

# Log-linear Models for Contingency Tables

## Edpsy/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology





# Outline

## Overview

Log-linear Models for 2-way  
Tables

---

Parameter Constraints

---

Log-linear Models for 3-way  
Tables

---

Inference for Log-linear Models

---

statistical versus practical  
significance

---

Log-linear Models for  
4+ -Way Tables

---

Logit-Log-linear Model  
Connection

---

Strategies in (Log-linear) Model  
Selection

---

In this set of notes:

- Loglinear models for 2-way tables.
- Loglinear models for 3-way tables.
- Statistical inference & model checking.
- Statistical versus Practical Significance.
- Higher-way tables.
- The logit-log-linear model connection. (We'll discuss further connections when we cover multcategory logit models).
- Model building (graphical models).
- Modelling ordinal associations, including linear  $\times$  linear associations.
- Modelling approach to testing conditional independence.
- Sparse data, including
  - ◆ Structural zeros
  - ◆ Sampling zeros
  - ◆ Effect on  $G^2$  and  $X^2$ .



# Log-linear models (or Poisson regression)

$$\log(\mu) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

where

$\mu$  = response variable = count (or rate)

- The most common use of log-linear models is for modeling counts in contingency tables; that is, the explanatory variables are all categorical.
- Log-linear models are used to model the association (or interaction structure) between/among categorical variables.
- The categorical variables (which in GLM terminology are explanatory variables) are the “responses” in the sense that we’re interested in describing the relationship between the variables.
- This use of “response” differs from our use in GLM. For log-linear models, the response variable are the cell frequencies (counts) in the contingency table.

## Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Log-linear Models for 2-way Tables

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Review of notations for 2-way Contingency tables:

- $I$  = the number of rows.
- $J$  = the number of columns.
- $I \times J$  contingency table.
- $N = IJ$  = the number of cells in the table.
- $n$  = the number of subjects (respondents, objects, etc.) cross-classified by 2 discrete (categorical) variables.



# Example: 1989 GSS data

(from Demaris).

Cross-classification of respondents according to

- **Choice** for president in the 1988 presidential election (Dukakis or Bush).
- **Political View** with levels liberal, moderate, conservative.

Political View	Vote Choice		Total
	Dukakis	Bush	
Liberal	197	65	263
Moderate	148	186	334
Conservative	68	242	310
	413	493	906

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# A More Recent Example

Cross-classification of respondents to the General Social Survey from 1996.

- **Choice** for president in the 1992: presidential election:  
“If you voted in 1992, did you vote for Clinton, Bush or Perot?”
- **Political View** with levels liberal, moderate, conservative.

“We hear a lot of talk these days about liberals and conservatives. I’m going to give you a scale . . . Where do you place yourself on this scale?”

The scale had 7 levels: extremely liberal, liberal, slightly liberal, moderate, etc., but I collapsed them in to 3.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# The Data

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Political View	Vote Choice			Total
	Bush	Clinton	Perot	
Liberal	70 (.15)	342 (.73)	56 (.12)	468 1.00
Moderate	195 (.31)	332 (.53)	101 (.16)	628 1.00
Conservative	382 (.55)	199 (.29)	117 (.17)	698 1.00
Total	647	873	274	1794



# Statistical Independence

Between row and column variable means that

- The joint probabilities  $\{\pi_{ij}\}$  of observations falling into a cell equal the product of the marginal probabilities,

$$\pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{for all } i = 1, \dots, I \quad \text{and} \quad j = 1, \dots, J$$

- The frequencies (cell counts) equal

$$\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j} \quad \text{for all } i, j$$

- The probabilities  $\pi_{ij}$  are the parameters of Binomial or Multinomial distribution.
- In Log-linear models, the response variable equals the counts and expected cell counts  $\{\mu_{ij}\}$ , rather than cell probabilities  $\{\pi_{ij}\}$ ; therefore, the random component is Poisson.
- Taking logarithms gives us a log-linear model of statistical independence

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i+}) + \log(\pi_{+j})$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Log-Linear Model of Statistical Independence

The “log-linear model of independence” for 2–way contingency tables is

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

- This is an ANOVA-type representation.
- $\lambda$  represents an “overall” effect or a constant.

It term ensures that  $\sum_i \sum_j \mu_{ij} = n$ .

- $\lambda_i^X$  represents the “main” or marginal effect of the row variable  $X$ . It represents the effect of classification in row  $i$ .

The  $\lambda_i^X$ ’s ensure that  $\sum_j \mu_{ij} = \mu_{i+} = n_{i+}$ .

- $\lambda_j^Y$  represents the “main” or marginal effect of the column variable  $Y$ . It represents the effect of classification in column  $j$ .

The  $\lambda_j^Y$ ’s ensure that  $\sum_i \mu_{ij} = \mu_{+j} = n_{+j}$ .

- The re-parameterization allow modeling association structure.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Log-linear model of Statistical Independence

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

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

& hypothesis test of statistical independence in 2-way tables.

- The estimated expected cell counts for the chi-squared test of independence equal (from beginning of course)

$$\hat{\mu}_{ij} = \frac{n_{i+}n_{+j}}{n}$$

- which also equal the estimated fitted values for the independence log-linear model.
- The significance of this: The  $X^2$  and  $G^2$  tests of independence are goodness-of-fit tests of the independence log-linear model.
- The null hypothesis of independence is equivalent to

The model  $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$  holds.

- and the alternative hypothesis of dependence is equivalent to

The model  $\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$  does not hold.



# Example

```
PROC GENMOD ORDER=DATA;
CLASS pview choice;
MODEL count = pview choice / LINK=log DIST=Poisson OBSTATS;
```

The observed and estimated fitted values

Political View	Vote Choice			Total
	Bush	Clinton	Perot	
Liberal	70 (168.78)	342 (227.74)	56 (71.478)	468
Moderate	195 (226.49)	332 (305.60)	101 (95.915)	628
Conservative	382 (251.73)	199 (339.66)	117 (106.61)	698
Total	647	873	274	1794

The fitted values satisfy the definition of independence perfectly. e.g.,  
for the (Liberal, Bush) cell  $\frac{n_{1+}n_{+1}}{n} = \frac{(468)(647)}{1794} = 168.78$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Independence also implies

that the odds ratios for every  $2 \times 2$  sub-table must equal 1.

For our example, fitted odds ratio for each of the 3 possible sub-tables for Bush and Clinton:

$$(168.78)(305.60)/(227.74)(226.49) = 1.00$$

$$(226.49)(339.66)/(305.60)(251.73) = 1.00$$

$$(168.78)(339.66)/(227.74)(251.73) = 1.00$$

The same is true for all possible  $(2 \times 2)$  sub-tables.

Fit statistics for the independence log-linear model:

Statistic	<i>df</i>	Value	<i>p</i> -value
$X^2$	4	252.10	< .0001
$G^2$	4	262.26	< .0001
Log Likelihood		7896.55	

These are the same the  $X^2$  and  $G^2$  we get when testing independence (i.e., what you get if you use *PROC FREQ*).

Any guesses as to what model might fit these data?

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Log-linear Model as a GLM

For  $I \times J$  Tables

**Random component.** The  $N = IJ$  counts in the cells of the contingency tables are assumed to be  $N$  independent observations of a Poisson random variable. Thus, we focus on expected values of counts:

$$E(\text{counts}) = \mu_{ij}$$

**Link** is log (canonical link for the Poisson distribution).

**Systematic component** is a linear predictor with discrete variables.

Loglinear model (of independence) is

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

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Log-Linear Model Parameters

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

- The row and column variables ( $X$  and  $Y$ , respectively) are both “response” variables (classification variables) in the sense that this model represents the relationship between the 2 variables.
- Log-linear models do not distinguish between “response” and explanatory (predictor) variables.
- When one variable is a response variable, then this influences (guides) the interpretation of parameters (as well as choice of model).

Case of an  $I \times 2$  tables where the column classification ( $Y$ ) is the “response” or outcome variable and the row classification ( $X$ ) is an explanatory variable.  
e.g., 1992 Presidential election

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# E.g., 1992 Presidential Election

Just consider Clinton and Bush and re-fit the independence model.

Let  $\pi = \text{Prob}(Y = 1) = \text{Prob}(\text{Clinton})$ , so

$$\begin{aligned}\text{logit}(\pi) &= \log(\mu_{i1}/\mu_{i2}) \\ &= \log(\mu_{i1}) - \log(\mu_{i2}) \\ &= (\lambda + \lambda_i^X + \lambda_1^Y) - (\lambda + \lambda_i^X + \lambda_2^Y) \\ &= \lambda_1^Y - \lambda_2^Y\end{aligned}$$

which does not depend on the row variable.

- The log-linear model of independence corresponds to the logit model with only an intercept term; that is,

$$\text{logit}(\pi) = \alpha$$

where  $\alpha = (\lambda_1^Y - \lambda_2^Y)$  is the same for all rows (levels of political view).

- Odds =  $\exp(\alpha) = \exp(\lambda_1^Y - \lambda_2^Y)$  is the same for all rows.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Interpretation of Parameters

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- When there are only two levels of the response variable, logit models are often preferable (they're simpler in the sense that they contain fewer terms in the model). This is especially true when we have 2 or more explanatory variables.
- Log-linear models are primarily used when modeling the relationship among 2 or more categorical responses.

