

# Chapters 9 and 10: Log-linear models

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 770: Categorical Data Analysis

## Prologue: log-linear models for $2 \times 2$ table

Assume  $X \in \{1, 2\}$  and  $Y \in \{1, 2\}$ , so the table has 4 cells:

	$Y = 1$	$Y = 2$
$X = 1$	$n_{11}$	$n_{12}$
$X = 2$	$n_{21}$	$n_{22}$

Assume *multinomial* sampling so

$$\mathbf{n} = (n_{11}, n_{12}, n_{21}, n_{22}) \sim \text{mult}\{n, \mathbf{p} = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})\}.$$

We write this  $\{n_{ij}\} \sim \text{mult}(n, \{\pi_{ij}\})$  for short.

Let's examine the additive model for this table in some detail...

## Log-linear models for $2 \times 2$ table

The additive model for  $E(n_{ij}) = n\pi_{ij}$  is

$$\log(n\pi_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y.$$

We set  $\lambda_2^X = \lambda_2^Y = 0$  for identifiability. Then the cell means are

	$Y = 1$	$Y = 2$
$X = 1$	$e^{\lambda + \lambda_1^X + \lambda_1^Y}$	$e^{\lambda + \lambda_1^X}$
$X = 2$	$e^{\lambda + \lambda_1^Y}$	$e^\lambda$

Under multinomial sampling  $\lambda$  is redundant and known through

$$\frac{e^{\lambda + \lambda_1^X + \lambda_1^Y}}{n} + \frac{e^{\lambda + \lambda_1^X}}{n} + \frac{e^{\lambda + \lambda_1^Y}}{n} + \frac{e^\lambda}{n} = 1.$$

That is

$$\lambda = \log n - \log \left\{ e^{\lambda_1^X + \lambda_1^Y} + e^{\lambda_1^X} + e^{\lambda_1^Y} + 1 \right\}.$$

Under the additive model,

$$\begin{aligned}\theta &= \frac{P(Y = 2|X = 2)/P(Y = 1|X = 2)}{P(Y = 2|X = 1)/P(Y = 1|X = 1)} \\ &= \frac{P(Y = 2, X = 2)/P(Y = 1, X = 2)}{P(Y = 2, X = 1)/P(Y = 1, X = 1)} \\ &= \frac{e^\lambda / e^{\lambda + \lambda_1^Y}}{e^{\lambda + \lambda_1^X} / e^{\lambda + \lambda_1^X + \lambda_1^Y}} = 1.\end{aligned}$$

This proves  $X \perp Y$ .

There are only two parameters in the model:  $\lambda_1^X$  and  $\lambda_1^Y$  to estimate three free probabilities in  $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ .

## Log-linear models for $2 \times 2$ table

Under the **interaction** model we have

$$\log(n\pi_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY},$$

where  $\lambda_{12}^{XY} = \lambda_{22}^{XY} = \lambda_{21}^{XY} = 0$ . This adds one more non-zero parameter to the model  $\lambda_{11}^{XY}$  for a total of three. There are only three degrees of freedom in the table for  $(n_{11}, n_{12}, n_{21}, n_{22})$  and thus the model is saturated; three parameters  $\lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY}$  to estimate three free probabilities in  $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ . Then

$$\begin{aligned}\theta &= \frac{P(Y = 2, X = 2)/P(Y = 1, X = 2)}{P(Y = 2, X = 1)/P(Y = 1, X = 1)} \\ &= \frac{e^\lambda / e^{\lambda + \lambda_1^Y}}{e^{\lambda + \lambda_1^X} / e^{\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}}} = e^{\lambda_{11}^{XY}}.\end{aligned}$$

The interaction term is a simple function of the odds ratio. We see that  $X \perp Y$  iff  $\lambda_{11}^{XY} = 0$  iff  $\lambda_{ij}^{XY} = 0$  for all  $i, j$ .

# Example

Subtable of Table 2.1 (p. 38):

	Fatal	Nonfatal
Placebo	18	171
Aspirin	5	99

SAS code:

```
data table;
  input Treat$ Outcome$ count @@;
  datalines;
  1 1 18  1 2 171  2 1 5  2 2 99
;
proc format;
  value $tc '1'='Placebo' '2'='Aspirin';
  value $oc '1'='Fatal' '2'='Nonfatal';
proc freq order=data; weight count;
  format Treat $tc. Outcome $oc.;
  tables Treat*Outcome / norow nocol nopercnt expected;
  exact chisq or;
proc genmod order=data; class Treat Outcome;
  model count = Treat Outcome Treat*Outcome / dist=poi link=log;
```

The FREQ Procedure

Table of Treat by Outcome

Treat	Outcome		
Frequency			
Expected	Fatal	Nonfatal	Total
Placebo	18	171	189
	14.836	174.16	
Aspirin	5	99	104
	8.1638	95.836	
Total	23	270	293

Statistics for Table of Treat by Outcome

Statistic	DF	Value	Prob
Chi-Square	1	2.0627	0.1509
Likelihood Ratio Chi-Square	1	2.2173	0.1365

## Pearson Chi-Square Test

```
-----  
Chi-Square                2.0627  
DF                        1  
Asymptotic Pr > ChiSq    0.1509  
Exact      Pr >= ChiSq    0.1782
```

## Odds Ratio (Case-Control Study)

```
-----  
Odds Ratio                2.0842
```

## Asymptotic Conf Limits

```
95% Lower Conf Limit      0.7506  
95% Upper Conf Limit      5.7872
```

## Exact Conf Limits

```
95% Lower Conf Limit      0.7151  
95% Upper Conf Limit      7.3897
```

Sample Size = 293

## Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.5951	0.1005	4.3981	4.7921	2090.40	<.0001
Treat	1	1	0.5465	0.1263	0.2990	0.7941	18.73	<.0001
Treat	2	0	0.0000	0.0000	0.0000	0.0000	.	.
Outcome	1	1	-2.9857	0.4584	-3.8841	-2.0873	42.43	<.0001
Outcome	2	0	0.0000	0.0000	0.0000	0.0000	.	.
Treat*Outcome	1 1	1	0.7344	0.5211	-0.2869	1.7557	1.99	0.1587
Treat*Outcome	1 2	0	0.0000	0.0000	0.0000	0.0000	.	.
Treat*Outcome	2 1	0	0.0000	0.0000	0.0000	0.0000	.	.
Treat*Outcome	2 2	0	0.0000	0.0000	0.0000	0.0000	.	.

