, each path between variables in set A and variables in set C involve at least 1 variable in set B....
Collapsibility for Multiway Tables

Graphically, each path between variables in set $A$ and variables in set $C$ involve at least 1 variable in set $B$. 

set $A$ \set B \set C 
\begin{array}{ccc}
U & W & Y \\
V & X & Z \\
\end{array}
Example of Collapsibility & Multiway Tables

A 2nd Possibility:

And others?
A 2nd Possibility:
Using Graphs to Guide Modeling


There are two signals (i.e., vertically oriented sinusoids):

- $H$ — A high frequency one.
- $L$ — A low frequency one.

On each trial for each potential signal, subjects rated on a 1 to 6 scale whether a signal was present or not where 1 indicates they were sure that no signal was presented and 6 indicates that they were sure that a signal was presented. Each subject performed 2,000 trials where there were 500 consisting of $2 \times 2$ combinations of $H$ and $L$ signals being present or absent.
Using Graphs to Guide Modeling

There are 2 response variables:

$X$ for the rating of the $H$ signal

$Y$ for the rating of the $L$ signal.

... and there were 2 factors (conditions) were $L$ and $H$ present and/or absent.
The Data

<table>
<thead>
<tr>
<th>Low Freq</th>
<th>Absent</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>X=1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>69 6 1 1 0 0</td>
<td>10 5 2 11 16 28</td>
</tr>
<tr>
<td>2</td>
<td>34 20 10 3 1 0</td>
<td>8 5 11 43 27 38</td>
</tr>
<tr>
<td>Absent</td>
<td>43 24 13 9 1 0</td>
<td>9 6 7 28 32 45</td>
</tr>
<tr>
<td>4</td>
<td>78 40 20 6 0 1</td>
<td>8 6 14 19 23 22</td>
</tr>
<tr>
<td>5</td>
<td>32 38 17 5 4 0</td>
<td>4 5 7 6 18 18</td>
</tr>
<tr>
<td>6</td>
<td>5 14 3 2 0 0</td>
<td>0 1 2 3 5 8</td>
</tr>
<tr>
<td>Present</td>
<td>4 1 0 0 0 0</td>
<td>5 0 1 4 4 9</td>
</tr>
<tr>
<td>2</td>
<td>5 3 2 1 0 0</td>
<td>0 1 3 6 9 27</td>
</tr>
<tr>
<td>3</td>
<td>8 6 3 1 0 0</td>
<td>2 3 2 11 27 20</td>
</tr>
<tr>
<td>4</td>
<td>36 25 18 3 1 0</td>
<td>9 12 11 10 23 31</td>
</tr>
<tr>
<td>5</td>
<td>83 69 26 6 1 0</td>
<td>16 7 5 19 23 40</td>
</tr>
<tr>
<td>6</td>
<td>127 50 12 7 2 0</td>
<td>21 14 13 20 21 61</td>
</tr>
</tbody>
</table>

With four variables, there are many possible models to fit. However, we don’t need to consider all models that could be fit to the data.
Assumptions?

Concerned about the assumptions?

- Independence of observations?
- Homogeneity?
Random Responding

Since $H$ and $L$ were fixed by the experimenter (i.e., “fixed by design”), all models should include terms $\lambda^{HL}$ for the $HL$ association. The simplest model would be that a subject responds randomly

$$X \quad Y$$

The log-linear model: $(HL, X, Y)$.

$G^2 = 2265.57$, $df = 130$, $p < .01$
Detectable Signals

The subject can detect the signals & detecting one does not influence detection of the other (i.e., subject does what the experimenter asked).

\[ X \quad Y \]
\[ H \quad L \]

Log-linear model: \((HL, XH, YL)\)

This is a “base” model to which we can add more complicated forms of associations.

\[ G^2 = 375.72, \quad df = 120, \quad p < .01 \]

If one signal or the other was not detectable, then we might have another base model (e.g., \((HL, XH, Y)\) or \((HL, LY, X)\)).
**Association to the unrelated signal**

In this model, responses to one signal are influenced by whether both signals are present and/or absent (i.e., the appropriate and inappropriate signal).

![Graphical representation]

The log-linear model \((HLX, HLY)\)

\[ G^2 = 221.43, \ df = 100, \ p < .01 \]
Association to the unrelated signal

- $X$ and $Y$ are conditionally independent given $H$ and $L$.
- A more restricted alternative model that also has this graphical representation, $(HL, HX, HY, LX, LY)$.
- Since we’re only considering models that “make sense” (i.e. that are interpretable), we wouldn’t include a model such as: 

![Graphical Representation]

\[ X \quad \bullet \quad \bullet \quad \bullet \quad Y \]

\[ H \quad L \]
Response-response association

We add to the base model (detectable signals) the possibility that a response regarding one signal is related to response to the other signal.

The log-linear model: \((HL, HX, LY, XY)\).

\[ G^2 = 159.27, \quad df = 95, \quad p < .01 \]
All pairwise associations.

The log-linear model \((HL, HX, HY, LX, LY, XY)\).
\[ G^2 = 113.82, \quad df = 85, \quad p = .02 \]

It’s graphical representation is

```
  X \hspace{1cm} Y
   \hspace{1cm}
  H \hspace{1cm} L
```

This is also the representation of many other log-linear models with dependencies, including model with 4-way interaction (i.e. saturated model).

This is the most complex graph, but there are interesting log-linear models that have this representation.
Another Model

We can add three-factor terms to the all pairwise association model. Some of these all have reasonable interpretations. For example, consider the model that adds $\lambda^{HLY}$. The $\lambda^{HLY}$ terms imply that delectability of the $L$ signal (measured by $Y$) is affected by the presence of the $H$ signal.

It’s graphical representation is

```
X --- Y
|   |
H - L
```

Fit of this model $G^2 = 98.06$, $df = 80$, $p = .08$
QQ Plot for (HLY, HY, LY, XY)
Example from Agresti, 1990; The variables:

- **G** for gender.
- **PMS** for premarital sex.
- **EMS** for extra marital sex.
- **M** for marital status (divorced, still married).

We’ll depart somewhat from the graphical models that we’ve discussed so far and talk about **directed relationships**.
Eg: 4-way Table with Time Ordering

The point in time at which values of variables were determined:

G PMS EMS M

Any variable to the right of others could be a response & those left of it explanatory.
Breaking down the analysis

We could analyze these data in three stages:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Response</th>
<th>Explanatory</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>PMS</td>
<td>Gender</td>
</tr>
<tr>
<td>(2)</td>
<td>EMS</td>
<td>Gender, PMS</td>
</tr>
<tr>
<td>(3)</td>
<td>M</td>
<td>Gender, PMS, EMS</td>
</tr>
</tbody>
</table>

To further guide the modeling consider the following figure, which might have been hypothesized as the existing causal structure for the variables.

![Causal Diagram]

G ➞ P ➞ M ➞ E
Stage 1

PMS is the response & G explanatory.

\[ G^2 [(G, P)] = 75.26, \text{ df } = 1, \text{ and } p < .0001. \]

Sample (marginal) odds ratio \( \hat{\theta}_{GP} = .27 \) (or \( 1/.27 = 3.70 \)).
Stage 2

Stage 2: EMS is the response and G & PMS are possible explanatory variables.