Odds ratios are functions of model parameters:

$$\begin{aligned}\log(\text{odds ratio}) = \log(\theta_{(12,12)}) &= \log\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) \\ &= (\lambda + \lambda_1^X + \lambda_1^Y) + (\lambda + \lambda_2^X + \lambda_2^Y) \\ &\quad - (\lambda + \lambda_1^X + \lambda_2^Y) - (\lambda + \lambda_2^X + \lambda_1^Y) \\ &= 0\end{aligned}$$

So the odds ratio,  $\theta = \exp(0) = e^0 = 1$ .



# Parameter Constraints

Identification constraints on parameters to be able to estimate them from data.

There is not a unique set of parameters.

There are  $I$  terms in the set  $\{\lambda_i^X\}$ , but 1 is redundant.

There are  $J$  terms in the set  $\{\lambda_j^Y\}$ , but 1 is redundant.

Possible (typical) constraints:

1. Fix 1 value in a set equal to a constant, usually 0.  
*SAS/GENMOD* sets the last one equal to 0, e.g.,  
 $\lambda_I^X = 0 \dots$  dummy coding (i.e.,  $X = 0, 1$ )
2. Fix the sum of the terms equal to a constant, usually 0.  
*SAS/CATMOD* uses zero sum or “ANOVA-type” constraints.  
e.g.,  $\sum_{i=1}^I \lambda_i^X = 0 \dots$  “effect” coding (i.e.,  $X = 1, 0, -1$ ).

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# What's Unique about the Parameters?

Overview

Log-linear Models for 2-way  
Tables

**Parameter Constraints**

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- The differences between them are unique:

$$(\hat{\lambda}_1^Y - \hat{\lambda}_2^Y) = \text{unique value}$$

$$(\hat{\lambda}_1^X - \hat{\lambda}_2^X) = \text{unique value}$$

- Since differences are unique,

$$\log(\text{odds}) = \log(\theta) = \text{unique value}$$

and

$$\text{odds} = \text{unique value}$$

- The goodness-of-fit statistics are unique.
- The fitted values are unique, which takes more space to show...



# Fitted Values are Unique

e.g., for  $2 \times 2$

$$\log(\hat{\mu}_{ij}) = \alpha + \lambda^X X_i + \lambda^Y Y_j$$

For Dummy Coding (i.e.,  $X_1 = 0, X_2 = 1$  and  $Y_1 = 0, Y_2 = 1$ ),

$$\log(\hat{\mu}_{ij}) = \begin{cases} \lambda & \text{for } (1, 1) \\ \lambda + \lambda^X & \text{for } (2, 1) \\ \lambda + \lambda^Y & \text{for } (1, 2) \\ \lambda + \lambda^X + \lambda^Y & \text{for } (2, 2) \end{cases}$$

For Effect Coding (i.e.,  $X_1 = -1, X_2 = 1$  and  $Y_1 = -1, Y_2 = 1$ ),

$$\log(\hat{\mu}_{ij}) = \begin{cases} \lambda^* - \lambda^{*X} - \lambda^{*Y} & \text{for } (1, 1) \\ \lambda^* + \lambda^{*X} - \lambda^{*Y} & \text{for } (2, 1) \\ \lambda^* - \lambda^{*X} + \lambda^{*Y} & \text{for } (1, 2) \\ \lambda^* + \lambda^{*X} + \lambda^{*Y} & \text{for } (2, 2) \end{cases}$$

What's the correspondence?

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Saturated Log-linear Model for 2-way Tables

If rows and columns are dependent, then

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

- $\lambda$ ,  $\lambda_i^X$ , and  $\lambda_j^Y$ , are the overall and marginal effect terms (as defined before).
- $\lambda_{ij}^{XY}$ 's
  - ◆ Represent the association between  $X$  and  $Y$ .
  - ◆ Reflect the departure or deviations from independence.
  - ◆ Ensure that  $\mu_{ij} = n_{ij}$
- Fits the data perfectly; the fitted values are exactly equal to the observed values.
- Has as many unique parameters as there are cells in the table (i.e.,  $N = IJ$ ), so  $df = 0$ .
- Called the “Saturated Model”.
- Is the most complex model possible for a 2-way table.
- Has independence as a special case (i.e., the model with  $\lambda_{ij}^{XY} = 0$  for all  $i$  and  $j$ ).

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Parameters and Odds Ratios

There is a functional relationship between the model parameters and odds ratios, which is how we are defining and measuring interactions.

$$\begin{aligned}\log(\theta_{ii',jj'}) &= \log(\mu_{ij}\mu_{i'j'} / \mu_{i'j}\mu_{ij'}) \\ &= \log(\mu_{ij}) + \log(\mu_{i'j'}) - \log(\mu_{i'j}) - \log(\mu_{ij'}) \\ &= (\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}) + (\lambda + \lambda_{i'}^X + \lambda_{j'}^Y + \lambda_{i'j'}^{XY}) \\ &\quad - (\lambda + \lambda_{i'}^X + \lambda_j^Y + \lambda_{i'j}^{XY}) - (\lambda + \lambda_i^X + \lambda_{j'}^Y + \lambda_{ij'}^{XY}) \\ &= \lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY}\end{aligned}$$

The odds ratio  $\theta$  measures the strength of the association and depends only on the interaction terms  $\{\lambda_{ij}^{XY}\}$ .

How many numbers do we need to completely characterize the association in an  $I \times J$  table?

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Parameters Needed to Describe Association

$(I - 1)(J - 1)$ , the number of unique  $\lambda_{ij}^{XY}$ 's.

Identification constraints:

- Fix 1 value equal to a constant, e.g.,

$$\lambda_{i1}^{XY} = \lambda_{1j}^{XY} = 0$$

- Fix the sum equal to a constant, i.e.,

$$\sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = 0$$

Count of unique parameters:

Terms	Number of Terms	Number of Constraints	Number Unique
$\lambda$	1	0	1
$\{\lambda_i^X\}$	$I$	1	$I - 1$
$\{\lambda_j^Y\}$	$J$	1	$J - 1$
$\{\lambda_{ij}^{XY}\}$	$IJ$	$I + J - 1$	$(I - 1)(J - 1)$
Total			$IJ = N$ cells of table

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Parameters Needed to Describe Association

Overview

Log-linear Models for 2-way  
Tables

**Parameter Constraints**

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

We generally hope to find models that are simpler than the data itself (simpler than the saturated model). Simpler models “smooth” the sample data and provide more parsimonious descriptions.

When we have 3 or more variables, we can include 2-way interactions and the model will not be saturated.



# Hierarchical Models

We'll restrict attention to hierarchical models.

- Hierarchical models include all lower-order terms that comprise the the higher-order terms in the model.

- Is this a hierarchical model?

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_{ij}^{XY}$$

- Restrict attention to hierarchical models because
  - ◆ We want interaction terms to represent just the association (dependency).
  - ◆ Without lower order terms, the statistical significance and (substantive) interpretation of interaction terms would depend on how variables were coded.
  - ◆ With hierarchical models, coding doesn't matter.
- If there is an interaction in the data, we do not look at the lower-order terms, but interpret the higher-order (interaction) terms. It can be misleading to look at the lower-order terms, because the values will depend on the coding scheme.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Log-linear Models for 3-way Tables

Example: We can add a third variable to our GSS presidential election data — gender.

Gender	Political View	Choice for President			
		Bush	Clinton	Perot	Total
Males	Liberal	26	121	24	171
	Moderate	82	128	52	262
	Conservative	202	75	74	351
Females	Liberal	44	221	32	297
	Moderate	113	204	49	366
	Conservative	180	124	43	347

The most saturated log-linear model for this table is

$$\log(\mu_{ijk}) = \lambda + \lambda_i^P + \lambda_j^C + \lambda_k^G + \lambda_{ij}^{PC} + \lambda_{ik}^{PG} + \lambda_{jk}^{CG} + \lambda_{ijk}^{PCG}$$

where  $G$  = gender,  $P$  = political view,  $C$  = choice.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Saturated Model for 3-way Table

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

More generally, the most complex log-linear model (the saturated model) for a 3-way table

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

Simpler models set higher order interaction terms equal to 0.

Overview of hierarchy of models for 3-way contingency tables  
(from complex to simple)



# Overview of Models for 3-way Table

3-way Association

$$(XYZ)$$

Homogeneous association

$$(XY, XZ, YZ)$$

Conditional Independence

$$(XY, XZ)$$

$$(XY, YZ)$$

$$(XZ, YZ)$$

Joint Independence

$$(XY, Z)$$

$$(XZ, Y)$$

$$(X, YZ)$$

Complete Independence

$$(X, Y, Z)$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Blue Collar worker data

(Andersen, 1985)

		Bad Management			Good Management		
		Worker's satisfaction		Total	Worker's satisfaction		Total
Supervisor's satisfaction	Low	High	Low		High	Low	
	Low	103	87	190	59	109	168
High	32	42	74	78	205	283	
	135	129	264	137	314	451	

$$\hat{\theta}_{bad} = 1.55 \quad \text{and} \quad 95\% \text{ CI for } \theta_{bad} (.90, 1.67)$$

$$\hat{\theta}_{good} = 1.42 \quad \text{and} \quad 95\% \text{ CI for } \theta_{good} (.94, 2.14)$$

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Blue Collar worker data Analyses

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- CMH for testing whether worker and supervisor satisfaction is conditionally independent given management quality  
= 5.42,  $p$ -value = .02

- The combined  $G^2$ 's from separate partial tables  
= 2.57 + 2.82 = 5.39,  $df = 1 + 1 = 2$ ,  $p$ -value = .02.