As promised,  $e^{0.7344} = 2.0842$  with CI  $(e^{-0.2869}, e^{1.7557}) = (0.7506, 5.7875)$ . We also obtain the  $p$ -value for the Wald test of  $H_0 : \lambda_{11}^{XY} = 0$  in the saturated model, 0.1587, slightly different than the Pearson or LRT tests obtained from PROC FREQ.

Now let's look at  $I \times J$  tables...

A log-linear model is a Poisson model with ANOVA structure for the log-means of counts in a contingency table.

- We start with  $I \times J$  tables and then consider multiway, e.g.  $I \times J \times K \times L$  tables.
- Useful to determine conditional dependence relationships between variables.
- Can be generalized to non-categorical predictors.
- No one categorical variable is the outcome.

## 8.1 Two-way models

Let  $n_{ij}$  be the counts in an  $I \times J$  contingency table.

	$Y = 1$	$Y = 2$	$\dots$	$Y = J$
$X = 1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1J}$
$X = 2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2J}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X = I$	$n_{I1}$	$n_{I2}$	$\dots$	$n_{IJ}$

The random, total number in the table is  $n_{++} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ .  
We assume that each cell in the table is independent Poisson,

$$n_{ij} \stackrel{ind.}{\sim} \text{Poisson}(\mu_{ij}).$$

Different parameterizations for  $\mu_{ij}$  lead to different distributions for  $(X, Y)$ .

The  $\mu_{11}, \mu_{12}, \dots, \mu_{IJ}$  are the rates at which the  $(X, Y)$  fall into the cross-classified categories.

Let  $(X, Y)$  be a pair of nominal or ordinal outcomes with  $X \in \{1, \dots, I\}$  and  $Y \in \{1, \dots, J\}$ . We will collect  $n$  such pairs *iid* from the population:  $(X_1, Y_1), \dots, (X_n, Y_n)$ .

Let  $n_{ij} = \sum_{k=1}^n I\{X_k = i, Y_k = j\}$  be the number of pairs  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$  that fall into the  $i^{\text{th}}$  category of  $X$  and the  $j^{\text{th}}$  category of  $Y$ .

We assume that data are collected over time and that the  $n_{ij}$  are independent Poisson random variables with means  $\mu_{ij}$ . At any time we can stop the collection process and have a snapshot of the contingency table at that time. For example, if  $n = n_{++} = 1000$  people are sampled and cross-classified, we have a snapshot after  $n = 1000$  individuals are sampled.

# Poisson is equivalent to multinomial sampling

We know  $\{n_{ij}\}$  is distributed as  $I \times J$  independent Poisson variables. But if we stop collecting data when  $n_{++} = n$ , what is the distribution? Recall that the sum of independent Poisson random variables is also Poisson with a rate that is the sum of the individual rates. So  $n \sim \text{Pois}(\sum_{i,j} \mu_{ij})$ .

$$\begin{aligned} p(n_{11}, \dots, n_{IJ} | n_{++} = n) &= \frac{p(n_{11}, \dots, n_{IJ}) I\{n_{++} = n\}}{P(n_{++} = n)} \\ &= \frac{I\{n_{++} = n\} \prod_{i,j} \frac{e^{-\mu_{ij}} \mu_{ij}^{n_{ij}}}{n_{ij}!}}{\frac{e^{-\sum_{i,j} \mu_{ij}} [\sum_{i,j} \mu_{ij}]^{\sum_{i,j} n_{ij}}}{[\sum_{i,j} n_{ij}]!}} \\ &= \binom{n}{n_{11} \cdots n_{IJ}} \prod_{i,j} \left[ \frac{\mu_{ij}}{\mu_{++}} \right]^{n_{ij}}. \end{aligned}$$

# Poisson is equivalent to multinomial sampling

This pmf, subject to  $n_{++} = n$ , is a multinomial distribution with parameters  $n$  and  $\mathbf{p} = (\mu_{11}/\mu_{++}, \dots, \mu_{IJ}/\mu_{++})$ .

Put another way, Poisson sampling is equivalent to multinomial sampling where at any time such that  $n_{++} = n$ ,  
 $\pi_{ij} = P(X = i, Y = j) = \mu_{ij}/\mu_{++}$ .

Thus, **fitting a Poisson model for the  $(\mu_{11}, \dots, \mu_{IJ})$  conditional on  $n_{++} = n$  is the same as fitting the multinomial.**

We will fit log-linear models using the Poisson distribution in PROC GENMOD.

## 9.1.1 Independence model

The *independence model* (also see Section 4.3.7) stipulates

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y.$$

For identifiability, we must place restrictions on the parameters, e.g.  $\lambda_I^X = \lambda_J^Y = 0$ . Then there are  $(I - 1) + (J - 1) + 1 = I + J - 1$  parameters to estimate:  $(\lambda_1^X, \dots, \lambda_{I-1}^X, \lambda_1^Y, \dots, \lambda_{J-1}^Y, \lambda)$ .

