Model Building for Log-linear and Logit Models

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS

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 interaction.
- 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.
- Interactions in 3–way contingency tables. Barlett (1935). The notion of interaction in contingency tables studied by Barlett is formally identical to the notions used in statistical physics. The development of graphical models for multi-way contingency data stems from a paper by Darroch, J.N., Lauritzen, S.L., & Speed, T.P. (1980).

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.

Darroch, J.N., Lauritzen, S.L., & Speed, T.P. (1980). Markov fields and log-linear models for contingency tables. *Annals of Statistics, 8*, 522–539.

Edwards, D. (2000). *Introduction to Graphical Modeling*, 2nd Edition. NY: Springer–Verlag.

Lauritzen, S.L. (1996). *Graphical Models*. NY: Oxford Science Publications.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, 2nd Edition. Chichester: Wiley.

Graphical Models & Contingency Tables

We'll be using graphs to

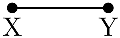
1. Help determine when marginal and partial associations are the same such that we can *collapse* a multi-way table into a smaller table (or tables) to study certain associations.
2. Represent substantive theories and hypotheses, which correspond to certain loglinear/logit models.

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.

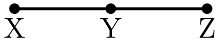


  The absence of an edge between two vertices indicates that the two variables are independent.

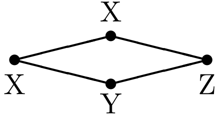
  We will be (mostly) restricting our attention to undirected relationships, so our lines won't have arrows on them (lines with arrows represent directed relationships).

More Terminology & Definitions

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

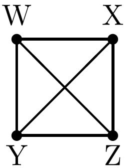


- Separated. Two variables are said to be separated if all paths between the two variables intersect a third variable (or set of variables).



Even More Definitions

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

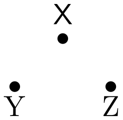


This is also known as a "complete graph" and if this is part of a larger graph, a "complete subgraph".

Fundamental Result (cornerstone of graphical modeling): Two variables are conditionally independent given any subset of variables that separates them.

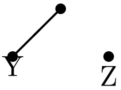
Graphs for Log-linear Models of. . .

The graph for the Complete Independence, $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$

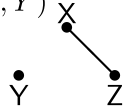


The graphs for Joint Independence,

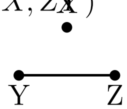
(XY, Z)



(XZ, Y)

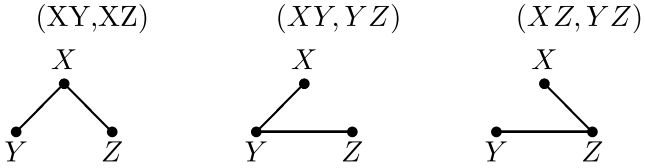


(X, ZY)



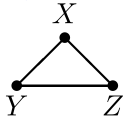
Graphs for Conditional Independence

The graphs for Conditional Independence:



Graphs for 3–Way Association model

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



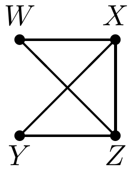
This is also a graph for Homogeneous Association, (XY,XZ,YX), which is also a model of dependence.

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.

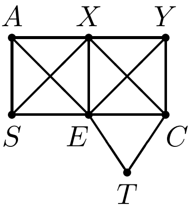
- Each graph implies at least one log-linear model. Unless otherwise specified, the model "read" from a graph will the most complex one.

What is the log-linear model for this graph?



More Association Graphs & Log-linear Models

What is the log-linear model for this graph?



More Association Graphs & Log-linear Models

What is the graph for this log-linear model?

$$(\mathrm{WY},\ \mathrm{YZ},\ \mathrm{ZX})$$

Are there other log-linear models with this graphical representation?
What is the graph for this log-linear model?

$$(\mathrm{WXY},\mathrm{WXZ})$$

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,

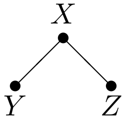
The X-Y marginal and partial odds ratios are identical if either the

- *Log-linear model (XY, ZY) holds, or*
- *Log-linear model (XY, XZ) holds.*

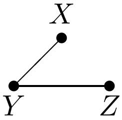
Collapsibility in Graphical Terms

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*



or



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:

Manage	Super	Worker	n_{ijk}	M, S, W	MS, W	MS, MW	no MSW
bad	low	low	103	50.15	71.78	97.16	102.26
bad	low	high	87	82.59	118.22	92.84	87.74
bad	high	low	32	49.59	27.96	37.84	32.74
bad	high	high	42	81.67	46.04	36.16	41.26
good	low	low	59	85.10	63.47	51.03	59.74
good	low	high	109	140.15	104.53	116.97	108.26
good	high	low	78	84.15	105.79	85.97	77.26
good	high	high	205	138.59	174.21	197.28	205.74

Partial and marginal odds ratios computed using fitted values.