Statistic	$df$	Bad Management		Good Management	
		Value	$p$ -value	Value	$p$ -value
$X^2$	1	2.56	.11	2.85	.09
$G^2$	1	2.57	.11	2.82	.09

- Mantel-Haenszel estimator of common odds ratio (for Worker-Supervisor or "W-S" odds ratio) = 1.47.
- Breslow-Day statistic = .065,  $df = 1$ ,  $p$ -value = .80.



# Complete Independence

There are no interactions; everything is independent of everything else.

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

Depending on author, this model is denoted by

- $(X, Y, Z)$  (e.g., Agresti) or  $[X, Y, Z]$
- $(X)(Y)(Z)$  or  $[X][Y][Z]$  (e.g., Fienberg).

Degrees of freedom are computed in the usual way:

$$\begin{aligned} df &= \# \text{ cells} - \# \text{ unique parameters} \\ &= \# \text{ cells} - (\# \text{ parameters} - \# \text{ constraints}) \\ &= IJK - 1 - (I - 1) - (J - 1) - (K - 1) \end{aligned}$$

In terms of associations, all partial odds ratios equal 1,

$$\theta_{XY(k)} = \theta_{ii',jj',(k)} = \mu_{ijk}\mu_{i'j'k} / \mu_{i'jk}\mu_{ij'k} = 1$$

$$\theta_{YZ(i)} = \theta_{(i),jj',kk'} = \mu_{ijk}\mu_{ij'k'} / \mu_{ij'k}\mu_{ijk'} = 1$$

$$\theta_{XZ(j)} = \theta_{ii',(j),kk'} = \mu_{ijk}\mu_{i'jk'} / \mu_{i'jk}\mu_{ijk'} = 1$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Joint Independence

Two variables are “jointly” independent of the third variable. For example,  $X$  and  $Y$  and jointly independent of  $Z$ ,

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

This model may be denoted as

- $(XY, Z)$  or  $[XY, Z]$ .
- $(XY)(Z)$  or  $[XY][Z]$ .

Degrees of Freedom:

$$\begin{aligned} df &= IJK - 1 - (I - 1) - (J - 1) - (K - 1) - (I - 1)(J - 1) \\ &= (IJ - 1)(K - 1) \end{aligned}$$

The partial or conditional odds ratios for  $XZ$  given  $Y$  and the odds ratios for  $YZ$  given  $X$  equal 1.

$$\theta_{XZ(j)} = \theta_{ii',(j),kk'} = \mu_{ijk}\mu_{i'jk'} / \mu_{i'jk}\mu_{ijk'} = 1$$

$$\theta_{YZ(i)} = \theta_{(i),jj',kk'} = \mu_{ijk}\mu_{ij'k'} / \mu_{ij'k}\mu_{ijk'} = 1$$

And what does  $\theta_{XY(k)}$  equal?

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Joint Independence (continued)

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

$$\begin{aligned} \log(\theta_{XY(k)}) &= \log(\theta_{ii',jj'(k)}) \\ &= \log(\mu_{ijk}\mu_{i'j'k} / \mu_{i'jk}\mu_{ij'k}) \\ &= (\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}) \\ &\quad + (\lambda + \lambda_{i'}^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{i'j'}^{XY}) \\ &\quad - (\lambda + \lambda_{i'}^X + \lambda_j^Y + \lambda_k^Z + \lambda_{i'j}^{XY}) \\ &\quad - (\lambda + \lambda_i^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{ij'}^{XY}) \\ &= \lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY} \end{aligned}$$

$$\theta_{XY(k)} = \theta_{ii',jj'(k)} = \exp(\lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY}) = \frac{e^{\lambda_{ij}^{XY}} e^{\lambda_{i'j'}^{XY}}}{e^{\lambda_{i'j}^{XY}} e^{\lambda_{ij'}^{XY}}}$$

- The interaction terms represent the association between  $X$  and  $Y$ .
- Since  $\theta_{XY(k)}$  does not depend on  $k$  (level of variable)  $Z$ , we know that  $\theta_{XY(1)} = \theta_{XY(2)} = \dots = \theta_{XY(K)}$

- Overview
- Log-linear Models for 2-way Tables
- Parameter Constraints
- Log-linear Models for 3-way Tables**
- Inference for Log-linear Models
- statistical versus practical significance
- Log-linear Models for 4+-Way Tables
- Logit-Log-linear Model Connection
- Strategies in (Log-linear) Model Selection



# Conditional Independence

Two variables are conditionally independent given the third variable.  
e.g., the model in which  $Y$  and  $Z$  are conditionally independent given

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

This model may be denoted by

- $(XY, XZ)$  or  $[XY, XZ]$ .
- $(XY)(XZ)$  or  $[XY][XZ]$ .

$$\begin{aligned} df &= IJK - 1 - (I - 1) - (J - 1) - (K - 1) \\ &\quad - (I - 1)(J - 1) - (I - 1)(K - 1) \\ &= I(J - 1)(K - 1) \end{aligned}$$

The partial odds ratios of  $YZ$  given  $X$  equals 1:

$$\begin{aligned} \log(\theta_{YZ(i)}) &= \log(\theta_{(i),jj',kk'}) \\ &= \log(\mu_{ijk}\mu_{ij'k'} / \mu_{ij'k}\mu_{ijk'}) \\ &= \log(\mu_{ijk}) + \log(\mu_{ij'k'}) - \log(\mu_{ij'k}) - \log(\mu_{ijk'}) = 0 \end{aligned}$$

$$\theta_{YZ(i)} = \theta_{(i),jj',kk'} = \exp(0) = e^0 = 1$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Conditional Independence: $\theta_{XY(k)}$ & $\theta_{XZ(j)}$

$$\theta_{XY(k)} = \theta_{ii',jj'(k)} = \exp(\lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY})$$

$$\theta_{XZ(j)} = \theta_{ii',(j),kk'} = \exp(\lambda_{ik}^{XZ} + \lambda_{i'k'}^{XZ} - \lambda_{i'k}^{XZ} - \lambda_{ik'}^{XZ})$$

- The partial odds ratios are completely characterized by the corresponding 2-way interaction terms from the model (and no other parameters).
- Neither of these depend on the level of the third variable.
- Since the partial odds ratios are equal across levels of the third variable,

$$\theta_{XY(1)} = \theta_{XY(2)} = \dots = \theta_{XY(K)}$$

and

$$\theta_{XZ(1)} = \theta_{XZ(2)} = \dots = \theta_{XZ(J)}$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Homogeneous Association

or the “no 3–factor interaction model”.

This is a model of association; it is not an “independence” model, but it is also not the most complex model possible.

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

This model may be denoted by

- $(XY, XZ, YZ)$  or  $[XY, XZ, YZ]$ .
- $(XY)(XZ)(YZ)$  or  $[XY][XZ][YZ]$ .

$$\begin{aligned} df &= IJK - 1 - (I - 1) - (J - 1) - (K - 1) - (I - 1)(J - 1) \\ &\quad - (I - 1)(K - 1) - (J - 1)(K - 1) \\ &= (I - 1)(J - 1)(K - 1) \end{aligned}$$

$df$  = the number of odds ratios to completely represent a 3–way association?

None of the partial odds ratios (necessarily) equal 1.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Homogeneous Association (continued)

The partial odds ratios are a direct function of the model parameters

The partial (conditional) odds ratios equal

$$\theta_{XY(k)} = \theta_{ii',jj',(k)} = \exp(\lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY})$$

$$\theta_{XZ(j)} = \theta_{ii',(j),kk'} = \exp(\lambda_{ik}^{XZ} + \lambda_{i'k'}^{XZ} - \lambda_{i'k}^{XZ} - \lambda_{ik'}^{XZ})$$

$$\theta_{YZ(i)} = \theta_{(i),jj'),kk'} = \exp(\lambda_{jk}^{YZ} + \lambda_{j'k'}^{YZ} - \lambda_{j'k}^{YZ} - \lambda_{jk'}^{YZ})$$

Each of the partial odds ratios for 2 variables given levels of the third variable

- depends only on the corresponding 2-way interaction terms.
- do not depend on levels of the third variable.
- are equal across levels of the third variable.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# 3-way Association (the saturated model)

This model has a three factor association

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

This model may be denoted by  $(XYZ)$  or  $[XYZ]$ .

$$df = 0$$

The partial odds ratios for two variables given levels of the third variable need not be equal. e.g.,

$$\begin{aligned} \log(\theta_{XY(k)}) &= \log(\theta_{ii',jj'(k)}) \\ &= \log(\mu_{ijk}\mu_{i'j'k} / \mu_{i'jk}\mu_{ij'k}) \\ &= \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ} \\ &\quad + \lambda + \lambda_{i'}^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{i'j'}^{XY} + \lambda_{i'k}^{XZ} + \lambda_{j'k}^{YZ} + \lambda_{i'j'k}^{XYZ} \\ &\quad - \lambda + \lambda_{i'}^X + \lambda_j^Y + \lambda_k^Z + \lambda_{i'j}^{XY} + \lambda_{i'k}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{i'jk}^{XYZ} \\ &\quad - \lambda + \lambda_i^X + \lambda_{j'}^Y + \lambda_k^Z + \lambda_{ij'}^{XY} + \lambda_{ik}^{XZ} + \lambda_{j'k}^{YZ} + \lambda_{ij'k}^{XYZ} \\ &= (\lambda_{ij}^{XY} + \lambda_{i'j'}^{XY} - \lambda_{i'j}^{XY} - \lambda_{ij'}^{XY}) \\ &\quad + (\lambda_{ijk}^{XYZ} + \lambda_{i'j'k}^{XYZ} - \lambda_{i'jk}^{XYZ} - \lambda_{ij'k}^{XYZ}) \end{aligned}$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# 3-way Association

A measure/definition of 3-way association is the ratio of partial odds ratios (ratios of ratios of ratios),

$$\Theta_{ii',jj',kk'} = \theta_{XY(k)} / \theta_{XY(k')}$$

which in terms of our model parameters equals

$$\begin{aligned} \Theta_{ii',jj',kk'} &= \frac{\theta_{ii',jj'(k)}}{\theta_{ii',jj'(k')}} \\ &= \exp(\lambda_{ijk}^{XYZ} + \lambda_{i'j'k}^{XYZ} + \lambda_{i'jk'}^{XYZ} + \lambda_{ij'k'}^{XYZ} \\ &\quad - \lambda_{i'jk}^{XYZ} - \lambda_{ij'k}^{XYZ} - \lambda_{ijk'}^{XYZ} - \lambda_{i'j'k'}^{XYZ}) \end{aligned}$$

That is, the 3-way association is represented by the 3-way interaction terms  $\{\lambda_{ijk}^{XYZ}\}$ .

