

---

# Three–Way Tables

## Edpsy/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology



# Outline

Overview

● Outline

● Examples of 3-Way Tables

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

- Types of association
  1. Marginal & Partial tables.
  2. Marginal & Conditional odds ratios.
  3. Marginal & Conditional Independence/Dependence.
    - (a) Marginal Independence and Conditional Dependence.
    - (b) Marginal Dependence and Conditional Independence.
    - (c) Marginal and Conditional Dependence.
  4. Homogeneous association.
- Inference for Large Samples.
  1. Cochran-Mantel-Haenszel tests — Conditional independence.
  2. Estimating common odds ratio.
  3. Breslow-Day statistic — Testing homogeneity.
  4. Comments.
- Inference for Small Samples (a few comments).

# Examples of 3–Way Tables

Overview

● Outline

● Examples of 3–Way Tables

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3–Way Tables

Tests of Conditional Independence

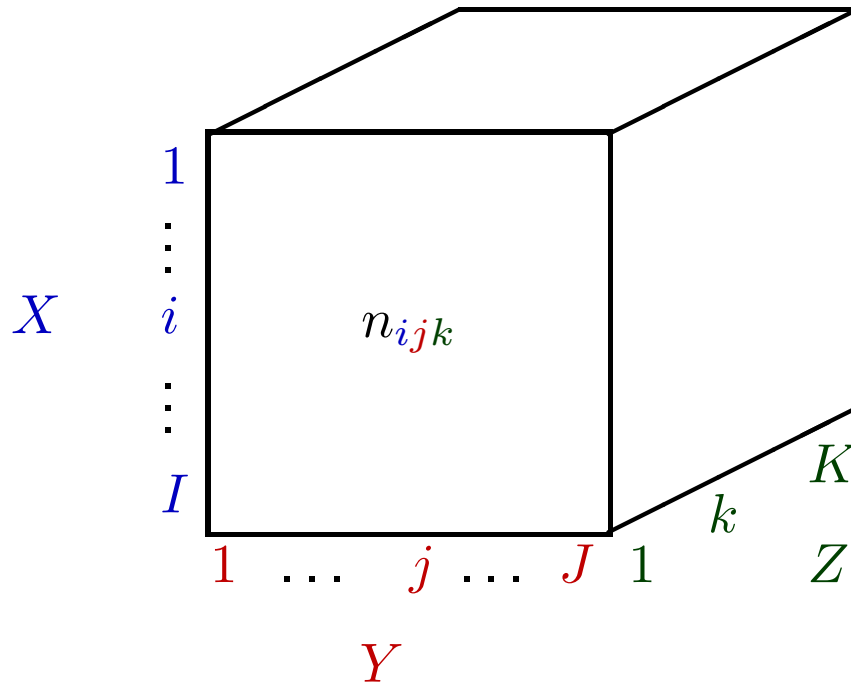
Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

- Smoking × Breathing × Age.
- Group × Response × Z (hypothetical).
- Boys Scouts × Delinquent × SES (hypothetical).
- Cal graduate admissions × gender × Department.
- Supervisor Job satisfaction × Worker Job satisfaction × Management quality.
- Race × Questions regarding media × Year.
- Employment status × Residence × Months after hurricane Katrina.

# 3-Way Contingency Table



Slices of this table are “*Partial Tables*”.

There are 3–ways to slice this table up.

1.  $K$  Frontal planes or  $XY$  for each level of  $Z$ .
2.  $J$  Vertical planes or  $XZ$  for each level of  $Y$ .
3.  $I$  Horizontal planes or  $YZ$  for each level of  $X$ .