Note that conditional on  $n$ , we have multinomial sampling and  $\mu_{ij} = e^\lambda e^{\lambda_i^X} e^{\lambda_j^Y} = n\pi_{i+}\pi_{+j}$ . That is, the intercept term  $\lambda$  adjusts the overall mean  $\mu_{++}$  in the Poisson model and is a function of  $n$  as well as the other model parameters. *However*, it is not true that  $e^\lambda = n$ ,  $e^{\lambda_i^X} = \pi_{i+}$  and  $e^{\lambda_j^Y} = \pi_{+j}$ .

# Independence model

In fact, we know that  $n_{++} \sim \text{Poisson}(\mu_{++})$  and that the MLE of this is  $\hat{\mu}_{++} = n_{++} = n$ . So we must have

$$n = \sum_{i=1}^I \sum_{j=1}^J e^{\hat{\lambda}} e^{\hat{\lambda}_i^X} e^{\hat{\lambda}_j^Y}.$$

So,

$$\hat{\lambda} = \log n - \log \sum_{i=1}^I \sum_{j=1}^J e^{\hat{\lambda}_i^X + \hat{\lambda}_j^Y}.$$

Under multinomial sampling (conditional on  $n_{++} = n$ ) the number of parameters  $(\lambda_1^X, \dots, \lambda_{I-1}^X, \lambda_1^Y, \dots, \lambda_{J-1}^Y)$  drops by 1, because  $\lambda$  is known, to  $(I-1) + (J-1)$ .

Conditional on  $n$ , the model satisfies  $\pi_{ij} = \pi_{i+}\pi_{+j}$ .

# Five hierarchical models

Model	Interpretation
$\log \mu_{ij} = \lambda$	$X \perp Y, \pi_{ij} = \pi$
$\log \mu_{ij} = \lambda + \lambda_i^X$	$X \perp Y, \pi_{ij} = \pi_i$
$\log \mu_{ij} = \lambda + \lambda_j^Y$	$X \perp Y, \pi_{ij} = \pi_j$
$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$	$X \perp Y, \pi_{ij} = \pi_i + \pi_j$
$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$	$X \not\perp Y$

We are typically only interested in the last two, as a means to test  $H_0 : X \perp Y$  versus  $H_1 : X \not\perp Y$ . This boils down to testing  $H_0 : \lambda_{ij}^{XY} = 0$  in the full interaction model.

The interaction model is given by

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

where  $\lambda_i^X = 0$ ,  $\lambda_j^Y = 0$ , and  $\lambda_{ij}^{XY} = \lambda_{ij}^{XY} = 0$  for  $i = 1, \dots, I$  and  $j = 1, \dots, J$ . So there are  $(I - 1) + (J - 1) + (I - 1)(J - 1) = IJ - 1$  parameters to estimate in the multinomial interaction model, *one for each cell*.

## Testing independence, example

The LRT for independence from Chapter 3 is equivalent to testing the additive (most flexible independence model) to the interaction model in the Poisson GLM framework. The difference in parameters is

$(I-1) + (J-1) + (I-1)(J-1) - [(I-1) + (J-1)] = (I-1)(J-1)$   
as we found before. Let's examine an example in detail.

From Chapter 2 in Christensen (1997) we have a sample of  $n = 52$  males with ages from 11 to 30 with knee operations via arthroscopic surgery. They are cross-classified according to  $X = 1, 2, 3$  for injury type (twisted knee, direct blow, or both) and  $Y = 1, 2, 3$  for surgical result (excellent, good, or fair-to-poor).

$n_{ij}$	Excellent	Good	Fair to poor	Totals
Twisted knee	21	11	4	36
Direct blow	3	2	2	7
Both types	7	1	1	9
Totals	31	14	7	$n = 52$

with theoretical probabilities:

$\pi_{ij}$	Excellent	Good	Fair to poor	Totals
Twisted knee	$\pi_{11}$	$\pi_{12}$	$\pi_{13}$	$\pi_{1+}$
Direct blow	$\pi_{21}$	$\pi_{22}$	$\pi_{23}$	$\pi_{2+}$
Both types	$\pi_{31}$	$\pi_{32}$	$\pi_{33}$	$\pi_{3+}$
Totals	$\pi_{+1}$	$\pi_{+2}$	$\pi_{+3}$	$\pi_{++} = 1$

SAS code:

```
data table;
  input Injury$ Result$ count @@;
  datalines;
1 1 21 1 2 11 1 3 4 2 1 3 2 2 2 2 3 2 3 1 7 3 2 1 3 3 1
;
proc format;
value $ic '1'='twisted' '2'='direct blow' '3'='both';
value $rc '1'='excellent' '2'='good' '3'='fair-to-poor';
proc freq order=data; weight count;
  format Injury $ic. Result $rc.;
  tables Injury*Result / chisq;
proc genmod order=data; class Injury Result;
  model count = Injury Result / dist=poi link=log;
```

# Output from PROC FREQ

Injury	Result			Total
	Frequency Percent	excellen t	good	
twisted	21 40.38	11 21.15	4 7.69	36 69.23
direct blow	3 5.77	2 3.85	2 3.85	7 13.46
both	7 13.46	1 1.92	1 1.92	9 17.31
Total	31 59.62	14 26.92	7 13.46	52 100.00

Statistics for Table of Injury by Result

Statistic	DF	Value	Prob
Chi-Square	4	3.2288	0.5203
Likelihood Ratio Chi-Square	4	3.1732	0.5293

# Output from PROC GENMOD

## Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	3.1732	0.7933
Scaled Deviance	4	3.1732	0.7933
Pearson Chi-Square	4	3.2288	0.8072
Scaled Pearson X2	4	3.2288	0.8072
Log Likelihood		61.9602	

## Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	0.1919	0.4845	-0.7577	1.1415	0.16	0.6921
Injury	1	1.3863	0.3727	0.6559	2.1167	13.84	0.0002
Injury	2	-0.2513	0.5040	-1.2390	0.7364	0.25	0.6180
Injury	3	0.0000	0.0000	0.0000	0.0000	.	.
Result	1	1.4881	0.4185	0.6679	2.3083	12.65	0.0004
Result	2	0.6931	0.4629	-0.2141	1.6004	2.24	0.1343
Result	3	0.0000	0.0000	0.0000	0.0000	.	.