There are analogous expressions for  $\theta_{XZ(j)}$  and  $\theta_{YZ(i)}$ .

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Summary of Hierarchy of Models

3-way Association

$$(XYZ)$$

Homogeneous association

$$(XY, XZ, YZ)$$

Conditional Independence

$$(XY, XZ)$$

$$(XY, YZ)$$

$$(XZ, YZ)$$

Joint Independence

$$(XY, Z)$$

$$(XZ, Y)$$

$$(X, YZ)$$

Complete Independence

$$(X, Y, Z)$$

Any model that lies below a given model may be a special case of the more complex model(s).

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+ -Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Example: Blue Collar Workers

Fitted values and observed values from some models

Manage	Super	Worker	$n_{ijk}$	$M, S, W$	$MS, W$	$MS, MW$	$MSW$	no
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	

Example:  $(M, S, W)$  and partial odds ratio for S and W

$$\hat{\theta}_{SW} = (50.15)(81.67)/(82.59)(49.59) = 1.00$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Fitted (partial) Odds Ratios

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Model	Fitted Partial Odds Ratio		
	W and S	M and W	M and S
$(M, S, W)$	1.00	1.00	1.00
$(MS, W)$	1.00	1.00	4.28
$(MS, MW)$	1.00	2.40	4.32
$(MS, WS, MW)$	1.47	2.11	4.04
$(MSW)$ -level 1	1.55	2.19	4.26
$(MSW)$ -level 2	1.42	2.00	3.90



# Inference for Log-linear Models

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

**Inference for Log-linear Models**

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

1. Chi-squared goodness of fit tests.
2. Residuals.
3. Tests about partial associations (e.g.,  $H_0 : \lambda_{ij}^{XY} = 0$  for all  $i, j$ ).
4. Confidence intervals for odds ratios.



# Chi-squared goodness-of-fit tests

where  $H_0$  is

- a model “holds”.
- a model gives a good (accurate) description/representation of the data.
- $\log(\mu_{ij}) = \text{some model}$   
(i.e., expected frequencies given by loglinear model).

For “large” samples chi-squared statistics to test this hypothesis, we compare the observed and estimated expected frequencies.

Likelihood ratio statistic:

$$G^2 = 2 \sum_i \sum_j \sum_k n_{ijk} \log \left( \frac{n_{ijk}}{\hat{\mu}_{ijk}} \right)$$

Pearson statistic:

$$X^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

**Inference for Log-linear Models**

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Chi-squared goodness-of-fit tests

If  $H_0$  is true and for larger samples, these statistics are approximately chi-squared distributed with degrees of freedom

$$df = (\# \text{ cells}) - (\# \text{ non-redundant parameters})$$

$$= (\# \text{ cells}) - (\# \text{ parameters}) + (\# \text{ id constraints})$$

Blue collar worker data:

Model	$df$	$G^2$	$p$ -value	$X^2$	$p$ -value
$(M, S, W)$	4	118.00	< .001	128.09	< .001
$(MS, W)$	3	35.60	< .001	35.72	< .001
$(MW, S)$	3	87.79	< .001	85.02	< .001
$(M, WS)$	3	102.11	< .001	99.09	< .001
$(MW, SW)$	2	71.90	< .001	70.88	< .001
$(MS, MW)$	2	5.39	.07	5.41	.07
$(MS, WS)$	2	19.71	< .001	19.88	< .001
$(MW, SW, MS)$	1	.065	.80	.069	.80

These are all global tests.

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Residuals

Local (miss)fit. A good model has small residuals.

We can use Pearson residuals

$$\begin{aligned} e_{ijk} &= \frac{(\text{observed} - \text{expected})}{\hat{\text{Var}}(\text{expected})} \\ &= \frac{(n_{ijk} - \hat{\mu}_{ijk})}{\hat{\mu}_{ijk}} \end{aligned}$$

or

$$\text{adjusted residual} = \frac{e_{ijk}}{(1 - h_{ijk})}$$

where  $h_{ijk}$  equals the leverage of cell  $(i, j, k)$ .

If the model holds, then adjusted residuals  $\approx N(0, 1)$

Adjusted residuals suggest a lack of fit of the model

- When there are few cells (small  $N$ ) and adjusted residuals  $> 2$ .
- When there are lots and lots of cells (larger  $N$ ) and adjusted residuals  $> 3$ .

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Residuals & Blue Collar Data

Manage	Super	Worker	$n_{ijk}$	$(MS, MW)$		$(MS, MS, WS)$	
				$\hat{\mu}_{ijk}$	adj res	$\hat{\mu}_{ijk}$	adj res
bad	low	low	103	97.16	1.60	102.26	.25
bad	low	high	87	92.84	-1.60	87.74	-.25
bad	high	low	32	37.84	-1.60	32.74	-.25
bad	high	high	42	36.16	1.60	41.26	.25
good	low	low	59	51.03	1.69	59.74	-.25
good	low	high	109	116.97	-1.69	108.26	.25
good	high	low	78	85.97	1.69	77.26	.25
good	high	high	205	197.28	-1.69	205.74	-.25

- $df$  for the model  $(MS, MW)$  equals 2 and therefore there are only 2 non-redundant residuals.
- $df$  for the model  $(MS, MW, WS)$  equals 1 and therefore there is only 1 non-redundant residual.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

**Inference for Log-linear Models**

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Hypothesis about partial association

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

The following are all equivalent statements of the null hypothesis considered here:

- There is no partial association between two variables given the level of the third variable.

e.g., There is no partial association between supervisor's job satisfaction and worker's satisfaction given management quality.

- The conditional or partial odds ratios equal 1.00.

e.g.,  $\theta_{SW(i)} = 1.00$ .

- The two-way interaction terms equal zero.

e.g.,  $\lambda_{jk}^{SW} = 0$ .



# Tests about partial association

To test partial association, we use the likelihood ratio statistic  $-2(L_O - L_1)$  to test the difference between a restricted and a more complex model.

e.g., The restricted model or  $M_O$  is  $(MS, MW)$  or

$$\log(\mu_{ijk}) = \lambda + \lambda_i^M + \lambda_j^S + \lambda_k^W + \lambda_{ij}^{MS} + \lambda_{ik}^{MW}$$

and the more complex model  $M_1$  is  $(MS, MW, MS)$

$$\log(\mu_{ijk}) = \lambda + \lambda_i^M + \lambda_j^S + \lambda_k^W + \lambda_{ij}^{MS} + \lambda_{ik}^{MW} + \lambda_{ij}^{SW}$$

The likelihood ratio statistic  $-2(L_O - L_1)$  equals the difference between the deviances of the 2 models, or equivalently the difference in  $G^2$  for testing model fit.

e.g.,

$$G^2 [(MS, MW)|(MS, MW, WS)] = G^2(MS, MW) - G^2(MS, MW, WS)$$

and  $df = df(MS, MW) - df(MS, MW, WS)$ .

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Example Testing partial association

Model	$df$	$G^2$	$p$	$H_O$	$\Delta df$	$\Delta G^2$	$p$
$(MW, SW, MS)$	1	.065	.80	—	—	—	—
$(MW, SW)$	2	71.90	< .001	$\lambda_{ij}^{MS} = 0$	1	71.835	< .001
$(MS, MW)$	2	5.39	.07	$\lambda_{jk}^{SW} = 0$	1	5.325	.02
$(MS, WS)$	2	19.71	< .001	$\lambda_{ik}^{MW} = 0$	1	19.645	< .001

Sample size and hypothesis tests:

- With small samples, “reality may be much more complex than indicated by the simplest model that passed the goodness-of-fit test.
- With large samples, “. . . statistically significant effects can be weak and unimportant.”