Overview

Marginal and Partial Tables

- 3–Way Contingency Table
- Partial Tables & Marginal Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3–Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

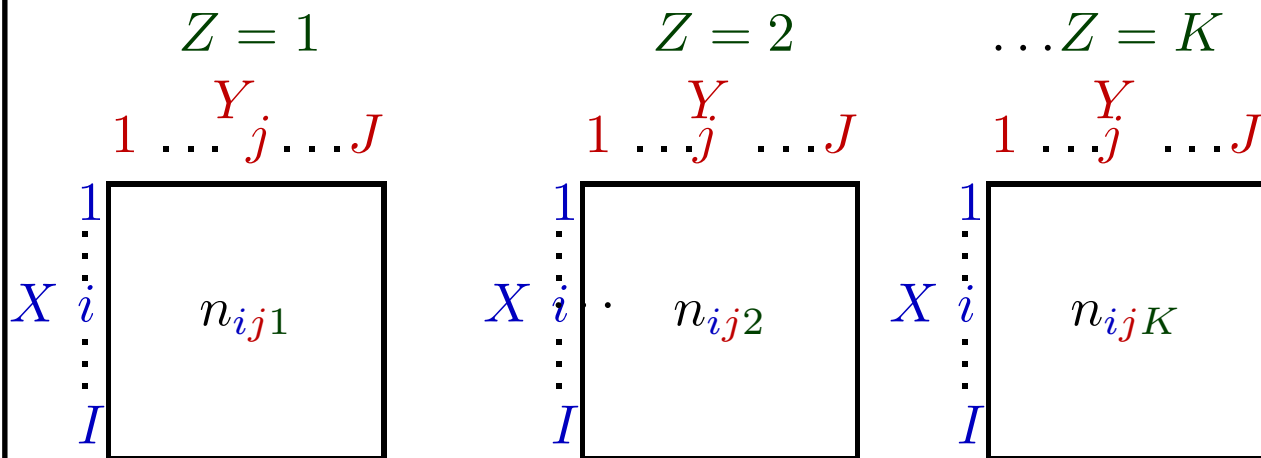
Testing Homogeneity of Odds Ratios

Concluding comments

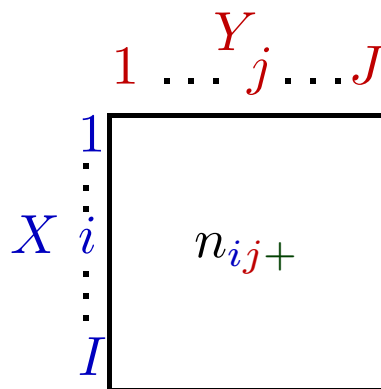
# Partial Tables & Marginal Tables

e.g.,  $XY$  tables for each level of  $Z$ ...

The Frontal planes of the box are  $XY$  tables for each level of  $Z$  are **Partial tables**:



Sum across the  $K$  levels of  $Z$  Yields the following **Marginal Table**



$$\text{where } n_{ij+} = \sum_{k=1}^K n_{ijk}$$

Overview

Marginal and Partial Tables

- 3-Way Contingency Table
- Partial Tables & Marginal Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

# Conditional or “Partial” Odds Ratios

Notation:

$n_{ijk}$  = observed frequency of the  $(i, j, k)$ th cell.

$\mu_{ijk}$  = expected frequency of the  $(i, j, k)$ th cell.

$$= n\pi_{ijk}$$

**Conditional Odds Ratios** are odds ratios between two variables for fixed levels of the third variable.

For fixed level of  $Z$ , the conditional  $XY$  association given  $k$ th level of  $Z$  is

$$\theta_{XY(k)} = \frac{\mu_{11k}\mu_{22k}}{\mu_{12k}\mu_{21k}} \quad \& \text{ more generally} \quad \theta_{ii',jj'(k)} = \frac{n_{ijk}n_{i'j'k}}{n_{i'jk}n_{ij'k}}$$

Conditional odds ratios are computed using the partial tables, and are sometimes referred to as measures of “*partial association*”.