- Pearson and LRT test statistics and  $df$  for independence from PROC FREQ are the same as the GOF tests of the additive model versus the *saturated* interaction model from PROC GENMOD fitting the Poisson models.
- $P(\chi_4^2 > 3.1732) = 0.5293$ ; compare to PROC FREQ.
- $\hat{\lambda} = 0.1919 = \log 52 - \log \sum_{i=1}^3 \sum_{j=1}^3 e^{\hat{\lambda}_i^X + \hat{\lambda}_j^Y}$  from the last 6 rows of the SAS GENMOD Analysis of Parameter Estimates.
- We accept that  $X \perp Y$ , i.e. that  $\pi_{ij} = \pi_{i+}\pi_{+j}$ .
- Testing whether we can drop either Result or Injury from the model significantly increases the difference in  $-2$  times the log-likelihood (on 2  $df$  for either test) and we reject the simpler models.

## 9.2 Three-way $I \times J \times K$ tables

Now we have  $n$  individuals cross-classified on three variables  $(X, Y, Z)$ . Let  $n_{ijk}$  be the number out of  $n = n_{+++}$  that are classified  $X = i$ ,  $Y = j$ , and  $Z = k$ . We assume

$$\{n_{ijk}\} \sim \text{mult}(n, \{\pi_{ijk}\}).$$

As before, including ANOVA parameters for the log-mean  $E(n_{ijk}) = n\pi_{ijk}$  will force certain types of dependence among  $(X, Y, Z)$ . The saturated model is

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

with the usual constraints. There are  $IJK - 1$  free parameters in the model to estimate  $IJK - 1$  free probabilities in the table. Shorthand:  $[XYZ]$ .

# Independence model

## 1. $[X][Y][Z]$

The **additive** model is

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

The additive model implies complete independence:

$$P(X = i, Y = j, Z = k) = P(X = i)P(Y = j)P(Z = k),$$

i.e.

$$\pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k}.$$

The shorthand for this model is  $[X][Y][Z]$ .

A test of the additive model versus the saturated model tests  $H_0 : X \perp Y \perp Z$ .

However, there are a number of models (7 total) between the additive (mutual independence) model and the saturated model, each implying a unique dependency structure among  $(X, Y, Z)$ .

## 2. $[XY][Z]$ , 3. $[XZ][Y]$ , or 4. $[YZ][X]$

There are three ways that one variable can be independent of the remaining two:  $(X, Y) \perp Z$ ,  $(X, Z) \perp Y$ , or  $(Y, Z) \perp X$ . These have shorthand  $[XY][Z]$ ,  $[XZ][Y]$ , or  $[YZ][X]$  respectively. These models imply  $\pi_{ijk} = \pi_{ij+}\pi_{++k}$ ,  $\pi_{ijk} = \pi_{i+k}\pi_{j+}$ , or  $\pi_{ijk} = \pi_{+jk}\pi_{i++}$  and have log-linear model representation:

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY},$$

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ},$$

or

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}.$$

$[XY][Z]$  implies

$$P(X = i, Y = j, Z = k) = P(X = i, Y = j)P(Z = k), \text{ etc.}$$

## 5. $[XZ][YZ]$ , 6. $[XY][ZY]$ , or 7. $[YX][ZX]$

There are three ways that two variables can be independent conditional on the other one:  $X \perp Y|Z$ ,  $X \perp Z|Y$ , or  $Y \perp Z|X$ .

These have shorthand  $[XZ][YZ]$ ,  $[XY][ZY]$ , or  $[YX][ZX]$  respectively. These models imply

$$P(X = i, Y = j|Z = k) = P(X = i|Z = k)P(Y = j|Z = k),$$

$$P(X = i, Z = k|Y = j) = P(X = i|Y = j)P(Z = k|Y = j), \text{ or}$$

$$P(Y = j, Z = k|X = i) = P(Y = j|X = i)P(Z = k|X = i) \text{ and}$$

have log-linear model representation:

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ},$$

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ},$$

or

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}.$$

# Hard-to-interpret models

Note that the shorthand summarizes the highest-order interactions included in the model as well as the dependence structure. This leaves two last models:

8.  $[XY][XZ][YZ]$  given by

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ},$$

and the **saturated** model

9.  $[XYZ]$  given by

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

Both of these imply rather complex dependency structures. Please see pp. 344-346 (more at end of notes). Models 1-7 yield simplified dependency structure for  $(X, Y, Z)$  and are preferred if one or more fit.

# Choosing among log-linear models is an art

- Many contingency tables will have many, sometimes mostly, empty or near-empty cells. The asymptotics involved in testing reduced models relative to the saturated model are then tenuous at best.
- Testing reduced models to (non-saturated) higher-order interaction models is a bit safer. *Browns tests of association* are a useful tool to find higher-order models from which to start from. See paper posted on course website if interested .
- An *ad hoc* but useful approach is to find models that minimize the AIC and check “winning” model fit through a residual analysis. That’s what we will do here.