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

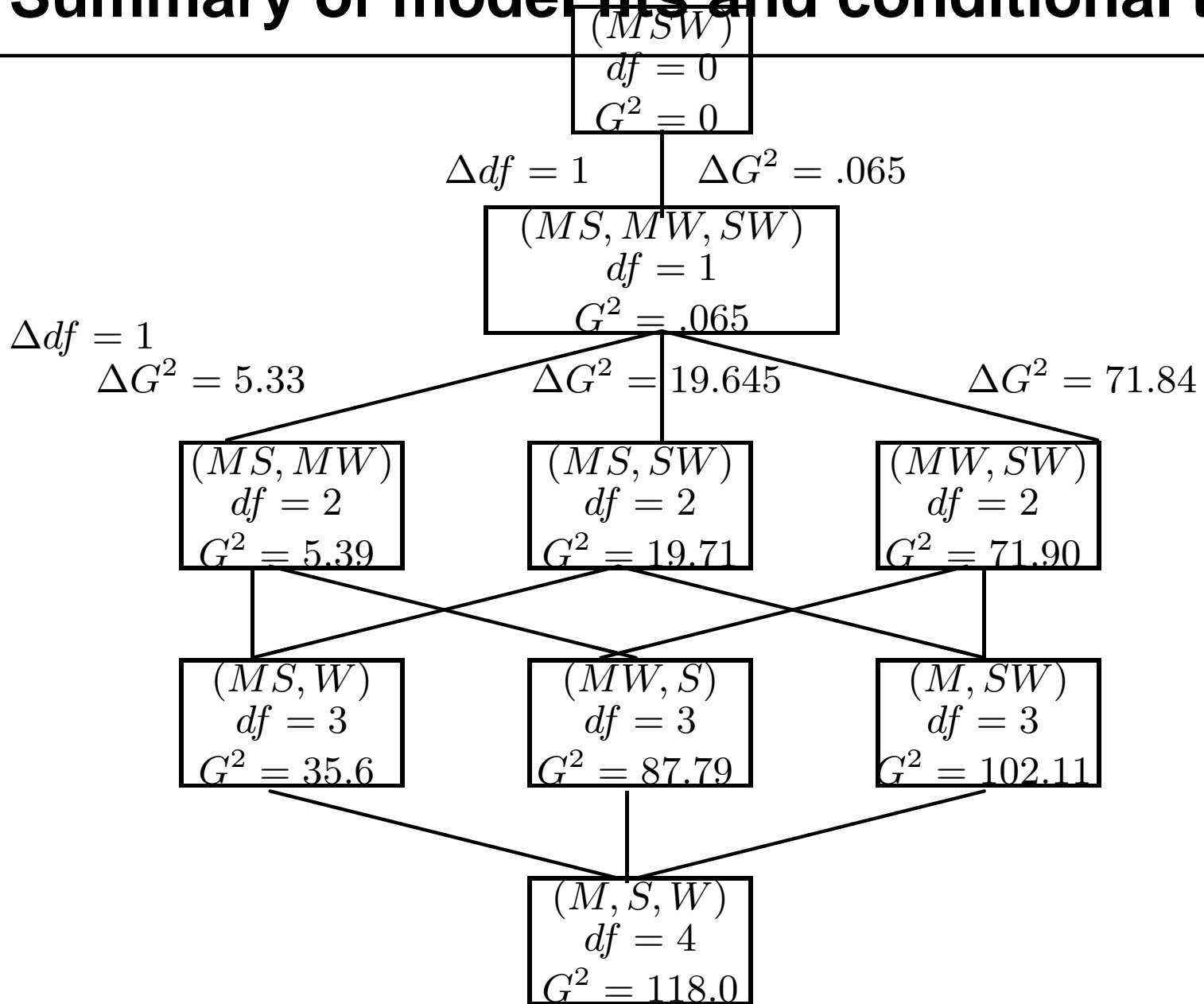
Log-linear Models for 4+Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Summary of model fits and conditional tests



Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

**Inference for Log-linear Models**

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Confidence Intervals for Odds Ratios

... a bit more informative than hypothesis testing alone.

We know that odds ratios are direct functions of log-linear model parameters.

e.g., Suppose that we are interested in an estimate of the  $SW$  partial odds ratio. Using the estimated parameters of the homogeneous association model ( $MS, MW, SW$ ) (estimated using *SAS/GENMOD*,

	Estimate	ASE
$\hat{\lambda}_{low,low}^{SW} =$	.3847	.1667
$\hat{\lambda}_{low,hi}^{SW} =$	.0000	.0000
$\hat{\lambda}_{hi,low}^{SW} =$	.0000	.0000
$\hat{\lambda}_{hi,hi}^{SW} =$	.0000	.0000

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Confidence Intervals for Odds Ratios

$$\begin{aligned}\log(\hat{\theta}_{SW(i)}) &= \hat{\lambda}_{\text{low,low}}^{SW} + \hat{\lambda}_{\text{hi,hi}}^{SW} - \hat{\lambda}_{\text{low,hi}}^{SW} - \hat{\lambda}_{\text{hi,low}}^{SW} \\ &= .3827 + 0.00 - 0.00 - 0.00 \\ &= .3827\end{aligned}$$

and  $\hat{\theta}_{SW(i)} = e^{.3827} = 1.4662$ .

A  $(1 - \alpha) \times 100\%$  confidence interval for  $\log(\theta_{SW(i)})$  is

$$\log(\hat{\theta}_{SW(i)}) \pm z_{\alpha/2}(ASE)$$

e.g., A 95% confidence interval for the log of the supervisor by worker satisfaction odds ratio is

$$.3827 \pm 1.96(.1667) \longrightarrow (.05596, .70943)$$

For confidence interval for the odds ratio take the anti-log of the interval for  $\log(\theta_{SW(i)})$  to get the confidence interval for the odds ratio. So the 95% confidence interval for the (partial) odds ratio  $\theta_{SW(i)}$  is

$$(e^{.05596}, e^{.70943}) \longrightarrow (1.058, 2.033)$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Statistical versus Practical Significance

## and Large Samples

For the blue collar worker data, two models that could be a good model (representation) of the data.

Criterion	In favor of ( $MS, MW$ )	In favor of ( $MS, MW, SW$ )
Model goodness of fit $G^2 =$	5.39	.065
	$df = 2, p = .07$	with $df = 1, p = .80$
Largest adjusted residual	1.69	.25
Likelihood ratio test of $\lambda_{jk}^{SW} = 0$	—	$G^2 = 5.325,$ $df = 1, p = .02$
Complexity	simpler	more complex

Question: Do we really need the SW partial association?  
Weak effect, but is significant due to large sample size  
( $n = 715$ ) relative to table size ( $N = 2 \times 2 \times 2 = 8$ )?

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Deciding on a Model: Subjective Judgment

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

When there is more than one reasonable model, additional things to consider in choosing a single “best” model are

1. Substantive importance and considerations.
2. Closeness between observed and fitted odds ratios.
3. Dissimilarity index.
4. Correlations between observed and fitted values.
5. Information Criteria (AIC, BIC & others).
6. Analysis of association.

Let's see if these help in making the decision between  $(MS, MW)$  and  $(MS, MW, SW)$ .



# Similarity between the Fitted Odds Ratios

If two models have nearly the same values for the odds ratio, then choose the simpler one.

What constitutes “nearly the same values” is a subjective decision.

Fitted partial odds ratios based two best model and the observed partial odds ratios for the worker satisfaction data:

		Fitted Odds Ratio		
		W–S	M–W	M–S
	Model			
	$(MS, MW)$	1.00	2.40	4.32
	$(MS, WS, MW)$	1.47	2.11	4.04
Observed or	$(MSW)$ –level 1	1.55	2.19	4.26
	$(MSW)$ –level 2	1.42	2.00	3.90

They seem similar. Whether they are “close” enough, that depends on purpose or uses you’ll make of the results.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Dissimilarity Index

For a table of with

Any dimension (e.g.,  $(I \times J)$ ,  $(I \times J \times K)$ , etc).

Cell counts equal to  $n_i = np_i$ .

Fitted counts equal to  $\hat{\mu}_i = n\hat{\pi}_i$ .

The “**Dissimilarity Index**” is a summary statistic of how close the fitted values of a model are to the data. It equals

$$D = \frac{\sum_i |n_i - \hat{\mu}_i|}{2n} = \frac{\sum_i |p_i - \hat{\pi}_i|}{2}$$

Properties of  $D$ :

- $0 \leq D \leq 1$ .
- $D$  = the proportion of sample cases that need to move to a different cell to have the model fit perfectly.
- Small  $D$  means that there is little difference between fitted values and observed counts.
- Larger  $D$  means that there is a big difference between fitted values and observed counts.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# More Properties of the Dissimilarity Index

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- $D$  is an estimate of the change,  $\Delta$ , which measures the lack-of-fit of the model in the population.

When the model fits perfectly *in the population*,

- ◆  $\Delta = 0$
- ◆  $D$  overestimates the lack-of-fit (especially for small samples).
- For large samples when the model does not fit perfectly,
  - ◆  $G^2$  and  $X^2$  will be large.
  - ◆  $D$  reveals when the lack-of-fit is important in a practical sense.
- Rule-of-thumb:  $D < .03$  indicates non-important lack-of-fit.



# Example of the Dissimilarity Index

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Bluecollar Example: For the model ( $MW, MS$ )

$$D = \frac{55.2306}{2(715)} = .039$$

We would need to move 3.9% percent of the observations to achieve a perfect fit of the model ( $MW, MS$ ) to observed (sample) data.

For the model ( $MW, MS, SW$ ),

$$D = \frac{5.8888}{2(715)} = .004$$

We would need to move .4% of the observations to achieve a perfect fit of the models ( $MW, MS, SW$ ) to the observed data.

Which one? Possibly the model of conditional independence.



# Correlations between Counts and Fitted Counts

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

A large value indicates that the observed and fitted are “close”.

Worker satisfaction example:

For the model of conditional independence ( $MW, MS$ ),

$$r = .9906$$

and for the model of homogeneous association

$$r = .9999$$



# Information Criteria

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- Indices (statistics) that weigh goodness-of-fit of model to data, complexity of the model, and in some cases sample size.
- Good way to choose among reasonable models.
- Does not require the models be nested.
- Akaike Information Criteria (AIC):

$$AIC = -2 \log(L) + 2(\text{number of parameters})$$

- Bayesian Information Criteria (BIC): References:

Raftery, A.E. (1985). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society, Series B*.