<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>$(GP, E)$</td>
<td>3</td>
<td>48.92</td>
<td>&lt; .001</td>
<td>56.77</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$(GP, PE)$</td>
<td>2</td>
<td>2.91</td>
<td>.23</td>
<td>2.95</td>
<td>.23</td>
</tr>
<tr>
<td>$(GP, GE, PE)$</td>
<td>1</td>
<td>.00</td>
<td>.98</td>
<td>.00</td>
<td>.98</td>
</tr>
</tbody>
</table>

a. Value = .0008.

Loglinear model $(GP, PE)$ fits pretty well.
The estimated $P$-$E$ odds ratio $\hat{\theta}_{EP} = 3.99$.
The marginal odds ratio is also equal to 3.99, and the reason why can be seen by looking at the figure for the model that fit:
Last Stage

Stage 3: M is the response, G, PMS and EMS are explanatory variables.

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>$G^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(EGP, EM, PM)$</td>
<td>5</td>
<td>18.16</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>$(EGP, EMP)$</td>
<td>4</td>
<td>5.25</td>
<td>.26</td>
</tr>
<tr>
<td>$(EGP, EMP, GM)$</td>
<td>3</td>
<td>.70</td>
<td>.88</td>
</tr>
</tbody>
</table>

- $(EGP, EM, PM)$ corresponds to the original figure.
- $(EGP, EMP)$ adds an interaction between EMS and PMS with respect to M, marital status.
- $(EGP, EMP, GM)$ adds a main effect for Gender with respect to predicting M.
- $(EGP, EMP)$ and $(EGP, EMP, GM)$ are more complex than implied by original figure.
Modeling Ordinal Relationships in 2–Way Tables

- Loglinear models for contingency tables treat all variables as nominal variables.
- If there is an ordering of the categories of the variables, this is not taken into account.
- Could rearrange the rows and/or columns of a table and we would get the same fitted odds ratios for the data as we would given the ordinal ordering of the rows and/or columns.
In between independence & saturated models

High School and Beyond: Consider **Program type** (Vocational/technical, general and academic) and **SES** (low, middle, high).

<table>
<thead>
<tr>
<th>SES</th>
<th>Program Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vo/Tech</td>
</tr>
<tr>
<td>Low</td>
<td>45</td>
</tr>
<tr>
<td>Middle</td>
<td>82</td>
</tr>
<tr>
<td>High</td>
<td>20</td>
</tr>
</tbody>
</table>

For the SES × Program type data, if the two variables are independent, then we have

\[
\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P
\]

\[G^2 = 53.72, \ df = 4, \ p < .001\], which leaves us with the saturated model.
In between independence & saturated models

\[
\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P
\]

\[
\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_{ij}^{SP}
\]

We can use ordering of SES levels and assign scores to them and we’ll guess at the ordering of the program types, which we can use our model.

Given scores for the rows \(\{u_1 \leq u_2 \ldots \leq u_I\}\) and scores for the columns \(\{v_1 \leq v_2 \leq \ldots \leq v_J\}\), then we can model the dependency between the variables:

\[
\log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \beta u_i v_j
\]

This only requires 1 extra parameter (i.e., model \(df = 3\)).

This model is know as the “linear by linear association model”.
Log Linear by Linear Association Model

\[ \log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \beta u_i v_j \]

- It’s called the “linear by linear association model,” because...
  
  For each row \( i \), the association is a linear function of the columns,

  \[ \lambda_{ij}^{SP} = (\beta u_i) v_j \]

  For each column \( j \), the association is a linear function of the rows.

  \[ \lambda_{ij}^{SP} = (\beta v_j) u_i \]
Log Linear by Linear Association Model (continued)

\[ \log(\mu_{ij}) = \lambda + \lambda_i^S + \lambda_j^P + \beta u_i v_j \]

- Only has 1 more parameter than the independence model (i.e., \( \beta \)), so it is “in between” independence and the saturated models.
- If \( \beta > 0 \), then \( X \) and \( Y \) are positively associated (i.e., \( X \) tends to go up as \( Y \) goes up).
- If \( \beta < 0 \), the \( X \) and \( Y \) are negatively associated.
Linear by Linear Association Model (continued)

- The odds ratio for any $2 \times 2$ sub-table is a direct function of the row and column scores and $\beta$.

$$\log \left( \frac{\mu_{ij} \mu_{i'j'}}{\mu_{i'j} \mu_{ij'}} \right) = \log(\mu_{ij}) + \log(\mu_{i'j'}) - \log(\mu_{i'j}) - \log(\mu_{ij'})$$

$$= \beta(u_i v_j + u_{i'} v_{j'} - u_{i'} v_j - u_i v_{j'})$$

$$= \beta(u_i - u_{i'})(v_j - v_{j'})$$

The strongest associations occur in the extreme corners of the table (largest differences between scores).
The smallest associations occur for rows and columns that have scores that are more nearly equal.
**Example of linear by linear model**

For the high school data example, it seems reasonable to assign equally spaced scores for the levels of SES:

\[ u_1 = 1, \quad u_2 = 2, \quad u_3 = 3 \]

For the program types, it seems reasonable to order them as:

\[ \text{Vo/Tech} \leq \text{General} \leq \text{Academic} \]

Guess that Vo/Tech and General should be closer together than are General and Academic; therefore, let’s try

\[ v_1 = 1, \quad v_2 = 2 \quad v_3 = 4 \]

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>( G^2 )</th>
<th>( p )</th>
<th>( \Delta df )</th>
<th>( \Delta G^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>4</td>
<td>53.715</td>
<td>&lt; .001</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>L by L</td>
<td>3</td>
<td>5.980</td>
<td>.10</td>
<td>1</td>
<td>47.74</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
Estimated Parameters & Odds Ratios

\[ \hat{\beta} = .32 \quad \text{and} \quad \exp(.32) = 1.38, \]

The odds ratio for a unit change in row and column scores equals 1.38 (e.g., odds ratio for low–middle SES and vo/tech–academic subtable).