If  $\theta_{XY(k)} \neq 1$ , then variables  $X$  and  $Y$  are “*Conditionally associated*”.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

- Conditional or “Partial” Odds Ratios

- Marginal Odds Ratios

- Example of Marginal vs Partial Odds Ratios

- Example: Partial Tables

- Marginal and Conditional Associations

- Four Situations

- Marginal

Independence/Conditional Dependence

- Marginal

Dependence/Conditional Independence

- Conditional Independence

- Example of Conditional Independence:  $CA\mathcal{L}$

- $CA\mathcal{L}$  Admissions Data by Department

- 3rd Example of Conditional Independence

- 3rd Example: Partial Tables

- Simpson's Paradox

- (Hypothetical) Example of Simpson's Paradox

- Picture of Simpson's Paradox

- Homogeneous Association

- Homogeneous Association Three-Way Tables

(continued)

# Marginal Odds Ratios

are the odds ratios between two variables in the marginal table.

For example, for the  $XY$  margin:

$$\mu_{ij+} = \sum_{k=1}^K \mu_{ijk}$$

and the “*Marginal Odds Ratio*” is

$$\theta_{XY} = \frac{\mu_{11+}\mu_{22+}}{\mu_{12+}\mu_{21+}} \quad \& \text{ more generally} \quad \theta_{ii',jj'} = \frac{\mu_{ij+}\mu_{i'j'+}}{\mu_{i'j+}\mu_{ij'+}}$$

With sample data, use  $n_{ijk}$  and  $\hat{\theta}$ .

**Marginal association can be very different from conditional association.**

The marginal odds ratios need not equal the partial (conditional) odds ratios.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or “Partial” Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence:  $CAL$

●  $CAL$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association Three-Way Tables

(continued)

# Example of Marginal vs Partial Odds Ratios

These data are from a study reported by Forthofer & Lehnen (1981) (Agresti, 1990). Measures on Caucasians who work in certain industrial plants in Houston were recorded.

- Response/outcome variable: breathing test result (normal, not normal).
- Explanatory variable: smoking status (never, current).
- Conditioning variable: age

Marginal Table (ignoring age):

Smoking Status	Test Result		
	Normal	Not Normal	
Never	741	38	779
Current	927	131	1058
	1668	169	1837

Marginal odds ratio:  $\hat{\theta} = 2.756$

$H_O : \theta = 1$  vs  $H_A : \theta \neq 1$  —  $G^2 = 32.382$ ,  $df = 1$ , &  $p$ -value  $< .001$ .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional

Independence:  $CAL$

●  $CAL$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables  
(continued)

# Example: Partial Tables

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional

Independence: *CAL*

● *CAL* Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

Smoking Status	Age < 40 Test Result		
	Normal	Not Normal	
Never	577	34	611
Current	682	57	739
	1259	91	1350

$$\hat{\theta} = 1.418$$

$$G^2 = 2.489$$

$$p\text{-value} = .115$$

Smoking Status	Age 40–59 Test Result		
	Normal	Not Normal	
Never	164	4	168
Current	245	74	319
	409	78	487

$$\hat{\theta} = 12.38$$

$$G^2 = 45.125$$

$$p\text{-value} < .001$$

Compare these odds ratios with the marginal odds ratio:

$$\hat{\theta} = 2.756$$

# Marginal and Conditional Associations

## Overview

## Marginal and Partial Tables

## Conditional and Marginal Odds Ratios

## ● Conditional or “Partial” Odds Ratios

## ● Marginal Odds Ratios

## ● Example of Marginal vs Partial Odds Ratios

## ● Example: Partial Tables

## ● Marginal and Conditional Associations

## ● Four Situations

## ● Marginal

## Independence/Conditional Dependence

## ● Marginal

## Dependence/Conditional Independence

## ● Conditional Independence

## ● Example of Conditional Independence: $\mathcal{CAL}$

## ● $\mathcal{CAL}$ Admissions Data by Department

## ● 3rd Example of Conditional Independence

## ● 3rd Example of Conditional Independence

## ● 3rd Example: Partial Tables

## ● Simpson's Paradox

## ● (Hypothetical) Example of Simpson's Paradox

## ● Picture of Simpson's Paradox

## ● Homogeneous Association

## ● Homogeneous Association Three-Way Tables

(continued)

- Independence = “No Association”.
- Dependence = “Association”.
- Marginal Independence means that  $\theta_{XY} = 1$
- Marginal Dependence means that  $\theta_{XY} \neq 1$
- Conditional Independence means that  $\theta_{XY(k)} = 1$  for all  $k = 1, \dots, K$ .
- Conditional Dependence means that  $\theta_{XY(k)} \neq 1$  for at least one  $k = 1, \dots, K$ .
- Marginal independence does **not** imply conditional independence.
- Conditional independence does **not** imply marginal independence.