Raftery, A. E. (1986). Choosing models for cross-classifications. *American Sociological Review, 51*, 145–146.

Spiegelhalter, D.J. and Smith, A.F.M. (1982). Bayes Factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society, Series B, 44*, 377–387.

1998 or 1999 special issue of *Sociological Methodology & Research* on the BIC statistic.



# The Bayesian Approach to Model Selection

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Another way of making the trade-off between a simple parsimonious model (practical significance) and a more complex and closer to reality model (statistical significance), besides using just  $G^2$  and  $df$ .

Suppose you are considering a model, say  $\mathcal{M}_o$ , and you are comparing it to the saturated model,  $\mathcal{M}_1$ .

Which model gives a better description of the main features of the reality as reflected in the data?

More precisely, which of  $\mathcal{M}_o$  and  $\mathcal{M}_1$  is more likely to be the “true” model?

Answer: posterior odds

$$B = \frac{P(\mathcal{M}_o|X)}{P(\mathcal{M}_1|X)}$$



# The BIC statistic

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Skipping many details....for large samples:

$$BIC = -2 \log B = G^2 - (df) \log N$$

where  $N$  =total number of observations.

If BIC is negative, accept  $\mathcal{M}_o$ ; it's preferable to the saturated model.

When comparing a set of models, choose the one with the smallest BIC value. (*The models do not have to be nested*). This procedure provides you with a consistent model in the sense that in large samples, it chooses the correct model with high probability.



# Example of Information Criteria

Example: Worker Job Satisfaction × Supervisor's Job Satisfaction × Quality of Management

$N = 715$

Model	$df$	$G^2$	$p$ -value	BIC	# of Parameters	AIC
(MSW)	0	0.00	1.000	.00	8	—
(MS)(MW)(SW)	1	.06	.800	-6.51	7	-13.94
(MW)(SW)	2	71.90	.000	58.76	6	59.90
(MS)(WS)	2	19.71	.000	6.57	6	7.71
(SM)(WM)	2	5.39	.068	-7.75	6	-6.61
(MW)(S)	3	87.79	.000	68.07	5	77.79
(WS)(M)	3	102.11	.000	82.07	5	92.11
(MS)(W)	3	35.60	.000	15.88	5	25.60
(M)(S)(W)	4	118.00	.000	91.71	4	110.00

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# “Analysis of Association” Table

Another aid in helping to decide between practical and statistical significance. This comes from Leo Goodman’s toolbox.

This is useful even when the sampling distribution of  $G^2$  is not well approximated by chi-squared distribution.

We use the  $G^2$  from independence as a measure of the total association in the data and see how much association is accounted for by certain effects.

Effect	Models	$\Delta df$	$\Delta G^2$	$p$ -value	%	Cummul. Percent
MS	(M,S,W) - (MS,W)	1	82.40	.000	69.8%	69.8%
MW	(MS,W) - (MS,MW)	1	30.21	.000	25.6%	95.4%
SW	(MS,MW) - (MS,MW,SW)	1	5.33	.021	4.6%	100.0%
MSW	(MS,MW,SW) - (MSW)	1	.06	.800	0.0%	0.0%
total	(M,S,W)	4	118.00			

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# “Analysis of Association” Table

Effect	Models	$\Delta df$	$\Delta G^2$	$p$ -value	%	Cummul. Percent
MS	(M,S,W) - (MS,W)	1	82.40	.000	69.8%	69.8%
MW	(MS,W) - (MS,MW)	1	30.21	.000	25.6%	95.4%
SW	(MS,MW) - (MS,MW,SW)	1	5.33	.021	4.6%	100.0%
MSW	(MS,MW,SW) - (MSW)	1	.06	.800	0.0%	0.0%
total	(M,S,W)	4	118.00			

- $\Delta G^2$  = the difference between goodness-of-fit statistics for the models indicated.
- $\Delta df$  = the corresponding difference between the models'  $df$ .
- The column labeled “ $p$ -value” really shouldn't be in this table.
- Percent =  $\Delta G^2 / 118.00$ . Note: 118.00 is the  $G^2$  from (M,S,W).
- Cumulative percent = sum of “Percent” of current and all rows above the current one.

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Log-linear Models for 4+–Way Tables

They are basically the same as models for 3–way tables, but just more complex. They can have many more 2– and 3–way associations, as well as higher–way associations.

Example: These data come from a study by Thornes & Collard (1979), described by Gilbert (1981), and analyzed by others (including Agresti, 1990; Meulman & Heiser, 1996).

A sample of men and woman who filed for petition for divorce (they weren't married to each other), and a similar sample of married people were asked

- “Before you were married to your (former) husband/wife, had you ever made love to anyone else?”
- “During your (former) marriage, (did you have) have you had any affairs or tried sexual encounters with another man/woman?”

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Example of 4-Way Table

These data form a 4-way,  $2 \times 2 \times 2 \times 2$  table with variables

**G** for gender

**E** or **EMS** for whether reported extramarital sex.

**P** or **PMS** for whether reported premarital sex.

**M** for marital status.

		Gender							
		Women				Men			
Marital Status	PMS:	Yes		No		Yes		No	
	EMS:	Yes	No	Yes	No	Yes	No	Yes	No
Divorce		17	54	36	214	28	60	17	68
Still Married		4	25	4	322	11	42	4	130

For these data, a good model (perhaps the best) is  $(GP, MEP)$

$$G^2 = 8.15 \quad df = 6 \quad p = .23$$

(we'll talk about how we arrived at this model later).

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Parameter Estimates for example

Estimated Parameters for the highest-way associations in the model (from *SAS/GENMOD*)

Parameter				<i>df</i>	Estimate	ASE	Wald	<i>p</i>
PG	yes	women		1	-1.3106	.1530	73.4249	< .001
PG	yes	men		0	0.0000			
PG	no	women		0	0.0000			
PG	no	men		0	0.0000			
MEP	div	yes	yes	1	-1.7955	.5121	12.2948	< .001
MEP	div	yes	no	1	0.0000			
MEP	div	no	yes	1	0.0000			
MEP	div	no	no	0	0.0000			
MEP	mar	yes	yes	1	0.0000			
MEP	mar	yes	no	0	0.0000			
MEP	mar	no	no	0	0.0000			
MEP	mar	no	no	0	0.0000			

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Interpretation of GP Partial Association

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Since  $\hat{\lambda}_{\text{women,yes}}^{GP} = -1.3106$ , given EMS and marital status the odds of PMS for women is

$$e^{-1.3106} = .2696$$

times the odds for men.

Alternatively, given EMS and marital status, the odds of PMS for men is

$$e^{1.3106} = 1/.2696 = 3.71$$

times larger than the odds for women.



# Using the Fitted Values

... from the (*GP, MEP*) model...

PMS:	Gender							
	Women				Men			
	Yes		No		Yes		No	
EMS:	Yes	No	Yes	No	Yes	No	Yes	No
Divorced	18.67	47.30	38.40	204.32	26.33	66.70	14.60	77.68
Married	6.22	27.80	5.80	327.49	8.78	39.20	2.20	124.51

For EMS=yes and martial status=divorced

$$\frac{\text{odds(PMS for woman)}}{\text{odds(PMS for man)}} = \frac{(18.67)(14.60)}{(38.40)(26.33)} = .2696$$

or for EMS=yes and martial status=married

$$\frac{\text{odds(PMS for woman)}}{\text{odds(PMS for man)}} = \frac{(6.22)(2.20)}{(5.80)(8.78)} = .2696$$

which also equals the value if we use EMS=no.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Martial–PMS–EMS partial association

Can use either our estimated parameters or using fitted values.  
Fitted values from the  $(GP, MEP)$  model

PMS:	Gender							
	Women				Men			
	Yes		No		Yes		No	
EMS:	Yes	No	Yes	No	Yes	No	Yes	No
Divorced	18.67	47.30	38.40	204.32	26.33	66.70	14.60	77.68
Married	6.22	27.80	5.80	327.49	8.78	39.20	2.20	124.51

The odds ratio for marital status and extramarital sex for those who did and those who did not have premarital sex.

**PMS=yes:** Of those who had premarital sex, the odds of divorce given the person had extramarital sex is

$$\hat{\theta}_{ME|PMS=yes} = \frac{(18.67)(27.80)}{(6.22)(47.30)} = 1.76$$

times larger than the odds of divorce given the person did not have extramarital sex.

Note: We could also use the fitted values for men

$$\hat{\theta}_{ME|PMS=yes} = \frac{(26.33)(39.20)}{(8.78)(66.70)} = 1.76$$



# Martial–PMS–EMS partial association

**PMS=no** Of those who did not have premarital sex, the odds of divorce given the person had extramarital sex is

$$\hat{\theta}_{ME|PMS=no} = \frac{(38.40)(327.49)}{(5.50)(204.32)} = 10.62$$

times larger than the odds of divorce given the person did not have extramarital sex.

**3–way EMP association:** The partial odds ratio for marital status and extramarital sex given the person did not have premarital sex are

$$\frac{10.62}{1.76} = 6.03$$

times larger than the partial odds ratio given the person did have premarital sex.

(We could have arrived at the same interpretation of the partial associations by using the parameters of the log-linear model.)

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# MEP Partial association

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

What do the odds ratios equal in terms of the model parameters?