# Example

$n = 2121$  individuals did not exercise or have cardiovascular disease during a  $4\frac{1}{2}$  year study. They are cross-classified below according to personality type A (e.g. workaholics) or B (e.g. relaxed, slouchy graduate students), cholesterol level normal or high, and diastolic blood pressure normal or high. Lets call these factors  $P$ ,  $C$ , and  $B$ .

Personality	Cholesterol	Diastolic blood pressure	
		Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

SAS code:

```
data drugs;
input type chol bp count @@;
datalines;
1 1 1 716 1 1 2 79
1 2 1 207 1 2 2 25
2 1 1 819 2 1 2 67
2 2 1 186 2 2 2 22
;
proc genmod order=data; class type chol bp;
model count = type|chol|bp / dist=poi link=log type3;
```

## LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
type	1	0.56	0.4544
chol	1	238.24	<.0001
type*chol	1	0.33	0.5642
bp	1	1109.42	<.0001
type*bp	1	0.82	0.3665
chol*bp	1	1.62	0.2029
type*chol*bp	1	0.61	0.4336

There are plenty of observations in each cell and a test of the saturated model versus  $[PC][PB][CB]$  should be approximately valid. Here we reject that the 3-way interaction is necessary to model dependence and accept the model  $[PC][PB][CB]$ . Let's refit this model via `model count = type|chol type|bp chol|bp / dist=poi link=log type3;`

Source	DF	Chi-Square	Pr > ChiSq
type	1	1.35	0.2458
chol	1	241.43	<.0001
type*chol	1	3.95	0.0469
bp	1	1114.32	<.0001
type*bp	1	2.37	0.1240
chol*bp	1	1.45	0.2286

We can further drop  $[CB]$  and so we fit `model count = type|chol type|bp / dist=poi link=log type3;`

Source	DF	Chi-Square	Pr > ChiSq
type	1	1.46	0.2269
chol	1	772.43	<.0001
type*chol	1	4.12	0.0423
bp	1	1645.33	<.0001
type*bp	1	2.54	0.1111

The  $p$ -value for dropping  $[PB]$  is 0.11, a bit too close to 0.05 for comfort. I'll stop here and accept the model  $[PC][PB]$ . We accept that given personality type A or B, cholesterol level is independent of blood pressure in this study population. Put another way, personality type has all the information about blood pressure in it; nothing is to be gained from knowing the cholesterol level. In fact, we can *collapse* the table over cholesterol level if we want without loss of information concerning the relationship between blood pressure and personality or worries about Simpson's paradox – more on this next time.

If we had *accepted* we could drop  $[PB]$  from the model, then the final model would be  $[PC][B]$ , blood pressure is independent of the other two, a much stronger assertion.

## Higher order tables

All of these ideas generalize to higher order tables. A particular (hierarchical) log-linear model corresponds to a dependence structure among factors in the table. The shorthand for the association involves the highest order interactions needed for reasonable fit in the model. For example, say we have factors  $A, B, C, D$  and the following model fits:

$$\log(n\pi_{ijkl}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{jk}^{BD} + \lambda_{kl}^{CD}.$$

The shorthand is  $[A][BD][CD]$ .  $A$  is independent of the other three and  $B \perp C|D$ .

## 9.4.2: Seat belt example (pp. 351–352)

$n = 68694$  passengers in autos and light trucks involved in accidents in Maine in 1991.

Gender	Location	Seat belt	Injury	
			No	Yes
Female	Urban	No	7287	996
		Yes	11587	759
	Rural	No	3246	973
		Yes	6134	757
Male	Urban	No	10381	812
		Yes	10969	380
	Rural	No	6123	1084
		Yes	6693	513

Fitting the model with all four 3-way interactions yields a  $p$ -value for  $[GBI]$  of 0.84. Replacing this term with  $[GB][GI][BI]$  yields

Source	DF	Chi-Square	Pr > ChiSq
g	1	1.86	0.1725
l	1	292.60	<.0001
g*l	1	86.24	<.0001
b	1	49.79	<.0001
g*b	1	864.76	<.0001
l*b	1	3.78	0.0519
g*l*b	1	15.19	<.0001
i	1	47313.0	<.0001
g*i	1	405.58	<.0001
l*i	1	736.58	<.0001
g*l*i	1	2.22	0.1358
b*i	1	898.90	<.0001
l*b*i	1	3.12	0.0772

## Model fitting...

So we replace  $[GLI]$  with  $[GL][GI][LI]$  and obtain:

Source	DF	Chi-Square	Pr > ChiSq
g	1	1.51	0.2186
l	1	309.33	<.0001
g*l	1	181.34	<.0001
b	1	49.79	<.0001
g*b	1	869.47	<.0001
l*b	1	3.31	0.0690
g*l*b	1	17.04	<.0001
i	1	47612.6	<.0001
l*i	1	735.91	<.0001
g*i	1	404.72	<.0001
b*i	1	900.36	<.0001
l*b*i	1	3.87	0.0491

The deviance from this model is 3.59 on 3 *df* yielding a *p*-value of 0.31. The model is  $[LBI][GLB][GI]$ . This model has no simple conditional independence interpretation, but rather is interpretable in terms of odds ratios; we'll explore this later.

This approach to model selection uses backwards elimination from a fairly complex model. The model with all four 3-way interactions is just one degree of freedom away from the saturated model. We will discuss methods for assessing fit shortly, namely residuals.