The extreme corners of our table, which correspond to the low & high SES levels and program types vo/tech & academic:

\[ \hat{\theta} = \exp[.3214(3 - 1)(4 - 1)] = \exp(.3214(6)) = 6.88 \]

The odds of attending an academic versus a vo/tech program if you’re high SES is 6.88 times the odds if you’re low SES.
SAS/GENMOD and Fitting the L by L model

DATA hsb;
    input ses $ hsp $ count u v ;
datalines;
low    general    50 1 2
low    academic   44 1 4
low    votech     45 1 1
mid    general    70 2 2
mid    academic   147 2 4
mid    votech     82 2 1
hi     general    25 3 2
hi     academic   117 3 4
hi     votech     20 3 1
PROC GENMOD data=hsb;
    class ses hsp;
    model count = ses hsp u*v / link=log dist=poi;
    title 'Log Linear x Linear Association Model';
**SAS/GENMOD and Fitting the L by L model**

Linear x Linear Association Model

The GENMOD Procedure

Model Information

- **Data Set**: WORK.HSB
- **Distribution**: Poisson
- **Link Function**: Log
- **Dependent Variable**: count

Number of Observations Read: 9
Number of Observations Used: 9

Class Level Information

- **ses**: 3 levels: hi, low, mid
- **hsp**: 3 levels: academic, general, votech

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>3</td>
<td>5.9798</td>
<td>1.9933</td>
</tr>
<tr>
<td>Scaled Deviance</td>
<td>3</td>
<td>5.9798</td>
<td>1.9933</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>3</td>
<td>5.6845</td>
<td>1.8948</td>
</tr>
<tr>
<td>Scaled Pearson X2</td>
<td>3</td>
<td>5.6845</td>
<td>1.8948</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td></td>
<td>2020.3156</td>
<td></td>
</tr>
</tbody>
</table>

Algorithm converged.
### SAS and Fitting the L by L model

**Linear x Linear Association Model**

**The GENMOD Procedure**

**Analysis Of Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>3.04</td>
<td>0.21</td>
<td>2.63 3.45</td>
<td>216.20</td>
</tr>
<tr>
<td>ses hi</td>
<td>1</td>
<td>-1.59</td>
<td>0.19</td>
<td>-1.95 -1.22</td>
<td>72.06</td>
</tr>
<tr>
<td>ses low</td>
<td>1</td>
<td>0.04</td>
<td>0.15</td>
<td>-0.26 0.34</td>
<td>0.07</td>
</tr>
<tr>
<td>ses mid</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00 0.00</td>
<td>.</td>
</tr>
<tr>
<td>hsp academic</td>
<td>1</td>
<td>-0.59</td>
<td>0.23</td>
<td>-1.04 -0.14</td>
<td>6.72</td>
</tr>
<tr>
<td>hsp general</td>
<td>1</td>
<td>0.58</td>
<td>0.14</td>
<td>0.30 0.86</td>
<td>16.44</td>
</tr>
<tr>
<td>hsp votech</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00 0.00</td>
<td>.</td>
</tr>
<tr>
<td>u*v</td>
<td>1</td>
<td>0.32</td>
<td>0.05</td>
<td>0.23 0.42</td>
<td>43.71</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00 1.00</td>
<td>.</td>
</tr>
</tbody>
</table>

**NOTE:** The scale parameter was held fixed.
R and Fitting the L by L model

```r
> hsb
   ses  hsp   count u  v row col
 1 low general  50  1  1  1  1
 2 low academic  44  1  4  1  3
 3 low votech   45  1  2  1  2
 4 mid general  70  2  1  2  1
 5 mid academic 147  2  4  2  3
 6 mid votech   82  2  2  2  2
 7 hi  general  25  3  1  3  1
 8 hi  academic 117  3  4  3  3
 9 hi  votech   20  3  2  3  2

summary(lin.by.lin <- glm(count ~ ses + hsp + u*v, data=hsb, family=poisson))

Note: ses & hsp are factors and u and v are numeric.
```
R and Fitting the L by L model

Coefficients: (2 not defined because of singularities) ← can ignore

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.86477    | 0.55481 | 1.559    | 0.119075 |
| seslow    | 1.62718    | 0.29396 | 5.535    | 3.11e−08 *** |
| sesmid    | 1.58699    | 0.18695 | 8.489    | < 2e−16 *** |
| hspgeneral| 1.17067    | 0.30386 | 3.853    | 0.000117 *** |
| hspvotech | 0.59214    | 0.22834 | 2.593    | 0.009508 **  |
| u         | NA         | NA      | NA       | NA       |
| v         | NA         | NA      | NA       | NA       |
| u:v       | 0.32143    | 0.04862 | 6.612    | 3.80e−11 *** |

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

(Dispersion parameter for poisson family taken to be 1)
Null deviance: 206.9648 on 8 degrees of freedom
Residual deviance: 5.9798 on 3 degrees of freedom
AIC: 70.679
R and Fitting the L by L model

> (a ← anova(independence,lin.by.lin) )

Analysis of Deviance Table

<table>
<thead>
<tr>
<th>Model</th>
<th>count ~ ses + hsp</th>
<th>count ~ ses + hsp + u * v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resid. Df</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Resid. Dev</td>
<td>53.715</td>
<td>5.980</td>
</tr>
<tr>
<td>Df Dev</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Deviance</td>
<td>47.735</td>
<td></td>
</tr>
</tbody>
</table>

> dim(a)

2 4

> 1-pchisq(a[2,4],a[2,3])

4.878986e-12

> exp(lin.by.lin$coefficients[8])

u:v

1.379101
Choice of Scores

- Sets of scores with the same spacing between them will lead to the same goodness-of-fit statistics, fitted counts, odds ratios, and \( \hat{\beta} \). For HSB data, the following set of scores for the columns (hsp) would yield that same result: \( v_1 = 0, v_2 = 1, v_3 = 3 \).

- Two sets of scores with the same relative spacing will lead to the same goodness-of-fit statistics, fitted counts, and odds ratios, but different estimates of \( \beta \). e.g.,

  \[
  v_1 = 2, \quad v_2 = 4, \quad v_3 = 8
  \]

  With these column (HSP) scores, \( \hat{\beta} = .1607 \).

  - Odds ratio for low & middle (or middle & high) and vo/tech & general

    \[
    \hat{\theta} = \exp[.1607(2-1)(4-2)] = \exp[.1607(2)] = \exp[.3214] = 1.38
    \]

  - Odds ratio for low & high SES and program types vo/tech & academic:

    \[
    \hat{\theta} = \exp[.1607(3-1)(8-2)] = \exp[.1607(12)] = 6.88
    \]
**Uniform Association Model**

When scores are consecutive integers (or equally spaced scores) are used, e.g.,

\[ u_1 = 1, \ u_2 = 2, \ldots, u_I = I \]
\[ v_1 = 1, \ v_2 = 2, \ldots, v_J = J \]

This special case of L by L model is the "Uniform Association Model."

The uniform association model for the HSB example:

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>( G^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>4</td>
<td>53.715</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>L by L</td>
<td>3</td>
<td>5.980</td>
<td>.10</td>
</tr>
<tr>
<td>Uniform Assoc</td>
<td>3</td>
<td>11.74</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

This model is called the Uniform Association Model, because the odds ratios for any two adjacent rows and any two adjacent columns equals

\[ \theta = \exp \left[ \beta (u_i - u_{(i-1)})(v_j - v_{(j-1)}) \right] = \exp(\beta) \]

The "Local Odds Ratio" equals \( \exp(\beta) \) and is the same for adjacent rows and columns.
GSS example of Uniform Association Model

Recall...

- **Item 1:** A working mother can establish just as warm and secure of a relationship with her children as a mother who does not work.

- **Item 2:** Working women should have paid maternity leave.