	Partial Odds Ratio			**Marginal** Odds Ratio		
Model	W–S	M–W	M–S	W–S	M–W	M–S
(M, S, W)	1.00	1.00	1.00	1.00	1.00	1.00
(MS, W)	1.00	1.00	4.28	1.00	1.00	4.28
(MS, MW)	1.00	2.40	4.32	1.33	2.40	4.32
(MS, WS, MW)	1.47	2.11	4.04	1.86	2.40	4.32
(MSW) level 1	1.55	2.19	4.26	1.86	2.40	4.32
(MSW) level 2	1.42	2.00	3.90			

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:

Race	Sex	Graduates	Sample Size
White	women	498	796
White	men	878	1625
Black	women	54	143
Black	men	197	660

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,

$$\theta_{SG(i)} = \exp(\beta_f^S - \beta_m^S)$$

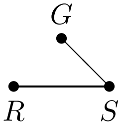
Collapsibility & Logit Models (continued)

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

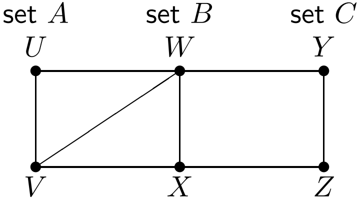
from Agresti

> *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.*

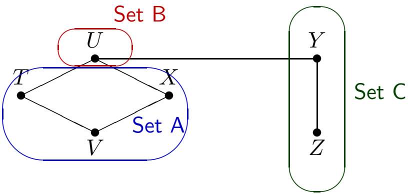
Graphically, each path between variables in set A and variables in set C involve at least 1 variable in set B. . . .

Collapsibility for Multiway Tables

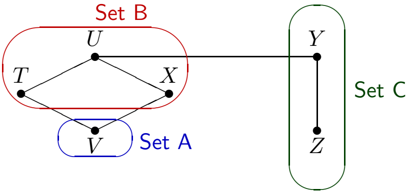
Graphically, each path between variables in set A and variables in set C involve at least 1 variable in set B.



Example of Collapsibility & Multiway Tables

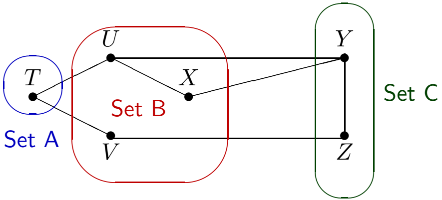


A 2nd Possibility:

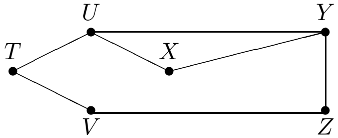


And others?

Example 2 of Collapsibility & Multiway Tables



A 2nd Possibility:



Using Graphs to Guide Modeling

Example: Data from a concurrent-task detection experiment. (Olzak, 1981; Olzak & Wickens, 1983; Wickens, 1989; Anderson, 2002; Kroonenberg & Anderson, 2006).

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

- 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 2×2 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

		High Frequency Signal											
Low		Absent						Present					
Freq	Y	$X=1$	2	3	4	5	6	1	2	3	4	5	6
	1	69	6	1	1	0	0	10	5	2	11	16	28
	2	34	20	10	3	1	0	8	5	11	43	27	38
Absent	3	43	24	13	9	1	0	9	6	7	28	32	45
	4	78	40	20	6	0	1	8	6	14	19	23	22
	5	32	38	17	5	4	0	4	5	7	6	18	18
	6	5	14	3	2	0	0	0	1	2	3	5	8
	1	4	1	0	0	0	0	5	0	1	4	4	9
	2	5	3	2	1	0	0	0	1	3	6	9	27
Present	3	8	6	3	1	0	0	2	3	2	11	27	20
	4	36	25	18	3	1	0	9	12	11	10	23	31
	5	83	69	26	6	1	0	16	7	5	19	23	40
	6	127	50	12	7	2	0	21	14	13	20	21	61

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.

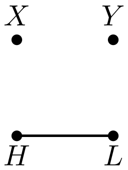
Assumptions?

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 λ^{HL} for the HL association. The simplest model would be that a subject responds randomly

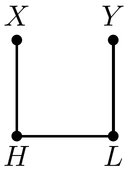


The log-linear model: (HL, X, Y).

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

Detectable Signals

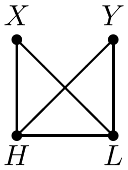
The subject can detect the signals & detecting one does not influence detection of the other (i.e., subject does what the experimenter asked).



Log-linear model: (HL, XH, YL)
This is a "base" model to which we can add more complicated forms of associations.
$G^2 = 375.72$, $df = 120$, $p < .01$
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

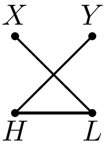
In this model, responses to one signal are influenced by whether both signals are present and/or absent (i.e., the appropriate and inappropriate signal).



The log-linear model (HLX, HLY)
$G^2 = 221.43$, $df = 100$, $p < .01$

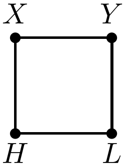
Association to the unrelated signal

- 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



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.



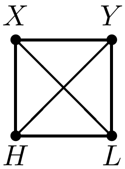
The log-linear model: (HL, HX, LY, XY).

$G^2 = 159.27$, $df = 95$, $p < .01$

All pairwise associations.

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

It's graphical representation is



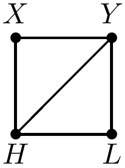
This is also the representation of many other log-linear models with dependencies, including model with 4-way interaction (i.e. saturated model).
This is the most complex graph, but there are interesting log-linear models that have this representation.

Another Model

We can add three–factor terms to the all pairwise association model. Some of these all have reasonable interpretations.