Let  $i$  index M (marital status),  $j$  index EMS,  $k$  index PMS, and  $l$  index Gender,

$$\begin{aligned} \log \left( \frac{\mu_{11kl}\mu_{22kl}}{\mu_{12kl}\mu_{21kl}} \right) &= \log(\mu_{11kl}) + \log(\mu_{22kl}) - \log(\mu_{12kl}) - \log(\mu_{21kl}) \\ &= \lambda_{11}^{ME} + \lambda_{22}^{ME} - \lambda_{12}^{ME} - \lambda_{21}^{ME} \\ &\quad + \lambda_{11k}^{MEP} + \lambda_{22k}^{MEP} - \lambda_{12k}^{MEP} - \lambda_{21k}^{MEP} \end{aligned}$$

$\hat{\lambda}_{11}^{ME} = 2.3960$  (Divorced and had EMS),  
all other  $\hat{\lambda}_{ij}^{ME}$ 's equal zero.

$\hat{\lambda}_{111}^{MEP} = -1.7955$  (Divorced, had EMS & had PMS),  
all other  $\hat{\lambda}_{ijk}^{MEP}$ 's equal zero.



# MEP Partial association

Of those who had PMS ( $k = 1 = \text{yes}$ ), the estimated odds ratio for marital status and extramarital sex equals

$$\exp(2.3626 - 1.7955) = \exp(.5671) = 1.76 \quad (95\%CI : 0.33, 9.21)$$

Of those who did not have PMS ( $k = 2 = \text{no}$ ), the estimated odds ratio for marital status and extramarital sex equals

$$\exp(2.3626) = 10.62$$

and the ratio of the odds ratios equals

$$\exp(2.3626 - .5671) = \exp(+1.7955) = 6.03 \quad \text{or} \quad \exp(-1.7955) = 1/6.03$$

$$(95\%CI : 7.92, 14.24)$$

Before summarizing the findings, how to compute CIs for these odds ratios. . .

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# $(1 - \alpha) \times 100\%$ CI for MEP Partial association

Of those who had PMS ( $k = 1 = \text{yes}$ ), the estimated odds ratio for marital status and extramarital sex equals

$$\exp(\hat{\lambda}_{11}^{ME} - \hat{\lambda}_{111}^{MEP}) = \exp(2.3626 - 1.7955) = \exp(.5671) = 1.76$$

Need the variances and covariance of parameters:

$$\Sigma = \begin{pmatrix} \sigma_{ME}^2 & \sigma_{ME,MEP} \\ \sigma_{ME,MEP} & \sigma_{MEP}^2 \end{pmatrix} = \begin{pmatrix} 0.14962 & -0.14962 \\ -0.14962 & 0.2622 \end{pmatrix}$$

$$\text{se}(\hat{\lambda}_{11}^{ME} - \hat{\lambda}_{111}^{MEP}) = \sqrt{\sigma_{ME}^2 + \sigma_{MEP}^2 - 2\sigma_{ME,MEP}}$$

$$= \sqrt{0.14962 + 0.2622 - 2(-0.14962)} = 0.84323$$

So a  $(1 - \alpha) \times 100\%$  CI for the log of the MEP partial for those who had PMS is

$$1.76 \pm 1.96(0.84323) \longrightarrow .5671 \pm 1.6527 \longrightarrow (-1.0856, 2.2198)$$

and for the MEP partial association for those who had PMS

$$(\exp(-1.0856), \exp(2.2198)) \longrightarrow (0.33, 9.21)$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Logit–Log-linear Model Connection

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Log-linear models:

- All variables are considered response variables; no distinction is made between response and explanatory variables (in terms of a variable's role/treatment in an analysis).
- Distribution = Poisson.
- Link = Log.

Logit Models:

- Represent how a binary response variable depends (or is related to) a set of explanatory variables.
- Distribution = Binomial.
- Link = Logit.



# Logit/Log-linear Model Connection

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

**Logit-Log-linear Model  
Connection**

Strategies in (Log-linear) Model  
Selection

Logit and log-linear models are related

Logit models are equivalent to certain log-linear model.

Log-linear models are more general than logit models.

More specifically,

1. For a log-linear model, you can construct logits for 1 (binary) response variable to help interpret the log-linear model.
2. Logit models with categorical explanatory variables have equivalent log-linear models.

The relationship is useful. . . use Logit models to interpret log-linear models



# Using Logit models to interpret loglinear models

For 2–way tables:

Interpret log-linear models by looking at differences between  $\lambda$ 's, which equal log of odds and functions of  $\lambda$ 's equal odds ratios.

For 3–way tables: The blue collar worker data and the homogeneous association model ( $MW, MS, SW$ ),

$$\log \mu_{ijk} = \lambda + \lambda_i^M + \lambda_j^S + \lambda_k^W + \lambda_{ij}^{MS} + \lambda_{ik}^{MW} + \lambda_{jk}^{SW}$$

If we focus on worker's job satisfaction, then we consider

$$\pi_{ij} = \text{Prob}(\text{Hi worker satisfaction} | M = i, S = j)$$

and the logit model for worker job satisfaction is

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \text{logit}(\pi_{ij}) \\ &= \log \left( \frac{P(\text{Hi worker satisfaction} | M = i, S = j)}{P(\text{Lo worker satisfaction} | M = i, S = j)} \right) \\ &= \log(\mu_{ij2} / \mu_{ij1}) \\ &= \log(\mu_{ij2}) - \log(\mu_{ij1}) \end{aligned}$$

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Using Logit models to interpret loglinear models

For 3–way tables (continued):

$$\begin{aligned}\text{logit}(\pi_{ij}) &= (\lambda_2^W - \lambda_1^W) + (\lambda_{i2}^{MW} - \lambda_{i1}^{MW}) + (\lambda_{j2}^{SW} - \lambda_{j1}^{SW}) \\ &= \alpha + \beta_i^M + \beta_j^S\end{aligned}$$

This is the additive effects logit model where

■  $\alpha = (\lambda_2^W - \lambda_1^W)$ . a constant.

■  $\beta_i^M = (\lambda_{i2}^{MW} - \lambda_{i1}^{MW})$ .

The relationship (effect) of management quality between (on) worker job satisfaction is the same at each level of supervisor's job satisfaction.

■  $\beta_j^S = (\lambda_{j2}^{SW} - \lambda_{j1}^{SW})$ .

The relationship (effect) of supervisor's job satisfaction between (on) worker job satisfaction is the same at each level of management quality.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Example of 4-way Table

Marital status  $\times$  EMS  $\times$  PMS  $\times$  Gender — A good model for these data is  $(GP, MEP)$ .

We can use a logit model formulation to help interpret the results of the  $(GP, MEP)$  log-linear model,

$$\begin{aligned} \log \mu_{ijkl} &= \lambda + \lambda_i^M + \lambda_j^E + \lambda_k^P + \lambda_l^G + \lambda_{ij}^{ME} + \lambda_{ik}^{MP} \\ &\quad + \lambda_{jk}^{EP} + \lambda_{kl}^{GP} + \lambda_{ijk}^{MEP} \end{aligned}$$

We will focus on marital status and form (log) odds of divorce,

$$\begin{aligned} \log\left(\frac{\pi_{1jkl}}{\pi_{2jkl}}\right) &= \log(\pi_{1jkl}) - \log(\pi_{2jkl}) \\ &= (\lambda_1^M - \lambda_2^M) + (\lambda_{1j}^{ME} - \lambda_{2j}^{ME}) \\ &\quad + (\lambda_{1k}^{MP} - \lambda_{2k}^{MP}) + (\lambda_{1jk}^{MEP} - \lambda_{2jk}^{MEP}) \\ &= \alpha + \beta_j^E + \beta_k^P + \beta_{jk}^{EP} \end{aligned}$$

and the estimated parameters for the logit model using the ones from the log-linear model. . . .

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Marital status × EMS × PMS × Gender

- Overview

---

- Log-linear Models for 2-way Tables

---

- Parameter Constraints

---

- Log-linear Models for 3-way Tables

---

- Inference for Log-linear Models

---

- statistical versus practical significance

---

- Log-linear Models for 4+-Way Tables

---

- Logit-Log-linear Model Connection**

---

- Strategies in (Log-linear) Model Selection

		Loglinear Model Parameters		Logit Model
		Marital Status		Parameters
		Divorced	Married	
		$\hat{\lambda}_1^M = -.4718$	$\hat{\lambda}_2^M = 0.00$	$\hat{\alpha} = -.4718$
<b>EMS</b>				
yes		$\hat{\lambda}_{11}^{ME} = 2.3626$	$\hat{\lambda}_{21}^{ME} = 0.0000$	$\hat{\beta}_1^E = 2.3626$
no		$\hat{\lambda}_{12}^{ME} = 0.0000$	$\hat{\lambda}_{22}^{ME} = 0.0000$	$\hat{\beta}_2^E = 0.0000$
<b>PMS</b>				
yes		$\hat{\lambda}_{11}^{MP} = 1.0033$	$\hat{\lambda}_{21}^{MP} = 0.0000$	$\hat{\beta}_1^P = 1.0033$
no		$\hat{\lambda}_{12}^{MP} = 0.0000$	$\hat{\lambda}_{22}^{MP} = 0.0000$	$\hat{\beta}_2^P = 0.0000$
<b>EMS</b>	<b>PMS</b>			
yes	yes	$\hat{\lambda}_{111}^{MEP} = -1.796$	$\hat{\lambda}_{211}^{MEP} = 0.0000$	$\hat{\beta}_{11}^{EP} = -1.796$
yes	no	$\hat{\lambda}_{112}^{MEP} = 0.0000$	$\hat{\lambda}_{212}^{MEP} = 0.0000$	$\hat{\beta}_{12}^{EP} = 0.0000$
no	yes	$\hat{\lambda}_{121}^{MEP} = 0.0000$	$\hat{\lambda}_{221}^{MEP} = 0.0000$	$\hat{\beta}_{12}^{EP} = 0.0000$
no	no	$\hat{\lambda}_{122}^{MEP} = 0.0000$	$\hat{\lambda}_{222}^{MEP} = 0.0000$	$\hat{\beta}_{22}^{EP} = 0.0000$



# Loglinear–Logit Model Equivalence

Marital status seems like a response/outcome variable, while the others seem to be more explanatory/predictor variables.