<table>
<thead>
<tr>
<th>Item 1</th>
<th>strongly agree</th>
<th>agree</th>
<th>neither</th>
<th>disagree</th>
<th>strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>strongly agree</td>
<td>1</td>
<td>97</td>
<td>96</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>agree</td>
<td>2</td>
<td>102</td>
<td>199</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>disagree</td>
<td>3</td>
<td>42</td>
<td>102</td>
<td>25</td>
<td>36</td>
</tr>
<tr>
<td>strongly disagree</td>
<td>4</td>
<td>9</td>
<td>18</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

C.J. Anderson  (Illinois)  Model Building for Logit and Log-Linear Models  Fall 2019  56.58/139
**GSS Results**

<table>
<thead>
<tr>
<th>Model/Test</th>
<th>df</th>
<th>$G^2$</th>
<th>$p$</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>12</td>
<td>44.96</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>$M^2$</td>
<td>1</td>
<td>36.261</td>
<td>&lt; .001</td>
<td>$r = .20$</td>
</tr>
<tr>
<td>Uniform Assoc</td>
<td>11</td>
<td>8.67</td>
<td>.65</td>
<td>$\hat{\beta} = .24$, $ASE = .0412$</td>
</tr>
<tr>
<td>RC(1) Assoc</td>
<td>6</td>
<td>4.77</td>
<td>.57</td>
<td>$\hat{\phi} = 1.63$</td>
</tr>
</tbody>
</table>

$H_0 : \beta = 0$ vs $H_a : \beta \neq 0$,

L.R. test: $G^2 = (44.96 - 8.67) = 36.29$, $df = 1$, $p < .01$

The estimated local odds ratio equals $e^{.24} = 1.28$.

For the extreme corners of the table, the estimated odds ratio equals $e^{.24(3)(4)} = 18.5$

Unlike the tests of ordinal association that are based on a correlation, these models provide us with estimated odds ratios for the table, as well as permit us to check residuals, etc.
**RC(M) Association Model**

Random: Poisson  
Link: log  
Predictor: multiplicative interaction

\[
\log(\mu_{ij}) = \lambda + \lambda_i^R + \lambda_j^C + \sum_{m=1}^{M} \phi_m \nu_{im}^R \nu_{jm}^C
\]

where

- \( \nu_i^R \) and \( \nu_j^C \) are estimated row and column scale values on the \( m \)th dimension
- \( \phi_m \) is the association parameter

ID constraints (typical): Location

\[
\sum_i \lambda_i^R = \sum_j \lambda_j^C + \sum_i \nu_{im}^R = \sum_j \nu_{jm}^C = 0
\]

Scaling

\[
\sum_i (\nu_{im}^R)^2 = \sum_j (\nu_{jm}^C)^2 = 1
\]

Orthogonality (for \( m > 1 \))

\[
\sum_i \nu_{im}^R \nu_{im'}^R = \sum_j \nu_{jm}^C \nu_{jm'}^C = 0
\]
GSS Results: RC(1) scale values

RC(1) Association Model + 95% CI on Scale Values

Category scores

mapaid

fechld

1
2
3
4
5

-0.5 0.0 0.5 1.0
Back to HSB

For the HSB data, using equally spaced scores we find that

\[ M^2 = 40.87, \quad df = 1, \quad p < .001, \quad \text{and} \quad r = .26 \]

However, when we fit the linear by linear association model with equal scores it did not fit the data (this is shown in the residuals, as well).

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>( G^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>4</td>
<td>53.715</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>L by L (unequal spacing)</td>
<td>3</td>
<td>5.980</td>
<td>.10</td>
</tr>
<tr>
<td>Uniform Assoc (equal spacing)</td>
<td>3</td>
<td>11.74</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>RC(1) Assoc (estimated)</td>
<td>1</td>
<td>1.74</td>
<td>.19</td>
</tr>
</tbody>
</table>

First dimension from correspondence analysis accounts for 96.84% of Pearson’s \( X^2 \) from independence.
R: Fitting RC(M) Association Model

There are 2 packages, `gnm` (Generalized non-linear models) and `logmult` where the former is more general and the latter is a wrapper function for `gnm` specially designed for RC(M) association models.

```r
library(logmult)
rc1 <- rc(hsb.tab, nd = 1, weighting=c("none"), rowsup = NULL, colsup = NULL, se = c("jackknife"), nreplicates = 100, family = poisson )
plot(rc1, main="RC(1) Association Model")
library(gnm)
rc1.gnm <- gnm(counts ~ ses + hsp + Mult(ses,hsp), data=hsb, family=poisson, verbose=TRUE)
```

I use a variety of different programs do this...
Other Options for Fitting LMA Models

- LEM (Vermunt) can fit log-linear, latent class, and LMA models (https://jeroenvermunt.nl/Software). This was the precursor to LatentGold software.
- LatentGold, although I have never used it for LMA models.
- Various SAS macros for RC association models.
- LMA models: PROC NLP or NLMIXED where input model and likelihood. This will fit a wider array of models. The hard part is setting up data and typing out model, and it is limited in terms of size of problems (i.e., size of cross-classification).
- More general LMA models: A SAS macro that uses pseudo-likelihood estimation (Paek & Anderson, 2016). This uses PROC MDC (“multinomial discrete choice”), which is in the econometrics package.
Correspondence Analysis

- It is a data analytic technique and not a statistical model (i.e., no significance test).
- Provides another way to represent association between 2-variables.
- An optimal scaling procedure that decomposes Pearson’s $X^2$ from independence.
- The scale values or scores from the 1st dimension yield the largest possible correlation between rows and columns. For HSB data this equals

$$r = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{52.06}{600}} = .29$$

- Applied to 2-way tables (there are generalizations for higher-way).
- Gives another way to visualize associations.
- Interpretation... let’s look at graph...
Correspondence Analysis (continued)
Correspondence Analysis (continued)

Correspondence Analysis (96.84% of $X^2$)

High School Program Types

Academic  |  VoTech

SES

High  |  Middle  |  Low

High School Program Types

C.J. Anderson (Illinois)  |  Model Building for Logit and Log-Linear Models  |  Fall 2019  |  65.67/139
Ordinal Tests of Independence

CMH test was one way to test of ordinal association (or independence), but now we have a model based method. Using the linear by linear association model

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

The likelihood ratio test and the Wald test of the hypothesis

$$H_0 : \beta = 0$$

is the same as testing

$$H_0 : \text{independence}$$

Using the likelihood ratio test,

$$G^2(I|L \times L) = G^2(I) - G^2(L \times L)$$
Ordinal Tests of Independence

For the HSB data:

\[ G^2(I|L \times L) = 53.715 - 5.98 = 47.73 \]

with \( df = 4 - 3 = 1 \), and \( p < .001 \).