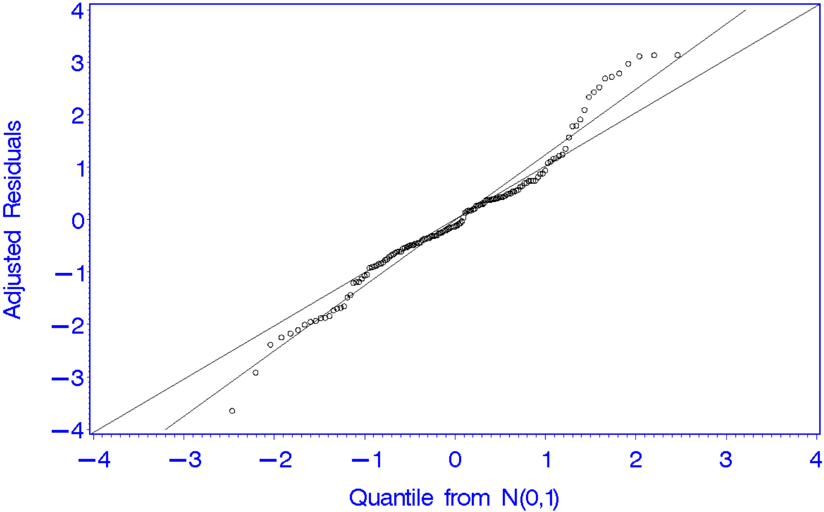
For example, consider the model that adds λ^{HLY}. The λ^{HLY} terms imply that delectability of the L signal (measured by Y) is affected by the presence of the H signal.

It's graphical representation is



Fit of this model $G^2 = 98.06$, $df = 80$, $p = .08$

QQ Plot for (HLY,HY,LY,XY)



Eg: 4-way Table with Time Ordering

Using a suggested ordering of the variables in terms of time and causal hypotheses and show how to "decompose" a model into smaller pieces.

Example from Agresti, 1990; The variables:

- **G** for gender.
- **PMS** for premarital sex.
- **EMS** for extra martial 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:

Stage	Response	Explanatory
(1)	PMS	Gender
(2)	EMS	Gender, PMS
(3)	M	Gender, PMS, EMS

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

$$
\begin{tikzcd}
 & & & E \arrow[dr] & \\
G \arrow[rr] & & P \arrow[ur] \arrow[rr] & & M
\end{tikzcd}
$$

Stage 1

PMS is the response & G explanatory.

$$G \longrightarrow P$$

$G^2\,[(G,P)] = 75.26$, $df = 1$, 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.

Model	df	G^2	p	X^2	p
(GP, E)	3	48.92	$< .001$	56.77	$< .001$
(GP, PE)	2	2.91	.23	2.95	.23
(GP, GE, PE)	1	.00[a]	.98	.00[a]	.98

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.

Model	df	G^2	p
(EGP, EM, PM)	5	18.16	$< .01$
(EGP, EMP)	4	5.25	.26
(EGP, EMP, GM)	3	.70	.88

- (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).

	Program Type		
SES	Vo/Tech	General	Academic
Low	45	50	44
Middle	82	70	147
High	20	25	117

For the SES $\times$ 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., β), 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 2×2 sub-table is a direct function of the row and column scores and β.

$$\begin{aligned}
\log\left(\frac{\mu_{ij}\mu_{i'j'}}{\mu_{i'j}\mu_{ij'}}\right) &= \log(\mu_{ij}) + \log(\mu_{i'j'}) - \log(\mu_{i'j}) - \log(\mu_{ij'}) \\
&= \beta(u_i v_j + u_{i'} v_{j'} - u_{i'} v_j - u_i v_{j'}) \\
&= \beta(u_i - u_{i'})(v_j - v_{j'})
\end{aligned}$$

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:

$$u_1 = 1, \qquad u_2 = 2, \qquad u_3 = 3$$

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

$$\text{Vo/Tech} \le \text{General} \le \text{Academic}$$

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

$$v_1 = 1, \qquad v_2 = 2 \qquad v_3 = 4$$

Example of linear by linear model

$$v_1 = 1, \qquad v_2 = 2 \qquad v_3 = 4$$

Model	df	G^2	p	Δdf	ΔG^2	p
Independence	4	53.715	$< .001$	—	—	—
L by L	3	5.980	.10	1	47.74	$< .001$

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:

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

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

Class	Levels	Values
ses	3	hi low mid
hsp	3	academic general votech

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3	5.9798	1.9933
Scaled Deviance	3	5.9798	1.9933
Pearson Chi-Square	3	5.6845	1.8948
Scaled Pearson X2	3	5.6845	1.8948
Log Likelihood		2020.3156	

Algorithm converged.

SAS and Fitting the L by L model

Linear x Linear Association Model
The GENMOD Procedure
Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	3.04	0.21	2.63	3.45	216.20	$< .0001$
ses	hi	1	-1.59	0.19	-1.95	-1.22	72.06	$< .0001$
ses	low	1	0.04	0.15	-0.26	0.34	0.07	.7903
ses	mid	0	0.00	0.00	0.00	0.00	.	.
hsp	academic	1	-0.59	0.23	-1.04	-0.14	6.72	.0095
hsp	general	1	0.58	0.14	0.30	0.86	16.44	$< .0001$
hsp	votech	0	0.00	0.00	0.00	0.00	.	.
u*v		1	0.32	0.05	0.23	0.42	43.71	$< .0001$
Scale		0	1.00	0.00	1.00	1.00		

NOTE: The scale parameter was held fixed.