## Let's revisit alligator food data

Let's reexamine the alligator food preference data. Call the factors  $F$ ,  $S$ ,  $L$ , and  $G$  for food, size, lake, and gender. The model with all 4 3-way interactions crashes the program (separation occurs). A bit of model building yields the following:

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
lake	3	2.88	0.4105
gender	1	9.88	0.0017
lake*gender	3	17.72	0.0005
size	1	2.80	0.0945
lake*size	3	4.14	0.2465
gender*size	1	23.85	<.0001
lake*gender*size	3	27.02	<.0001
food	4	85.71	<.0001
lake*food	12	49.13	<.0001
size*food	4	21.09	0.0003

This gives the model  $[GLS][LS][SF][LF]$  and the interpretation  $G \perp F|L, S$ . Males and females eat similarly within a lake and size category.

On model building:

- Brown's tests of association (not discussed) give large models to start backwards elimination from. BMDP implements these.
- Another approach is to try backward elimination from models with all higher  $k$ -way interactions (e.g. 3-way).
- $G^2$  is model deviance, the drop in  $-2 \log \mathcal{L}$  from reduced model to saturated model; Agresti uses  $G^2$  for model building.

## 10.3.1: Diagnostics

Let's consider  $I \times J \times K$  tables for illustration. The ideas immediately generalize.

A table has observed cell counts  $n_{ijk}$  and predicted under the model  $n\hat{\pi}_{ijk}$  where  $\pi_{ijk}$  is given by, e.g.,

$$\log(n\pi_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ},$$

for model  $[XY][XZ]$ . The  $ijk^{th}$  raw residual is  $n_{ijk} - n\hat{\pi}_{ijk}$ . A standardized version based on Poisson sampling is given by

$$e_{ijk} = \frac{n_{ijk} - n\hat{\pi}_{ijk}}{\sqrt{n\hat{\pi}_{ijk}}}.$$

## Example, personality-type data

The standardized Pearson residual is  $r_{ijk} = e_{ijk} / \sqrt{1 - \hat{h}_{ijk}}$ . One can find cells for which  $|r_{ijk}| > 3$  and flag them as being ill-fit, or simply compare the raw counts  $n_{ijk}$  to the fitted values  $n\hat{\pi}_{ijk}$ .

```
proc genmod order=data; class type chol bp;  
  model count = type|chol type|bp / dist=poi link=log r;
```

### Observation Statistics

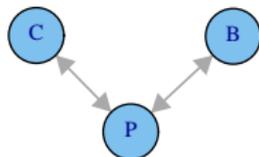
Observation	Resraw	Reschi	Resdev	StResdev	StReschi	Reslik
1	1.5063291	0.0563535	0.0563337	0.3724586	0.3725894	0.3725864
2	-1.506329	-0.167882	-0.16841	-0.37376	-0.372589	-0.372827
3	-1.506329	-0.104318	-0.104444	-0.373039	-0.372589	-0.372625
4	1.5063291	0.3107738	0.3075386	0.3687106	0.3725894	0.3698952
5	5.0786106	0.1780138	0.1778292	1.4298604	1.431345	1.4313221
6	-5.078611	-0.598194	-0.605433	-1.448667	-1.431345	-1.434386
7	-5.078611	-0.3674	-0.369046	-1.437757	-1.431345	-1.431768
8	5.0786106	1.2346018	1.179489	1.3674495	1.431345	1.3840886

The StReschi have the  $r_{ijk}$ . All are within  $|r_{ijk}| < 3$ .

## 10.1 Association graphs and collapsing tables

An association graph plots each factor as a vertex and connects factors according to interaction terms in the log-linear model.

Recall the the example that looked at personality type  $P$ , blood pressure  $B$ , and cholesterol  $C$ . We found the model  $[PC][PB]$  fit. This has association graph:

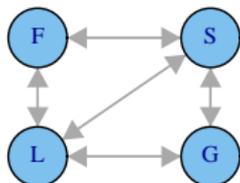


The two variables  $C$  and  $B$  are separated by  $P$ . All paths from  $C$  to  $B$  go through  $P$ . This implies that  $C \perp B | P$ .

# Collapsibility

From page 380: “Suppose that a model for a multiway table partitions variables into three mutually exclusive subsets  $A$ ,  $B$ , and  $C$  such that  $B$  separates  $A$  and  $C$ . After collapsing the table over the variables in  $C$ , parameters relating to variables in  $A$  and parameters relating  $A$  to  $B$  are unchanged.” Also:  $A \perp C|B$ .

Alligator food example: the model  $[GLS][SF][LF]$  fit the data. Then  $A = \{G\}$ ,  $C = \{F\}$  and  $B = \{L, S\}$  from the association graph:



We can collapse the table over gender and examine associations among  $F$ ,  $L$ ,  $S$  without worrying about Simpson's paradox (recall we dropped gender from the model with food as the outcome). Also:  $F \perp G|L, S$ .

## 10.2.2 Student survey of drug use

Example: Table 10.1 (p. 381). Five factors:  $M$ ,  $C$ ,  $A$ ,  $G$ ,  $R$ .