The Wald test:

\[ \left( \frac{\hat{\beta}}{ASE} \right)^2 = \left( \frac{.3199}{.0485} \right)^2 = 43.55 \]

The CMH test is the efficient score test for this same hypothesis

\[ M^2 = 40.87, df = 1, p < .001 \]
More Association Models for HSB

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>$G^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence</td>
<td>4</td>
<td>53.715</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Uniform Assoc (equal spacing)</td>
<td>3</td>
<td>11.74</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>L by L (unequal spacing)</td>
<td>3</td>
<td>5.980</td>
<td>.10</td>
</tr>
<tr>
<td>Nominal HSP × Ordinal SES (equal spaced SES)</td>
<td>2</td>
<td>2.30</td>
<td>.32</td>
</tr>
<tr>
<td>RC association (scores estimated)</td>
<td>1</td>
<td>1.74</td>
<td>.19</td>
</tr>
</tbody>
</table>

The estimated parameters for the SES × HSP association in the nominal × ordinal model:

$\hat{\beta}_{\text{votech}} = .000, \quad \hat{\beta}_{\text{general}} = -.005, \quad \hat{\beta}_{\text{academic}} = .864$

RC association model estimates the scores for both SES and HSP, as well as $\beta$ (the “association parameter”).

<table>
<thead>
<tr>
<th>HSP</th>
<th>est. score</th>
<th>SES</th>
<th>est. score</th>
</tr>
</thead>
<tbody>
<tr>
<td>VoTech</td>
<td>-.423</td>
<td>Low</td>
<td>-.669</td>
</tr>
<tr>
<td>General</td>
<td>-.393</td>
<td>Middle</td>
<td>-.071</td>
</tr>
<tr>
<td>Academic</td>
<td>.816</td>
<td>High</td>
<td>.740</td>
</tr>
</tbody>
</table>

and $\hat{\beta} = 1.000$
Comments on models for ordinal variables

- This approach is not restricted to models for 2–way tables and log-linear models. You add use scores in log-linear and/or logit model for higher–way tables.

- There are more general models where the scores are estimated from the data. For 2–way tables, this includes Goodman’s “row effects” model ($R$), “column effects” model ($C$), “row + column” effects model ($R + C$), and the row–column model $RC$. There are generalizations of these models to multiple dimensions and higher–way tables.

- There are also models for ordinal response variables that take into account the ordering of the categories.


- Log multiplicative models with latent variable interpretations (Anderson & Vermunt, 2000; Anderson, 2002; Anderson & Yu, 2007; Anderson, Li & Vermunt, 2007; Anderson, Verkuilen & Peyton, 2012; Anderson (2013); papers by group in Amsterdam and by group at Columbia).
Wickens & Olzak revisited

A good model for Wickens & Olzak data is \((HLY, HY, LY, XY)\),

\[
\log(\mu_{ijkl}) = \lambda + \lambda_i^H + \lambda_j^L + \lambda_{ij}^{HL} + \lambda_k^X + \lambda_{kl}^{XY} + \lambda_{ik}^{LY} + \lambda_{ijl}^{HLY}
\]

Let \(u_k = 1, \ldots, 6\) and \(v_\ell = 1, \ldots, 6\) be scores for the high and low responses, respectively. We can use these instead of nominal responses:

\[
\log(\mu_{ijkl}) = \lambda + \lambda_i^H + \lambda_j^L + \lambda_{ij}^{HL} + \lambda_k^X + \lambda_{kl}^{XY} + \lambda_i^H u_k + \lambda_j^L v_\ell + \lambda_{ij}^{HL} v_\ell
\]

This model doesn’t fit particularly well \((G^2 = 259.1267, df = 100, p < .01)\), but one with estimated scores does.
Eg. of a Latent Variable Model

\[
\log(\mu_{ijkl}) = \lambda + \lambda_{ij}^{HL} + \lambda_k^X + \lambda_l^Y + \lambda_{kl}^{XY} + \sigma_1^2 \nu_{1(ij)}^{HL} \nu_k + \sigma_2^2 \nu_{2(ij)}^{HL} \nu_l
\]

\[G^2 = 138.35, \ df = 98, \ p = .01, \ D = .082.\]

But there were 2 subjects and this graph describes both. For the other subject ("subject A"), \[G^2 = 111.12, \ df = 97, \ p = .15, \ D = .086.\]
Estimated Scores

![Graph showing estimated scores against rating scale](image-url)
## Parameter Estimates— the low signal

For both subjects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>Parameter estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>rating</td>
<td>1</td>
<td>( \hat{\nu}_1^{\text{rating}} = -0.42 )</td>
<td>(.01)</td>
</tr>
<tr>
<td>rating</td>
<td>2</td>
<td>( \hat{\nu}_2^{\text{rating}} = -0.38 )</td>
<td>(.01)</td>
</tr>
<tr>
<td>rating</td>
<td>3</td>
<td>( \hat{\nu}_3^{\text{rating}} = -0.24 )</td>
<td>(.01)</td>
</tr>
<tr>
<td>rating</td>
<td>4</td>
<td>( \hat{\nu}_4^{\text{rating}} = 0.00 )</td>
<td>(.02)</td>
</tr>
<tr>
<td>rating</td>
<td>5</td>
<td>( \hat{\nu}_5^{\text{rating}} = 0.31 )</td>
<td>(.02)</td>
</tr>
<tr>
<td>rating</td>
<td>6</td>
<td>( \hat{\nu}_6^{\text{rating}} = 0.73 )</td>
<td>(.02)</td>
</tr>
<tr>
<td>signal</td>
<td>high absent/low absent</td>
<td>( \hat{\nu}_{2(11)}^{(HL)} = -0.49 )</td>
<td>(.00)</td>
</tr>
<tr>
<td>signal</td>
<td>high present/low absent</td>
<td>( \hat{\nu}_{2(21)}^{(HL)} = -0.49 )</td>
<td>(.00)</td>
</tr>
<tr>
<td>signal</td>
<td>high absent/low present</td>
<td>( \hat{\nu}_{2(12)}^{(HL)} = 0.61 )</td>
<td>(.02)</td>
</tr>
<tr>
<td>signal</td>
<td>high present/low present</td>
<td>( \hat{\nu}_{2(22)}^{(HL)} = 0.37 )</td>
<td>(.03)</td>
</tr>
<tr>
<td>( \Theta_2 )</td>
<td></td>
<td>( \hat{\sigma}_2^2 = 3.30 )</td>
<td>(.12)</td>
</tr>
</tbody>
</table>
Parameter Estimates— the high signal

The $\nu$’s for ratings are the same as previous slide.

<table>
<thead>
<tr>
<th></th>
<th>Subject A</th>
<th>Subject B</th>
</tr>
</thead>
<tbody>
<tr>
<td>high absent/low absent</td>
<td>$\hat{\nu}^{(HL)}_{1(11)A}$ = $-.58$ (.00)</td>
<td>$\hat{\nu}^{(HL)}_{1(11)B}$ = $-.50$ (n.a.)</td>
</tr>
<tr>
<td>high present/low absent</td>
<td>$\hat{\nu}^{(HL)}_{1(21)A}$ = $.41$ (.00)</td>
<td>$\hat{\nu}^{(HL)}_{1(21)B}$ = $.50$ (n.a.)</td>
</tr>
<tr>
<td>high absent/low present</td>
<td>$\hat{\nu}^{(HL)}_{1(12)A}$ = $-.39$ (.00)</td>
<td>$\hat{\nu}^{(HL)}_{1(12)B}$ = $-.50$ (n.a.)</td>
</tr>
<tr>
<td>high present/low present</td>
<td>$\hat{\nu}^{(HL)}_{1(22)A}$ = $.58$ (.00)</td>
<td>$\hat{\nu}^{(HL)}_{1(22)B}$ = $.50$ (n.a.)</td>
</tr>
<tr>
<td>$\Theta_1$</td>
<td>$\hat{\sigma}^2_{1A}$ = $2.69$ (.18)</td>
<td>$\hat{\sigma}^2_{1B}$ = $7.11$ (.49)</td>
</tr>
</tbody>
</table>
Tests of Conditional Independence

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).
   1.1 Comparing conditional independence model to homogeneous association model.
   1.2 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.
## Nature of Variables: Ordinal &/or Nominal

We have 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 (Generalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Column</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>
Some Data that We’ll Use

For illustration, we’ll use some High School & Beyond data, i.e., the cross-classification of gender (G), SES (S) and high school program type (P).