R and Fitting the L by L model

```
> 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: 70.670

R and Fitting the L by L model

```
> (a ← anova(independence,lin.by.lin) )
Analysis of Deviance Table
 Model 1: count ~ ses + hsp
 Model 2: count ~ ses + hsp + u * v
    Resid. Df  Resid. Dev  Df  Deviance
 1  4              53.715
 2  3               5.980   1    47.735
> 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 β. e.g.,

$$v_1 = 2, \qquad v_2 = 4 \qquad v_3 = 8$$

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.,

$$u_1 = 1, \quad u_2 = 2, \quad \ldots, u_I = I$$
$$v_1 = 1, \quad v_2 = 2, \quad \ldots, v_J = J$$

This special case of L by L model is the "Uniform Association Model."
The uniform association model for the HSB example:

Model	df	G^2	p
Independence	4	53.715	$< .01$
L by L	3	5.980	$.10$
Uniform Assoc	3	11.74	$< .01$

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.

		Item 2				
		strongly agree	agree	neither	disagree	strongly disagree
Item 1		1	2	3	4	5
strongly agree	1	97	96	22	17	2
agree	2	102	199	48	38	5
disagree	3	42	102	25	36	7
strongly disagree	4	9	18	7	10	2

GSS Results

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

$H_o : \beta = 0$ 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 $e^{\cdot 24} = 1.28$.

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

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

- ν_i^R and ν_j^C are estimated row and column scale values on the mth dimension
- ϕ_m is the association parameter

ID constraints (typical): Location

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

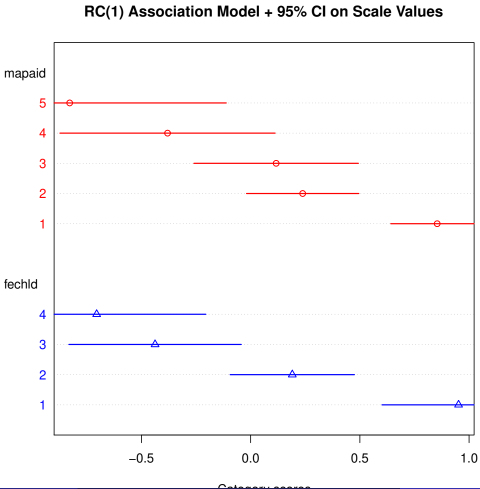
Scaling

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

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



Back to HSB

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

$$M^2 = 40.87, \qquad df = 1, \qquad p < .001, \qquad \text{and} \qquad r = .26$$

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).

Model		df	G^2	p
Independence		4	53.715	$< .01$
L by L	(unequal spacing)	3	5.980	$.10$
Uniform Assoc	(equal spacing)	3	11.74	$< .01$
RC(1) Assoc	(estimated)	1	1.74	$.19$