So rather than fitting a log-linear model, we could treat the data as if we have independent Binomial samples, and fit a logit model where the (binary) response variable is marital status and the explanatory variables are Gender, EMS, and PMS.

		Marital Status			
Gender	PMS	EMS	Divorced	Married	total
Women	yes	yes	17	4	21
		no	54	25	79
	no	yes	36	4	40
		no	214	322	536
Men	yes	yes	28	11	39
		no	60	42	102
	no	yes	17	4	21
		no	68	130	198

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+–Way Tables

Logit–Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Loglinear–Logit Model Equivalence (continued)

The saturated logit model for these data

$$\log \frac{P(\text{divorced}_{ijk})}{P(\text{married}_{ijk})} = \alpha + \beta_i^G + \beta_j^E + \beta_k^P + \beta_{ij}^{GE} + \beta_{ik}^{GP} + \beta_{jk}^{EP} + \beta_{ijk}^{GEP}$$

Logit Model	<i>df</i>	$G^2$	<i>p</i>
E,G,P	4	13.63	.001
GP,E	3	13.00	< .001
EG,P	3	10.75	.010
EP,G	3	.70	.873
EG,GP	2	10.33	< .001
EP,GP	2	.44	.803
EG,EP	2	.29	.865
EG,EP,GP	1	.15	.700
EGP	0	0.00	1.00

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3–way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+–Way Tables

Logit–Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# The “best” logit model is $(EP, G)$

$$\text{logit}(\pi_{ijk}) = \alpha + \beta_i^G + \beta_j^E + \beta_k^P + \beta_{jk}^{EP}$$

which is different from the logit model that we used to interpret our log-linear model  $(GP, EMP)$ , i.e.,

$$\text{logit}(\pi_{ijk}) = \alpha + \beta_j^E + \beta_k^P + \beta_{jk}^{EP}$$

The  $(GP, EMP)$  log-linear model is not equivalent to any logit model that we could fit to the data with marital status as the response variable because. . . .

- When we consider the data as 8 independent Binomial samples, the “row” margin corresponding to the total number of observations for each Gender  $\times$  EMS  $\times$  PMS combination is “fixed.”
- When we fit a log-linear model to the data, we should always include parameters to ensure that the GEP margin is fit perfectly.
- If marital status is our response variable, we are not interested in the relationship between/among Gender, EMS, and PMS, except with respect to how they are related to marital status.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# The log-linear equivalent to $(G, EP)$ logit

$$\text{logit}(\pi_{ijk}) = \alpha + \beta_i^G + \beta_j^E + \beta_k^P + \beta_{jk}^{EP}$$

is the  $(GEP, MEP, GM)$  log-linear model,

$$\begin{aligned} \mu_{ijkl} = & \lambda + \lambda_i^G + \lambda_j^E + \lambda_k^P + \lambda_{ij}^{GE} + \lambda_{ik}^{GP} + \lambda_{jk}^{EP} + \lambda_{ijk}^{GEP} \\ & + \lambda_l^M + \lambda_{il}^{GM} + \lambda_{jl}^{EM} + \lambda_{kl}^{PM} + \lambda_{ljk}^{MEP} \end{aligned}$$

- When odds are computed for marital status using a log-linear model with  $\lambda_{ijk}^{GEP}$ , all terms associated with this association and lower order terms drop out; that is,

$$\lambda, \lambda_i^G, \lambda_j^E, \lambda_k^P, \lambda_{ij}^{GE}, \lambda_{ik}^{GP}, \lambda_{jk}^{EP}, \lambda_{ijk}^{GEP}$$

- The log-linear model  $(GEP, MEP, GM)$  will have the exact same  $df$  and fit statistics as the  $(EP, G)$  logit model.
- The estimated parameters of the logit model are equal to differences of estimated log-linear model parameters.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# The log-linear/logit equivalents

For all the logit models that we fit to these data, here are the corresponding loglinear models:

Logit Model	Loglinear Model	<i>df</i>	$G^2$	<i>p</i>
E,G,P	EGP,ME,MG,MP	4	13.63	.001
GP,E	EGP,MGP,ME	3	13.00	< .001
EG,P	EGP,MEG,MP	3	10.75	.010
EP,G	EGP,MEP,MG	3	.70	.873
EG,GP	EGP,MEG,MGP	2	10.33	< .001
EP,GP	EGP,MEP,MGP	2	.44	.803
EG,EP	EGP,MEG,MEP	2	.29	.865
EG,EP,GP	EGP,MEG,MEP,MGP	1	.15	.700
EGP	EGPM	0	0.00	1.00

Overview

Log-linear Models for 2-way Tables

Parameter Constraints

Log-linear Models for 3-way Tables

Inference for Log-linear Models

statistical versus practical significance

Log-linear Models for 4+-Way Tables

Logit-Log-linear Model Connection

Strategies in (Log-linear) Model Selection



# Strategies in (Log-linear) Model Selection

First, when to use logit models and when to use log-linear models.

When one variable is a response variable and the rest are explanatory variables, you can use either logit models or log-linear models; however, the logit models are easier (better) to use.

The logit models can be fit directly and are advantageous in this situation in that the logit model is simpler; that is, the logit model formulations have fewer parameters than the equivalent log-linear model.

If the response variable has more than 2 levels, you can use a multicategory logit model (to be covered in a later lecture).

If you use log-linear models, the highest-way associations among the explanatory variables should be included in all models.

Whether you use logit or log-linear formulations, the results will be the same regardless of which formulation you use.

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Two or Model Response Variables

... Then the log-linear model should be used.

Log-linear models are more general than logit models.

In the Marital status  $\times$  Gender  $\times$  EMS  $\times$  PMS example, with the log-linear models we can examine not only how marital status is related to EMS, PMS and gender, but we can also examine associations between (for example) gender and EMS or PMS.

These classes are multivariate logit models:

- “Standard” type (see McCullah & Nelder)
- IRT models are multivariate logit models.
- Other kinds (see Anderson & Böckenholt, 2000; Anderson & Yu, 2007).

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Model selection strategies with Log-linear model

The more variables, the more possible models that exist.

We'll talk about **strategies for more of an exploratory** study here and later we'll talk more specifically about strategies for hypothesis/substantive theory guided studies (i.e., association graphs).

- (1.) Determine whether some variables are **responses** and others are explanatory variables.
  - ◆ Terms for associations among the explanatory variables should always be included in the model.
  - ◆ Focus your model search on models that relate the responses to explanatory variables.
  
- (2.) If a margin is **fixed by design**, then a term corresponding to that margin should always be included in the log-linear model (to ensure that the marginal fitted values from the model equal to observed margin). This reduces the set of ~~models that need to be considered.~~

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection



# Model selection strategies with Log-linear model

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+ -Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

- (3.) Try to determine the level of complexity that is necessary by fitting models with
  - ◆ marginal/main effects only.
  - ◆ all 2-way associations.
  - ◆ all 3-way associations.
  - ⋮
  - ◆ all highest-way associations.
- You can use a backward elimination strategy (analogous to one we discussed for logit models) or a stepwise procedure (but don't use computer algorithms for doing this).

Example of backward elimination and as promised how it was decided that  $(EGP, GP)$  was a good log-linear model for the  $EMS \times PMS \times Gender \times Marital Status$  data...



# Backward Elimination

Overview

Log-linear Models for 2-way  
Tables

Parameter Constraints

Log-linear Models for 3-way  
Tables

Inference for Log-linear Models

statistical versus practical  
significance

Log-linear Models for  
4+-Way Tables

Logit-Log-linear Model  
Connection

Strategies in (Log-linear) Model  
Selection

Stage	Model	$G^2$	$df$	Best Model
Initial	(EMP,EGM,EGP,GMP)	0.15	1	
1	(EMP,EGM,GMP)	0.19	2	*
	(EMP,EGM,EGP)	0.29	2	
	(EMP,EGP,GMP)	0.44	2	
	(EGM,EGP,GMP)	10.33	2	
2	(GP,EMP,EGM)	0.37	3	*
	(EG,EMP,GMP)	0.46	3	
	(EP,EGM,GMP)	10.47	3	
3	(EG,GM,GP,EMP)	0.76	4	*
	(EP,GP,MP,EGM)	10.80	4	
	(EMP,EGM)	67.72	4	
4	(GM,GP,EMP)	5.21	5	*
	(EG,GP,EMP)	5.25	5	
	(EG,GM,GP,EM,EP,MP)	13.63	5	
	(EG,GM,EMP)	70.10	5	
5	(GP,EMP)	8.15	6	*
	(GM,GP,EM,EP,MP)	18.13	6	
	(GM,EMP)	83.38	6	
6	(GP,EM,EP,MP)	21.07	7	*
	(G,EMP)	83.41	7	

(from Agresti, 1990).