Model with all 10 3-factor interactions fits well with  $G^2 = 5.3$  on 6  $df$   $p$ -value is 0.5. Reduced model with all 10 2-factor interactions also fits well with  $G^2 = 15.3$  on 16  $df$  and  $p$ -value is 0.5 (again).

```
data drug;
input g r a c m count @@;
datalines;
0 1 1 1 1 405 0 1 1 1 0 268
0 1 1 0 1 13 0 1 1 0 0 218
0 1 0 1 1 1 0 1 0 1 0 17
0 1 0 0 1 1 0 1 0 0 0 117
1 1 1 1 1 453 1 1 1 1 0 228
1 1 1 0 1 28 1 1 1 0 0 201
1 1 0 1 1 1 1 1 0 1 0 17
1 1 0 0 1 1 1 1 0 0 0 133
0 0 1 1 1 23 0 0 1 1 0 23
0 0 1 0 1 2 0 0 1 0 0 19
0 0 0 1 1 0 0 0 0 1 0 1
0 0 0 0 1 0 0 0 0 0 0 12
1 0 1 1 1 30 1 0 1 1 0 19
1 0 1 0 1 1 1 0 1 0 0 18
1 0 0 1 1 1 1 0 0 1 0 8
1 0 0 0 1 0 1 0 0 0 0 17
;
```

# Model fitting

```
proc genmod;
  class g r a c m;
  model count=g|r|a g|r|c g|r|m g|a|c g|a|m g|c|m r|a|c r|a|m r|c|m a|c|m / link=log dist=poi type3;
proc genmod;
  class g r a c m;
  model count=g|r g|a g|c g|m r|a r|c r|m a|c a|m c|m / link=log dist=poi type3;
```

## LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
g	1	5.98	0.0144
r	1	828.44	<.0001
g*r	1	0.84	0.3597
a	1	378.56	<.0001
g*a	1	3.38	0.0661
c	1	20.19	<.0001
g*c	1	0.98	0.3230
m	1	248.74	<.0001
g*m	1	9.82	0.0017
r*a	1	4.98	0.0256
r*c	1	0.44	0.5056
r*m	1	3.59	0.0582
a*c	1	185.86	<.0001
a*m	1	91.62	<.0001
c*m	1	498.13	<.0001

We can remove [RC]. Then [GR]. Then [GC]. (Not shown).

# Final model

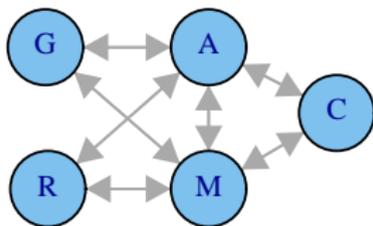
```
proc genmod;
  class g r a c m;
  model count=g|a g|m r|a r|m a|c a|m c|m / link=log dist=poi type3;
```

Source	DF	Chi-Square	Pr > ChiSq
g	1	6.20	0.0127
a	1	428.92	<.0001
g*a	1	5.51	0.0189
m	1	264.33	<.0001
g*m	1	8.90	0.0029
r	1	834.63	<.0001
r*a	1	4.78	0.0288
r*m	1	2.99	0.0836
c	1	25.49	<.0001
a*c	1	187.38	<.0001
a*m	1	92.05	<.0001
c*m	1	497.00	<.0001

The final model is  $[GA][GM][RA][RM][AC][AM][CM]$ . This model has  $G^2 = 17.54$  on 19 *df* for a *p*-value of 0.55.

# Interpretation of final model

The association graph looks like:



We see that  $C \perp G \perp R | M, A$ . For example, cigarette use is independent of gender given marijuana and alcohol use. Question: which individual factors can we collapse over and retain valid inferences for the remaining four parameters? Hint: there's three of them.

What if we accept that  $r * m$  is not needed above ( $p = 0.083$ )? Then race is connected to  $G$ ,  $M$ , and  $C$  only through alcohol. We would have  $R \perp (G, M, C) | A$ , i.e.  $R \perp G | A$ ,  $R \perp M | A$ , and  $R \perp C | A$ .

## 9.2.3 Interpretation for $[XY][XZ][YZ]$

In  $I \times J \times K$  table.

For  $1 \leq i \leq I - 1$  and  $1 \leq j \leq J - 1$  define

$$\theta_{ij(k)} = \frac{\pi_{i,j,k} \pi_{i+1,j+1,k}}{\pi_{i,j+1,k} \pi_{i+1,j,k}} = \frac{\left[ \frac{P(Y=j, X=i | Z=k)}{P(Y=j+1, X=i | Z=k)} \right]}{\left[ \frac{P(Y=j, X=i+1 | Z=k)}{P(Y=j+1, X=i+1 | Z=k)} \right]}.$$

There are  $(I - 1)(J - 1)$  local odds ratios at each level of  $Z = k$ . This completely determines the dependence structure among  $X, Y | Z = k$ .

For model  $[XY][XZ][YZ]$  we have

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

This implies

$$\log \theta_{ij(k)} = \lambda_{i,j}^{XY} + \lambda_{i+1,j+1}^{XY} - \lambda_{i,j+1}^{XY} - \lambda_{i+1,j}^{XY}.$$

So  $\theta_{ij(1)} = \theta_{ij(2)} = \dots = \theta_{ij(K)}$  for all  $i$  and  $j$ , the model of homogeneous association.

# Interpretation for $[XY][XZ][YZ]$

Similarly,  $[XY][XZ][YZ]$  implies  $\theta_{(1)jk} = \theta_{(2)jk} = \dots = \theta_{(I)jk}$  for all  $j$  and  $k$ , and  $\theta_{i(1)k} = \theta_{i(2)k} = \dots = \theta_{i(J)k}$  for all  $i$  and  $k$ . This is the difference between  $[XY][XZ][YZ]$  and the saturated model  $[XYZ]$  in which there is no homogeneous association.