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.

```
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).
- Log-linear by linear for larger problems using pseudo-likelihood estimation is in the R packages plRasch (Anderson, Li & Vermunt, 2007) or more general pleLMA (Anderson, 2022).
- A SAS macro that uses pseudo-likelihood estimation (Paek & Anderson, 2016). This uses PROC MDC ("multinomial discrete choice"), which is in the econometrics package.
- "Network Psychometrics" Group at Amsterdam: http://psychonetrics.org/

Demonstration of the pleLMA Package

This package fits extensions of RC association models to low $(K > 2)$ or high dimensional tables. It it as input sections of an adjacency matrix from a graphical model and iteratively fits multinomial logistic regression models (i.e., discrete choice models).

The models arise from many different underlying processes, including IRT models.

The use of pleLMA (published on CRAN) will be demonstrated in class. A forthcoming paper on association models will be (temporarily) posted on course web-site. The package comes with an extensive vignette describing the package and how to use it.

Correspondence Analysis

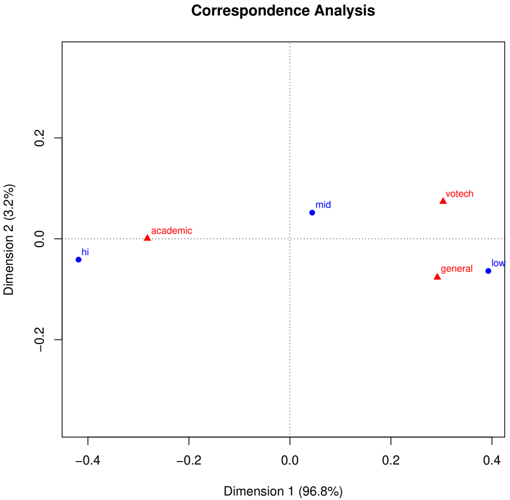
- It is a data analytic technique and it not a statistical model (i.e., no significance test).
- Provides another way to represent association between 2-variables.
- An optimal scaling procedure that decomposes Pearson's X^2 from independence.
- The scale values or scores from the 1st dimension yield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

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

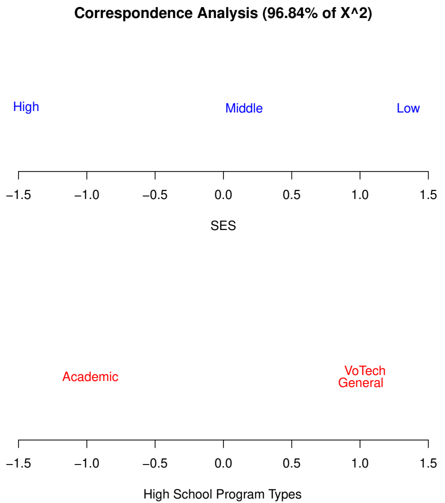
.

- 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)



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

$$H_o : \beta = 0$$

is the same as testing

$$H_o : \text{independence}$$

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

Model		df	G^2	p
Independence		4	53.715	$< .01$
Uniform Assoc	(equal spacing)	3	11.74	$< .01$
L by L	(unequal spacing)	3	5.980	.10
Nominal HSP $\times$ Ordinal SES	(equal spaced SES)	2	2.30	.32
RC association	(scores estimated)	1	1.74	.19

- The estimated parameters for the SES $\times$ HSP association in the nominal $\times$ ordinal model

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

- RC association model estimates the scores for both SES and HSP, as well as β (the "association parameter").

HSP	est. score	SES	est. score
VoTech	-.423	Low	-.669
General	-.393	Middle	-.071
Academic	.816	High	.740

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.
- Other ordinal models (Vermunt, J.K. (2001). *Sociological Methodology*).
- 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),

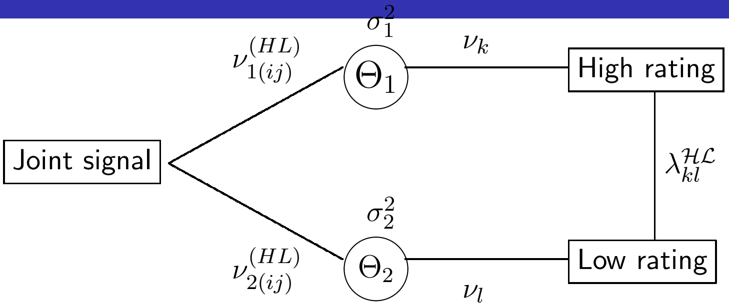
$$\log(\mu_{ijkl}) = \lambda + \lambda_i^H + \lambda_j^L + \lambda_{ij}^{HL} + \lambda_k^X + \lambda^Y + \lambda_{kl}^{XY} + \lambda_{ik}^{HX} + \lambda_{jl}^{LY} + \lambda_{ijl}^{HLY}$$

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:

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

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

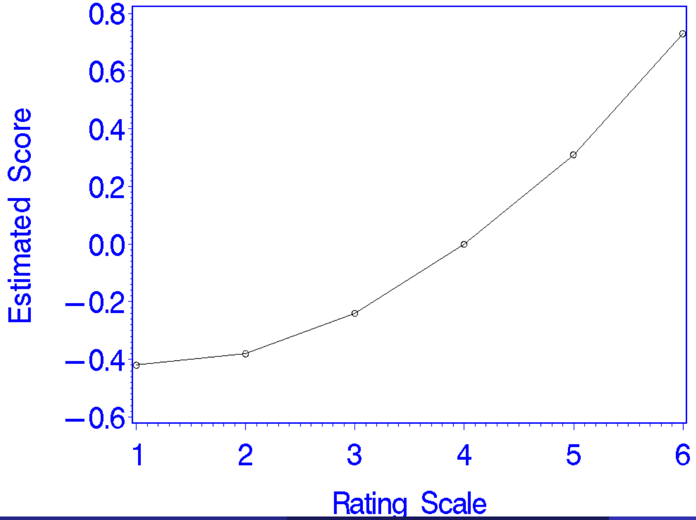


$$\log(\mu_{ijkl}) = \lambda + \lambda_{ij}^{HL} + \lambda_k^X + \lambda^Y + \lambda_{kl}^{XY} + \sigma_1^2\nu_{1(ij)}^{HL}\nu_k + \sigma_2^2\nu_{2(ij)}^{HL}\nu_l$$

$G^2 = 138.35$, $df = 98$, $p = .01$, $D = .082$.

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



Parameter Estimates— the low signal

For both subjects

Variable	Level	Parameter estimate	Standard error
rating	1	$\hat{\nu}_1^{rating} = -.42$	(.01)
rating	2	$\hat{\nu}_2^{rating} = -.38$	(.01)
rating	3	$\hat{\nu}_3^{rating} = -.24$	(.01)
rating	4	$\hat{\nu}_4^{rating} = .00$	(.02)
rating	5	$\hat{\nu}_5^{rating} = .31$	(.02)
rating	6	$\hat{\nu}_6^{rating} = .73$	(.02)
signal	high absent/low absent	$\hat{\nu}_{2(11)}^{(HL)} = -.49$	(.00)
signal	high present/low absent	$\hat{\nu}_{2(21)}^{(HL)} = -.49$	(.00)
signal	high absent/low present	$\hat{\nu}_{2(12)}^{(HL)} = .61$	(.02)
signal	high present/low present	$\hat{\nu}_{2(22)}^{(HL} = .37$	(.03)
Θ_2		$\hat{\sigma}_2^2 = 3.30$	(.12)

Parameter Estimates— the high signal

The ν's for ratings are the same as previous slide.

	Subject A			Subject B		
high absent/low absent	$\hat{\nu}^{(HL)}_{1(11)A}$	$= -.58$	(.00)	$\hat{\nu}^{(HL)}_{1(11)B}$	$= -.50$	(n.a.)
high present/low absent	$\hat{\nu}^{(HL)}_{1(21)A}$	$= .41$	(.00)	$\hat{\nu}^{(HL)}_{1(21)B}$	$= .50$	(n.a.)
high absent/low present	$\hat{\nu}^{(HL)}_{1(12)A}$	$= -.39$	(.00)	$\hat{\nu}^{(HL)}_{1(12)B}$	$= -.50$	(n.a.)
high present/low present	$\hat{\nu}^{(HL)}_{1(22)A}$	$= .58$	(.00)	$\hat{\nu}^{(HL)}_{1(22)B}$	$= .50$	(n.a.)