### Females

<table>
<thead>
<tr>
<th>SES</th>
<th>High School Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>middle</td>
<td>44</td>
<td>144</td>
</tr>
<tr>
<td>high</td>
<td>12</td>
<td>79</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>273</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES</th>
<th>VoTech</th>
<th>General</th>
<th>Academic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>30</td>
<td>31</td>
<td>28</td>
<td>89</td>
</tr>
<tr>
<td>middle</td>
<td>38</td>
<td>40</td>
<td>77</td>
<td>155</td>
</tr>
<tr>
<td>high</td>
<td>8</td>
<td>14</td>
<td>61</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>85</td>
<td>166</td>
<td>327</td>
</tr>
</tbody>
</table>

### Males

<table>
<thead>
<tr>
<th>SES</th>
<th>High School Program</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>middle</td>
<td>44</td>
<td>144</td>
</tr>
<tr>
<td>high</td>
<td>12</td>
<td>79</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
<td>273</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES</th>
<th>VoTech</th>
<th>General</th>
<th>Academic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>low</td>
<td>30</td>
<td>31</td>
<td>28</td>
<td>89</td>
</tr>
<tr>
<td>middle</td>
<td>38</td>
<td>40</td>
<td>77</td>
<td>155</td>
</tr>
<tr>
<td>high</td>
<td>8</td>
<td>14</td>
<td>61</td>
<td>83</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>85</td>
<td>166</td>
<td>327</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 $df = df(XZ, YZ) - df(XY, XZ, YZ)$. 
The likelihood ratio test (example)

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

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

<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 ) (G^2 ) (p)</td>
<td>(\Delta df) (\Delta G^2) (p)</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>
</tbody>
</table>
Notes Regarding Likelihood Ratio Test

- This test assumes that $(XY, XZ, YZ)$ holds.
- This single test is preferable 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, from SAS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>ASE</th>
<th>Wald</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_{lv}^{SP}$</td>
<td>1.8133</td>
<td>.3233</td>
<td>31.450</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>$\lambda_{lg}^{SP}$</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

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

which corresponds to the $(XY, XZ, YZ)$ log-linear model.
Notes Regarding Likelihood Ratio Test

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 $H_0 : \lambda_{ij}^{XY} = 0$ in the log-linear model.
Direct Goodness-of-Fit Test

We compare the fit of the conditional independence model to the saturated model; that is,

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

The null hypothesis for this test statistic is

\[ H_0 : \text{all } \lambda_{ij}^{XY} = 0 \text{ and all } \lambda_{ijk}^{XYZ} = 0 \]

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

\[ H_0 : \text{all } \lambda_{ij}^{SP} = 0 \text{ and all } \lambda_{ijk}^{GSP} = 0 \]

<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>
</tbody>
</table>
Notes on Direct Goodness-of-Fit Test

A direct goodness-of-fit test does not assume that \((XY, XZ, YZ)\) holds, while using \(G^2 \left[ (XZ, YZ) | (XY, XZ, YZ) \right]\) 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., \(G^2 \left[ (XZ, YZ) | (XY, XZ, YZ) \right]\)).
**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 \( Y \).

Notes:

- 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.
**Ordinal Conditional Association**

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

$$\log(\mu_{ijk}) = \lambda + \lambda^S_i + \lambda^P_j + \lambda^G_k + \beta u_i v_j + \lambda^{SG}_{ik} + \lambda^{PG}_{jk}$$

<table>
<thead>
<tr>
<th>Model</th>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$df$ $G^2$ $p$</td>
<td>$\Delta df$ $\Delta G^2$ $p$</td>
</tr>
<tr>
<td>$(GS, GP, SP)$</td>
<td>4 1.970 .74</td>
<td>— —</td>
</tr>
<tr>
<td>$(GS, GP, SP) - L \times L$</td>
<td>7 7.476 .38</td>
<td>3 5.505 .138</td>
</tr>
<tr>
<td>$(GS, GP)$</td>
<td>8 55.519 &lt; .0001</td>
<td>1 48.043 &lt; .001</td>
</tr>
</tbody>
</table>

From before...

<table>
<thead>
<tr>
<th>Model</th>
<th>Goodness-of-fit Test</th>
<th>Likelihood Ratio Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$df$ $G^2$ $p$</td>
<td>$\Delta df$ $\Delta G^2$ $p$</td>
</tr>
<tr>
<td>$(GS, GP, SP)$</td>
<td>4 1.970 .74</td>
<td>— —</td>
</tr>
<tr>
<td>$(GS, GP)$</td>
<td>8 55.519 &lt; .0001</td>
<td>4 53.548 &lt; .001</td>
</tr>
</tbody>
</table>
Example Continued... 

- The null hypothesis for the likelihood ratio test statistic (in the last row of top table) is $H_O : \beta = 0$ with $df = 1$; whereas, in the lower table, it is

$$H_O : \text{all } \lambda_{ij}^{SP} = 0 \text{ with } df = 4$$

- Comparing $G^2/df$ for the two tests,

$$53.548/4 = 13.387 \text{ versus } 48.043/1 = 48.043$$

- **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.
Estimated Partial Odds Ratios

\( \hat{\beta} = .3234 \). 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,

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

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...

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

Next, 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.
Nominal–Ordinal Case

For example, if only have put in scores for SES, we fitting the 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.

<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, 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>
<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>

For the nominal–ordinal model, from SAS: $\hat{\beta}_v^P = -0.8784$, $\hat{\beta}_g^P = -0.8614$, $\hat{\beta}_a^P = 0$, and from R: $\hat{\beta}_v^P = -0.0170$, $\hat{\beta}_g^P = 0.87$ $\Rightarrow$ 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 (Generalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Likelihood</td>
</tr>
<tr>
<td>Row</td>
<td>Column</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>

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**.
Generalized Cochran-Mantel-Haenszel Tests

- 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 $G^2 [(XZ, YZ)|(XY, XZ, YZ)]$).