---

## Section 9.5.1: $[XY][XZ][YZ]$ and logistic regression

Now let's say  $Y$  is the outcome and is dichotomous. Then

$$\begin{aligned} \log \frac{P(Y = 1|X = i, Z = k)}{P(Y = 2|X = i, Z = k)} &= \frac{P(Y = 1, X = i, Z = k)}{P(Y = 2, X = i, Z = k)} \\ &= \log n\pi_{i1k} - \log n\pi_{i2k} \\ &= \left[ \lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ} \right] \\ &\quad - \left[ \lambda + \lambda_i^X + \lambda_2^Y + \lambda_k^Z + \lambda_{i2}^{XY} + \lambda_{ik}^{XZ} + \lambda_{2k}^{YZ} \right] \\ &= \left[ \lambda_1^Y - \lambda_2^Y \right] + \left[ \lambda_{i1}^{XY} - \lambda_{i2}^{XY} \right] + \left[ \lambda_{1k}^{YZ} - \lambda_{2k}^{YZ} \right] \\ &\equiv \beta_0 + \beta_i^X + \beta_k^Z, \end{aligned}$$

which corresponds to an additive logistic regression model.

# Which: log-linear model or logistic regression?

If all's we care about is how  $(X, Y)$  relates to outcome  $Y$ , then logistic regression model is *okay*.

If we are concerned with dependence structure among  $(X, Y, Z)$ , then log-linear modeling is appropriate.

Table 9.11 gives the equivalent logistic regression model to several log-linear models:

log-linear model	logit model with outcome $Y$
$[Y][XY]$	logit $P(Y = 1) = \alpha$
$[XY][XZ]$	logit $P(Y = 1) = \beta_i^X$
$[YZ][XZ]$	logit $P(Y = 1) = \beta_k^Z$
$[XY][XZ][YZ]$	logit $P(Y = 1) = \beta_i^X + \beta_k^Z$
$[XYZ]$	logit $P(Y = 1) = \beta_i^X + \beta_k^Z + \beta_{ik}^{XZ}$

Question: where are  $[X][Y][Z]$ ,  $[X][YZ]$ ,  $[Z][XY]$ , and  $[XY][YZ]$ ?

## More on collapsibility

Recall “personality type” data, which had three factors:  $P$ ,  $C$ , and  $B$ . We decided  $[PC][PB]$  fit the data.

Fitting  $[PC][PB]$  yields  $\lambda_{11}^{PC} = -0.2176$  and  $\lambda_{11}^{PB} = -0.2409$ .

Fitting  $[PC]$ , i.e. *collapsing over blood pressure*, yields  $\lambda_{11}^{PC} = -0.2176$  (same as above).

Fitting  $[PB]$ , i.e. *collapsing over cholesterol*, yields  $\lambda_{11}^{PB} = -0.2409$  (same as above).

In model  $[PC][PB]$  we have

$$\theta_{11(k)} = \frac{P(P = 1, C = 1|B = k)P(P = 2, C = 2|B = k)}{P(P = 1, C = 2|B = k)P(P = 2, C = 1|B = k)}.$$

## More on collapsibility

In terms of the log-linear model parameters,

$$\log \theta_{11(k)} = [\lambda_{11}^{PC} + \lambda_{1k}^{PB}] + [\lambda_{22}^{PC} + \lambda_{2k}^{PB}] - [\lambda_{12}^{PC} + \lambda_{1k}^{PB}] - [\lambda_{21}^{PC} + \lambda_{2k}^{PB}] = \lambda_{11}^{PC},$$

independent of  $k$ ! This is because  $\lambda_{12}^{PC} = \lambda_{21}^{PC} = \lambda_{22}^{PC} = 0$  for identifiability.

So  $\hat{\theta}_{11(k)} = e^{-0.2176} = 0.80$ . The odds of having normal cholesterol is reduced 20% for personality type A (within each level of blood pressure).

Collapsing over blood pressure yielding model  $[PC]$  gives  $\theta_{11} = \lambda_{11}^{PC}$  from the reduced model, which has *exactly the same outcome*  $\hat{\theta}_{11} = 0.80$ .

As required by the collapsibility theorem, the marginal and conditional interpretations are the same. No information is lost by collapsing the table.

# Seat belt example revisited

The final model was  $[GLB][LBI][GI]$ . Ugh!

Can we say *anything* succinctly here?

Let's see how the gender/injury odds ratio changes with levels of location and belt use. Define

$$\theta_{11(kl)} = \frac{P(G = 1, I = 1|L = k, B = l)P(G = 2, I = 2|L = k, B = l)}{P(G = 1, I = 2|L = k, B = l)P(G = 2, I = 1|L = k, B = l)}.$$

In terms of log-linear model parameters,

$$\begin{aligned}\log \theta_{11(kl)} &= \left[ \lambda_{11l}^{GLB} + \lambda_{1kl}^{ILB} + \lambda_{11}^{GI} \right] + \left[ \lambda_{21l}^{GLB} + \lambda_{2kl}^{ILB} + \lambda_{22}^{GI} \right] \\ &\quad - \left[ \lambda_{11l}^{GLB} + \lambda_{2kl}^{ILB} + \lambda_{12}^{GI} \right] - \left[ \lambda_{21l}^{GLB} + \lambda_{1kl}^{ILB} + \lambda_{21}^{GI} \right] \\ &= \lambda_{11}^{GI},\end{aligned}$$

independent of  $L = k$  and  $B = l$ , the model of *homogeneous association*.

What is the association graph for  $[GLB][LBI][GI]$ ?

From the seatbelt output,  $\hat{\theta}_{11(kl)} = e^{-0.5459} = 0.58$ . The odds of not being injured for females is 0.58 times the odds for males within each  $(B, L)$  strata.

Fitting the table *collapsed over B and L*, i.e. fitting  $[GI]$ , we obtain the *marginal* odds ratio  $\hat{\theta}_{11} = e^{-0.4128} = 0.66$ .

The marginal interpretation is not the same (but not *that* different!) as the conditional interpretation. The conditions of the collapsibility theorem are not satisfied here, and so the interpretation changes upon collapsing the table.