Θ_1	$\hat{\sigma}^2_{1A}$	$= 2.69$	(.18)	$\hat{\sigma}^2_{1B}$	$= 7.11$	(.49)

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. Comparing conditional independence model to homogeneous association model.
    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:

Variable		Type of Test		
Row	Column	Likelihood Ratio	Wald	(Generalized) CMH
Nominal	Nominal			
Nominal	Ordinal			
Ordinal	Ordinal			

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	High School Program			
SES	VoTech	General	Academic	Total
low	15	19	16	50
middle	44	30	70	144
high	12	11	56	79
Total	71	60	142	273

Males	High School Program			
SES	VoTech	General	Academic	Total
low	30	31	28	89
middle	38	40	77	155
high	8	14	61	83
Total	76	85	166	327

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.,

$$H_O : \qquad \text{all} \quad \lambda_{ij}^{XY} = 0$$

The likelihood ratio test statistic is

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

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,

$$H_O : \quad \text{all} \quad \lambda_{ij}^{SP} = 0$$

	Goodness-of-fit Test			**Likelihood Ratio Test**		
Model	df	G^2	p	Δdf	ΔG^2	p
(GS, GP, SP)	4	1.970	$.74$	—	—	—
(GS, GP)	8	55.519	$< .0001$	4	53.548	$< .001$

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 λ_{ij}^{XY}'s. For our example, the result is pretty unambiguous; that is, from SAS

Parameter	Estimate	ASE	Wald	p
λ_{lv}^{SP}	1.8133	.3233	31.450	$< .0001$
λ_{lg}^{SP}	1.6600	.3033	29.952	$< .0001$
λ_{mv}^{SP}	1.1848	.2786	18.079	$< .0001$
λ_{mg}^{SP}	.8004	.2639	9.198	$.0024$

- 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 $H_O : \lambda_{ij}^{XY} = 0$ 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,

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

The null hypothesis for this test statistic is

$$H_O: \quad \text{all} \quad \lambda_{ij}^{XY} = 0 \qquad \text{and} \quad \text{all} \quad \lambda_{ijk}^{XYZ} = 0$$

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

$$H_O: \quad \text{all} \quad \lambda_{ij}^{SP} = 0 \qquad \text{and} \quad \text{all} \quad \lambda_{ijk}^{GSP} = 0$$

	Goodness-of-fit Test			Likelihood Ratio Test		
Model	df	G^2	p	Δdf	ΔG^2	p
(GS, GP, SP)	4	1.970	.74	—	—	—
(GS, GP)	8	55.519	$< .0001$	4	53.548	$< .001$

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

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

	Goodness-of-fit Test			Likelihood Ratio Test		
Model	df	G^2	p	Δdf	ΔG^2	p
(GS, GP, SP)	4	1.970	.74	—	—	—
(GS, GP, SP)–L$\times L$	7	7.476	.38	3	5.505	.138
(GS, GP)	8	55.519	$< .0001$	1	48.043	$< .001$

From before. . .

	Goodness-of-fit Test			Likelihood Ratio Test		
Model	df	G^2	p	Δdf	ΔG^2	p
(GS, GP, SP)	4	1.970	.74	—	—	—

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

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

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

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

- 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 β 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. . .

Variable		Type of Test		
		Likelihood		(Generalized)
Row	Column	Ratio	Wald	CMH
Nominal	Nominal	X	X	
Nominal	Ordinal			
Ordinal	Ordinal	X	X	

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 β 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 β_j^P are estimated parameters.

	Goodness-of-fit Test			Likelihood Ratio Test		
Model	df	G^2	p	Δdf	ΔG^2	p
(GS,GP,SP)	4	1.970	.74	—	—	—
(GS,GP)	8	55.519	$< .0001$	4	53.548	$< .001$
(GS,GP,SP)–L$\times L$	7	7.476	.38	3	5.505	.14
(GS,GP)	8	55.519	$< .0001$	1	48.043	$< .001$
(GS,GP,SP) with u_i	6	4.076	.62	2	2.106	.35
(GS,GP)	8	55.519	$< .0001$	2	51.443	$< .001$

For the nominal–ordinal model, from SAS: $\hat{\beta}^P_{votech} = -.8784$, $\hat{\beta}^P_{gen} = -.8614$, $\hat{\beta}^P_{academic} = 0$, and from R: $\hat{\beta}^P_{votech} = -0,^P_{general} = 0.0170$ & $\hat{\beta}^P_{academic} = 0.87$
$\Longrightarrow$ the "best" scores for VoTech and General programs are much closer together

So we have now discussed,

Variable		Type of Test		
		Likelihood		(Generalized)
Row	Column	Ratio	Wald	CMH
Nominal	Nominal	X	X	
Nominal	Ordinal	X	X	
Ordinal	Ordinal	X	X	

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

Statistic	Alternative Hypothesis	df	Value	p
1	Nonzero correlation	1	46.546	$< .001$
2	Row Mean Scores Differ	2	49.800	$< .001$
3	General Association	4	51.639	$< .001$

Example: Nominal–Ordinal Generalized CMH

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

$$[1(15)+2(44)+3(12)]\,/(15+44+12)=139/71=1.96$$

	High	SES			
Gender	School	Low	Middle	High	
Gender	Program	1	2	3	Mean
Females	VoTech	15	44	12	$139/71=1.96$
	General	19	30	11	$112/60=1.87$
	Academic	16	70	56	$324/142=2.28$