- **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 $df = 1$. 
Nominal–Ordinal Generalized CMH

- Suppose that $X$ (row) is nominal and $Y$ (column) is ordinal.
- Responses on each row can be summarized by the row mean score.
- 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 $H_O: \mu_{Y_j} = \mu_Y$ is true (i.e., 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 $Y \sim \mathcal{N}(\mu_{Y_j}, \sigma^2)$, a 1–way ANOVA would be an appropriate test; that is, the nominal–ordinal generalized CMH statistic is analogous to a 1–way ANOVA.
- Using midranks are used as scores in the generalized CMH statistic is equivalent to the Kruskal–Wallis (non-parametric) test for comparing mean ranks.
Example of Nominal–Ordinal Generalized CMH

In SAS or R output, the 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>
Example: Nominal–Ordinal Generalized CMH

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

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

<table>
<thead>
<tr>
<th>Gender</th>
<th>High School</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>VoTech</td>
<td>15</td>
<td>44</td>
<td>12</td>
<td>139/71 = 1.96</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>19</td>
<td>30</td>
<td>11</td>
<td>112/60 = 1.87</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td>16</td>
<td>70</td>
<td>56</td>
<td>324/142 = 2.28</td>
</tr>
<tr>
<td>Males</td>
<td>VoTech</td>
<td>30</td>
<td>38</td>
<td>8</td>
<td>130/76 = 1.71</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>31</td>
<td>40</td>
<td>14</td>
<td>153/85 = 1.80</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td>28</td>
<td>77</td>
<td>61</td>
<td>365/166 = 2.20</td>
</tr>
</tbody>
</table>
Nominal–Nominal Generalized CMH

- CMH test statistic is a test of “general association”.
- 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 $H_0: \lambda_{ij}^{XY} = 0$ in the $(XY, XZ, YZ)$ log-linear model.
- If the null is true, then the statistic is approximately chi–squared distributed with $df = (I - 1)(J - 1)$.

<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>
## Summary: Tests of Ordinal Association

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type of Test (Generalized)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Likelihood Ratio</td>
</tr>
<tr>
<td>Row</td>
<td>Column</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>

...and for the curious and sake of completeness...
Summary: Tests of Ordinal Association

Models for SES × Gender × Program type:

<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 that appears to fit the data:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^S + \lambda_j^P + \lambda_k^G + \beta u_i v_j$$
Sparse Data

and Incomplete Tables Methodology

- Types of empty cells (sampling and structural zeros).
- Effects of sampling zeros and strategies for dealing with them.
- 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

- **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 \( n_{ij} > 0 \).

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

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

Survey of teenagers regarding their health concerns (Fienberg):

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

The probability of a male with menstrual problems = 0.
Incomplete Data: What to do

- It is important to recognize that a table is incomplete.

- Determine 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

- Problems that can be encountered when modeling sparse tables.
- The effect of spareness on hypothesis testing.

Problems in modeling Sparse Tables.
There are two major ones:

- Maximum likelihood estimates of loglinear/logit models may not exist.
- 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):

\[
\begin{array}{cccc|cccc|cccc}
& Z = 1 & & Z = 2 & & Z = 3 \\
Y = & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 & 1 & 2 & 3 & 4 \\
X = 1 & 5 & 0 & 7 & 8 & 9 & 8 & 3 & 12 & 6 & 3 & 5 & 11 \\
X = 2 & 10 & 0 & 6 & 7 & 8 & 3 & 0 & 9 & 0 & 2 & 8 & 11 \\
\end{array}
\]

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

\[
\begin{array}{|c|c|c|}
\hline
X & 1 & 2 \\
\hline
77 & 64 \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|c|}
\hline
Y & 1 & 2 & 3 & 4 \\
\hline
38 & 16 & 29 & 58 \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|}
\hline
Z & 1 & 2 & 3 \\
\hline
43 & 52 & 46 \\
\hline
\end{array}
\]
The 2–way margins

<table>
<thead>
<tr>
<th></th>
<th>$Y$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>$X = 1$</td>
<td>20</td>
<td>11</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>$X = 2$</td>
<td>18</td>
<td>5</td>
<td>14</td>
<td>27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$Z$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$X = 1$</td>
<td>20</td>
<td>32</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>$X = 2$</td>
<td>23</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$Y$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z = 1$</td>
<td>15</td>
<td>0</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>$Z = 2$</td>
<td>17</td>
<td>11</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>$Z = 3$</td>
<td>6</td>
<td>5</td>
<td>13</td>
<td>22</td>
</tr>
</tbody>
</table>

- Since $n_{+21} = 0$, $YZ$ partial odds ratios involving this cell equal 0 or $+\infty$.
- The $YZ$ margin has a zero $\rightarrow$ no MLE estimate of $\lambda_{21}^{YZ}$.
- Suppose that $n_{121} > 0$, could you fit $(XY, YZ)$? Could you fit the saturated model $(XYZ)$.
Example from Tettegah & Anderson (2007)

- Recognition of the victim.
- Expression of empathic Concern for the victim.
- Managing the situation with the victim.
- Problem-Solving strategies.

\( N = 178 \)

<table>
<thead>
<tr>
<th>Solve</th>
<th>Manage</th>
<th>Mention</th>
<th>No Concern</th>
<th>Yes Concern</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td></td>
<td>38</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td></td>
<td>51</td>
<td>4</td>
<td>16</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td>2</td>
<td>1</td>
<td>21</td>
<td>17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What models can and cannot be fit to these data?
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.

Note: This is using the Wicken’s data.

R does not given any warning message.
I. Signs of a problem

The estimated standard errors of parameters and fitted counts are really, really large 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.

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

\[
\hat{\mu}_{221} = 5.94 \times 10^{-11}, \quad \log(\hat{\mu}_{121}) = -23.3468, \quad \text{std err} = 87,417
\]

R also have ridiculously large S.E.s, i.e., \(\hat{\mu}\)s for these cells ±29 and se = 3,966.26
Spareness & Odds Ratio Estimates

- Spareness can cause
  - The sampling distribution of fit statistics will be poorly approximated by the chi–squared distribution.
  - Odds ratio estimates to be severely biased
- Solution: add .5 to each cell in the table.
- Adding .5 shrinks the estimated odds ratios that are $\infty$ to finite values and increases estimates that are 0.
- Qualifications: For unsaturated models, adding .5 will over smooth the data.
- Remedies/Strategies/Comments...
Spar森ess & Odds Ratio Estimates

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., $1^{-8}$) to all cells in the table.
- Do a sensitivity analysis by adding different numbers of varying sizes to the cells (e.g., $1^{-8}$, $1^{-5}$, .01, .1). Examine fit statistics and parameter estimates to see if they change very much.
**Example: Hypothetical Data**

and the \((X, YZ)\) loglinear model:

<table>
<thead>
<tr>
<th>Number added</th>
<th>(G^2)</th>
<th>(X^2)</th>
<th>Converge?</th>
<th>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>

Alternative: Use an alternative estimation procedure (i.e., Bayesian).
Log-Linear Models & Empathy Data

Independence

Vignette
Manage
Solve
Concern
Mention

All 2-way interactions

Vignette
Manage
Solve
Concern
Mention

Unrelated to Vignette

Vignette
Manage
Solve
Concern
Mention

$G^2 = 137.46$
$df = 26, p < .01$

$G^2 = 13.69$
$df = 16, p = .62$

$G^2 = 15.24$
$df = 29, p = .76$

⇒ Collapse Data over Vignettes.
Effect of Sparseness on $X^2$ and $G^2$

Guidelines:

- 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 $\hat{\mu} < 5$.
  