Males	VoTech	30	38	8	$130/76=1.71$
	General	31	40	14	$153/85=1.80$
	Academic	28	77	61	$365/166=2.20$

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_O : all $\lambda_{ij}^{XY} = 0$ in the (XY, XZ, YZ) log-linear model.
- If the null is true, then the statistic is approximately chi–squared distributed with $df = (I-1)(J-1)$.

High School & Beyond example (all CMH tests):

Statistic	Alternative Hypothesis	df	Value	p
1	Nonzero correlation	1	46.546	$< .001$
2	Row Mean Scores Differ	2	49.800	$< .001$
3	General Association	4	51.639	$< .001$

Summary: Tests of Ordinal Association

Variable		Type of Test		
Row	Column	Likelihood Ratio	Wald	(Generalized) CMH
Nominal	Nominal	X	X	X
Nominal	Ordinal	X	X	X
Ordinal	Ordinal	X	X	X

. . . and for the curious and sake of completeness. . .

Summary: Tests of Ordinal Association

Models for SES × Gender × Program type:

Model	df	G^2	p
(GS,GP,SP)	4	1.970	.741
(GS,GP)	8	55.519	$< .0001$
(GP,SP)	6	8.532	.202
(SG,SP)*	6	3.312	.769
(GP,S)	10	62.247	$< .0001$
(GS,P)	10	57.027	$< .0001$
(G,SP)*	8	10.040	.262
(G,SP)–L × L*	11	16.221	.133
(G,P,S)	12	63.754	$< .0001$

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.

	1985			
1975	yes	no	not sampled	total
yes	175	190	230	595
no	139	1518	982	2639
not sampled	64	595	—	659
total	378	2303	1212	3893

Structurally incomplete

Survey of teenagers regarding their health concerns (Fienberg):

Health Concern	Gender: Male	Gender: Female
Sex/Reproduction	6	16
Menstrual problems	—	12
How healthy am I?	49	29
None	77	102

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):

	$Z=1$				$Z=2$				$Z=3$			
$Y=$	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

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

X	
1	2
77	64

Y			
1	2	3	4
38	16	29	58

Z		
1	2	3
43	52	46

The 2–way margins

	Y = 1	2	3	4
$X=1$	20	11	15	31
$X=2$	18	5	14	27

	Z = 1	2	3
$X=1$	20	32	25
$X=2$	23	20	21

	Y = 1	2	3	4
$Z=1$	15	0	13	15
$Z=2$	17	11	3	21
$Z=3$	6	5	13	22

- Since $n_{+21}=0$, YZ partial odds ratios involving this cell equal 0 or $+\infty$.
- The YZ margin has a zero $\longrightarrow$ no MLE estimate of λ_{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.

$N = 178$

		Mention			
		No		Yes	
		Concern		Concern	
Solve	Manage	No	Yes	No	Yes
No	No	38	0	3	0
	Yes	51	4	16	26
Yes	No	0	0	0	0
	Yes	2	1	21	17

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.

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,

$$\hat{\lambda}_{21}^{YZ} = -23.9833, \qquad \text{ASE} = 87,417.4434$$

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

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

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

R also have ridiculously large S.E.s, i.e., $\hat{\mu}$s for these cells ± 29 and $se = 3,966.26$

Sparseness & Odds Ratio Estimates

- Sparseness 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 ∞ to finite values and increases estimates that are 0.
- Qualifications: For unsaturated models, adding $.5$ will over smooth the data.
- Remedies/Strategies/Comments. . .

Sparseness & 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

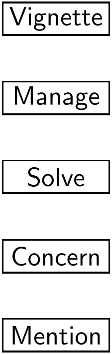
and the (X, YZ) loglinear model:

Number added	G^2	X^2	Converge?	ASE for $\hat{\lambda}_{21}^{YZ}$
—	16.86	13.38	no	87,417.44
0.00000001	15.43	17.92	yes	7,071.07
0.000001	16.83	13.37	yes	223.61
0.0001	16.87	13.38	yes	22.37
0.1	18.86	13.78	yes	2.30

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

Log-Linear Models & Empathy Data

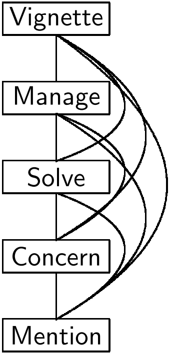
Independence



$G^2 = 137.46$

$df = 26,\ p < .01$

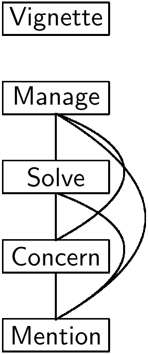
All 2-way interactions



$G^2 = 13.69$

$df = 16,\ p = .62$

Unrelated to Vignette



$G^2 = 15.24$

$df = 29,\ p = .76$

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 $G^2(M_O) - G^2(M_1)$ 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 large.
- Exact tests and exact analyses for models.
- An alternative test statistic: the Cressie-Read statistic

Cressie-Read statistic

Cressie, N, & Read, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, 43*, 440–464.

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 λ defines a specific statistic (note: λ 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

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 anomolous cells.
2. Excluding "problem" sampling zeros from an analysis.
3. Check collapsibility across *categories* of a variable.
4. Quasi–independence.
5. Symmetry and quasi–symmetry.
6. Marginal homogeneity.
7. Bradley–Terry–Luce model for paired comparisons.
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,

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

  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:

Health Concern	Gender Male	Gender Female
Sex/Reproduction	6	16
Menstrual problems	—	12
How healthy am I?	49	29
None	77	102

- 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

$$
\begin{aligned}
I(2,1) &= 1 \quad \text{for the (2,1) cell} \\
&= 0 \quad \text{otherwise}
\end{aligned}
$$

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 δ_{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 λ_{ij}^{HG} term from the model, but we still include the indicator variable for the $(2,1)$ cell.

What happens to degrees of freedom?

$$
\begin{aligned}
df &= (\#\text{ of cells}) - (\#\text{ non-redundant parameters}) \\
&= (\text{usual } df \text{ for the model}) - (\#\text{ cells fit perfectly})
\end{aligned}
$$

Independence: Teens and Health Concerns

$$\begin{aligned} df &= (I-1)(J-1)-1 \\ &= (4-1)(2-1)-1 = 2 \end{aligned}$$

$G^2 = 12.60$, and $X^2 = 12.39$, 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

$$\begin{array}{rrcrr} \hat{\lambda} = & 4.5466 & & & \\ \hat{\lambda}^H_1 = & -2.0963 & & \hat{\lambda}^G_1 = & -1.1076 \\ \hat{\lambda}^H_2 = & -2.0671 & & \hat{\lambda}^G_1 = & 0.0000 \\ \hat{\lambda}^H_3 = & -0.8307 & & & \\ \hat{\lambda}^H_4 = & 0.0000 & & \hat{\delta}_{21} = & -22.9986 \end{array}$$

For the $(2,1)$ cell,

$$\hat{\mu}_{21} = \exp(4.5466 - 2.0671 - .1076 - 22.9986) \sim 0$$

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.

  . . . 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).

	1953		1971	
Religion	Boys	Both	Boys	Both
Protestant	104	42	165	142
Catholic	65	44	100	130
Jewish	4	3	5	6
Other	13	6	32	23

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:

Model	df	G^2	p	X^2	p
(RY,G)	7	31.67	$< .001$	31.06	$< .001$
(RY,GY)	6	11.25	.08	11.25	.08
(RY,GR)	4	21.49	$< .001$	21.12	$< .001$
(RY,GY,GR)	3	0.36	.95	.36	.95

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

	1953		1971	
Religion	Boys	Both	Boys	Both
Protestant	.75	-1.05	.91	-.91
Catholic	-.84	1.18	-1.42	1.42
Jewish	-.29	.41	-.22	.22
Other	.12	-.17	.85	-.85

The 3 largest residuals $\longrightarrow$ 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)

$$\begin{aligned}\log(\mu_{ijk}) &= \lambda + \lambda_i^R + \lambda_j^Y + \lambda_k^G + \lambda_{ij}^{RY} + \lambda_{jk}^{GY} \\ &\quad + \delta_{212} I(2,1,2) + \delta_{221} I(2,2,1) + \delta_{222} I(2,2,2)\end{aligned}$$

where the indicator variables are defined as

$$\begin{aligned} I(2,j,k) &= 1 \quad \text{if Catholic and} \quad j \neq 1 \quad \text{and} \quad k \neq 1 \\ &= 0 \quad \text{otherwise}\end{aligned}$$

- 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.

1in

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

Holland Code	Gender Men	Women	Total
Realistic	13	1	14
Investigative	31	24	55
Artistic	2	2	4
Social	1	24	25
Enterprising	2	1	3
Conventional	3	9	12

Farmer, Rotella, Anderson & Wardrop

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,

Adjusted residuals

Holland Code	Independence
Realistic	-3.76
Investigative	-2.15
Artistic	$-.16$
Social	4.78
Enterprizing	$-.73$
Conventional	1.55

Two largest: Realistic and Social.
$\longrightarrow$ fit these perfectly but allow independence in the rest of the table,...

Farmer, Rotella, Anderson & Wardrop

Realistic and Social.

$\longrightarrow$ 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 $H_O : \beta^{\text{social}} = -\beta^{\text{realistic}} = \beta$.

Comparing Adjusted Residuals

The adjusted residuals look pretty good for new model

	Model	
Holland Code	Independence	Association
Realistic	-3.76	$.37$
Investigative	-2.15	-0.86
Artistic	$-.16$	$.02$
Social	4.78	$.27$
Enterprizing	$-.73$	$-.56$
Conventional	1.55	1.77

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 $=$ 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).

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

Four Religion Variables				1953		1971	
Protestant	Catholic	Jewish	Other	Boy	Both	Boy	Both
1	1	1	1	—	—	—	—
1	1	1	0	—	—	—	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0	0	0	104	42	165	142
0	1	1	1	—	—	—	—
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	1	0	0	65	44	100	130
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	1	0	4	3	5	6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0	0	0	1	13	6	32	23

Models for Snow Shovelling

Since G (gender) is considered the response and all log-linear models must include λ^{YPCJO} terms (and lower order ones).

Here are some of the fit to the data models.

Model	df	G^2
Fit previously		
(YPCJO,GY)	6	11.2
(YPCJO,GY,GPCJO)	3	0.4
New ones		
(YPCJO,GY,GO)	5	9.8
(YPCJO,GY,GJ)	5	10.9
(YPCJO,GY,GC)	5	1.4
(YPCJO,GY,GP)	5	4.8

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)