  Empathy data: 37.5% of cells equal 0.

- The permissible size of $\hat{\mu}$ decreases as the size of the table $N$ increases.

- 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 $n/N$). Empathy data: sample size/size of table $= 178/16 = 11 \ldots$ maybe OK.

- No single rule covers all situations.

- $X^2$ tends to be valid with smaller $n$ and sparser tables than $G^2$. 
Effect of Sparseness on $X^2$ and $G^2$

Guidelines (continued):

- $G^2$ 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 $n/N$).

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

$G^2$ for model fit may not be well approximated by the chi-squared distribution, but the distribution of difference between $G^2$'s for two nested models maybe.

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

So if margins have cells $> 5$, the chi-squared approximation of $G^2(M_0) - G^2(M_1)$ should be reasonable.
Effect of Sparseness on $X^2$ and $G^2$

Guidelines (continued):

- **Empathy log-linear models:**
  \[ G^2(\text{unrelated to vignette}|\text{All two-way}) = 15.24 - 13.69 = 1.55, \]
  \[ df = 13, \ p \text{ large}. \]

- **Exact tests and exact analyses for models.**

- **An alternative test statistic:** the Cressie-Read statistic
Cressie-Read statistic


They proposed a family of statistics of the form

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

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 \(G^2\).
Example: Teens and Health Concerns

- The 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 log-linear model as

\[
\log(\mu_{ij}) = \begin{cases} 
0 & \text{for the (2,1) cell} \\
\lambda + \lambda_i^H + \lambda_j^G + \lambda_{ij}^{HG} & \text{for the rest}
\end{cases}
\]

- Or equivalently we define an indicator variable...
Example: Teens and Health Concerns

\( I(2,1) = 1 \) for the (2,1) cell
\( = 0 \) otherwise

A single equation for the saturated log-linear 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?

\[
df = (\# \text{ of cells}) - (\# \text{ non-redundant parameters})
\]
\[
= (\text{usual } df \text{ for the model}) - (\# \text{ cells fit perfectly})
\]
Independence: Teens and Health Concerns

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

\[ G^2 = 12.60, \quad \text{and} \quad X^2 = 12.39, \] 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

\[ \hat{\lambda} = 4.5466 \]
\[ \hat{\lambda}_1^H = -2.0963 \quad \hat{\lambda}_1^G = -1.1076 \]
\[ \hat{\lambda}_2^H = -2.0671 \quad \hat{\lambda}_1^G = 0.0000 \]
\[ \hat{\lambda}_3^H = -0.8307 \]
\[ \hat{\lambda}_4^H = 0.0000 \quad \hat{\delta}_{21} = -22.9986 \]

For the (2, 1) cell,

\[ \hat{\mu}_{21} = \exp(4.5466 - 2.0671 - 0.1076 - 22.9986) \sim 0 \]
Anomalous Cells

- A model fits a table well, except for one or a few cells.

- The methodology for incomplete tables can be used to show that except for these cells, the model fits.
  \[\ldots\] 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.
Example Anomalous Cells

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>

Gender (\(G\)) is the response/outcome variable and Year (\(Y\)) and Religion (\(R\)) are explanatory:
Example Anomalous Cells

Gender (G) is the response/outcome variable and Year (Y) and Religion (R) are explanatory:

<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>
A closer look at models

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

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

with \( df = 3 \) and \( p = .01 \). Let's take a closer look at (RY,GY).

- The Pearson residuals from the (RY,GY) log-linear model

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

The 3 largest residuals → mothers who are Catholic. The model under predicts “both” and over predicts “boys”.
Deal Anomalous Cells

**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

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

**Why do we only need 3 indicators to “remove” the row for Catholic mothers?**

This model has \(df = 4, G^2 = 1.35, \) and \(X^2 = 1.39.\) So the (RY, GY) model fits well without the second row of the table.
Logit Model Example

Data are from Farmer, Rotella, Anderson & Wardrop (1998). Individuals are from a longitudinal study who had chosen a career in science. They were cross-classified according to their gender and the primary Holland code describing the type of career in science that they had chosen.

Interest was in testing whether women and men differed, and if so describing the differences. We’ll treat gender as a response variable.

<table>
<thead>
<tr>
<th>Holland Code</th>
<th>Gender</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men</td>
<td>Women</td>
</tr>
<tr>
<td>Realistic</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Investigative</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Artistic</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Social</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Enterprising</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Conventional</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>
The logit model corresponding to the (H,G) log-linear 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,

<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>Enterprizing</td>
<td>-0.73</td>
</tr>
<tr>
<td>Conventional</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Two largest: Realistic and Social.

\[\rightarrow\] fit these perfectly but allow independence in the rest of the table.
Farmer, Rotella, Anderson & Wardrop

Realistic and Social.

→ fit these perfectly but allow independence in the rest of the table.

The logit model without Realistic and Social:

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

where

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

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

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

Recall that the residuals from the independence models for realistic and social are both quite large but opposite signs.
Capturing the Association

Let’s define a new variable to capture the suspected association structure,

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

and fit the model

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

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 \( H_0 : \beta_{\text{social}} = -\beta_{\text{realistic}} = \beta \).
Comparing Adjusted Residuals

The adjusted residuals look pretty good for new model

<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>-0.86</td>
</tr>
<tr>
<td>Artistic</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 shovelling 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

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

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

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

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

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

### A 6–way, incomplete table

<table>
<thead>
<tr>
<th>Four Religion Variables</th>
<th>1953</th>
<th>1971</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>Catholic</td>
<td>Jewish</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Models for Snow Shovelling

Since G (gender) is considered the response and all log-linear models must include $\lambda^{YP\text{CJO}}$ terms (and lower order ones). 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>$(YP\text{CJO},GY)$</td>
<td>6</td>
<td>11.2</td>
</tr>
<tr>
<td>$(YP\text{CJO},GY,GPCJO)$</td>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>New ones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(YP\text{CJO},GY,GO)$</td>
<td>5</td>
<td>9.8</td>
</tr>
<tr>
<td>$(YP\text{CJO},GY,GJ)$</td>
<td>5</td>
<td>10.9</td>
</tr>
<tr>
<td>$(YP\text{CJO},GY,GC)$</td>
<td>5</td>
<td>1.4</td>
</tr>
<tr>
<td>$(YP\text{CJO},GY,GP)$</td>
<td>5</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Conclusions for Snow Shovelling

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).
Summary

- Association Graphs.
  - Introduction.
  - Collapsibility.
  - Representing models.

- Modeling ordinal association.
  - linear by linear association, (and RC(M) association model & correspondence analysis)
  - ordinal tests of independence.

- Testing conditional independence.

- Effects of sparse data.

- Model fitting details.

- A hybrid models (log-linear with numerical predictors)