# Four Situations

Situation	Marginal	Conditional	Comment
1	Independence	Independence	Not interesting
2	Independence	Dependence	“Conditional Dependence”
3	Dependence	Independence	“Conditional Independence”
4	Dependence	Dependence	“Conditional Dependence ”

Conditional dependence includes a number of different cases, which we have terms to refer to them:

- Simpson’s paradox.
- Homogeneous association.
- 3–way association.

We’ll take a look at examples of situations 2, 3 and 4 and each of these cases of conditional dependence.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or “Partial” Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence: *CAL*

● *CAL* Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson’s Paradox

● (Hypothetical) Example of Simpson’s Paradox

● Picture of Simpson’s Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# Marginal Independence/Conditional Dependence

## Marginal Table

Group	Response		
	yes	no	
A	30	30	60
B	30	30	60
	60	60	120

$$\theta = 1$$

$$\log(\theta) = 0$$

## Partial Tables:

$Z = 1$

Group	Response		
	yes	no	
A	5	15	20
B	15	5	20
	20	20	40

$\theta = 1/9$   
 $\log(\theta) = -2.197$

$Z = 2$

Group	Response		
	yes	no	
A	10	10	20
B	10	10	20
	20	20	40

$\theta = 1$   
 $\log(\theta) = 0$

$Z = 3$

Group	Response		
	yes	no	
A	15	5	20
B	5	15	20
	20	20	40

$\theta = 9$   
 $\log(\theta) = 2.197$

Association is in opposite directions in tables  $Z = 1$  and  $Z = 3$ .

# Marginal Dependence/Conditional Independence

or just “*Conditional Independence*”

- This situation and concept is not unique to categorical data analysis.
- Conditional independence is very important and is the basis for many models and techniques including
  - ◆ Latent variable models (e.g., factor analysis, latent class analysis, item response theory, etc.).
  - ◆ Multivariate Graphical models, which provide ways to decompose models and problems into sub-problems.
- Back to categorical data. . . .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or “Partial” Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence:  $CAL$

●  $CAL$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# Conditional Independence

Hypothetical Example from Agresti, 1990:

Marginal Table:

Boy Scout	Delinquent		
	Yes	No	
Yes	36	364	400
No	60	340	400
	96	704	800

$$\hat{\theta} = .56$$

$$G^2 = 6.882$$

$$p\text{-value} = .01$$

Partial Tables — condition on socioeconomic status

SES = Low

Boy Scout	Delinquent		$\hat{\theta} = 1.00$
	Yes	No	
Yes	10	40	50
No	40	160	200
	50	200	250

SES = Medium

Boy Scout	Delinquent		$\hat{\theta} = 1.00$
	Yes	No	
Yes	18	132	150
No	18	132	150
	36	264	300

SES = High

Boy Scout	Delinquent		$\hat{\theta} = 1.00$
	Yes	No	
Yes	8	192	200
No	2	48	50
	10	240	250

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence: *CAL*

● *CAL* Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association Three-Way Tables

(continued)

# Example of Conditional Independence: $CA\mathcal{L}$

■ University of California, Berkeley Graduate Admissions (1973). Data from Freedman, Pisani, & Purves (1978).

■ Question: Is there sex discrimination in admission to graduate school?

■ The data for two departments (B & C) of the 6 largest are

Gender	Admitted		
	Yes	No	
Female	219	399	618
Male	473	412	885
	692	811	1503

$$\hat{\theta} = .48$$

$$1/\hat{\theta} = 2.09$$

$$95\% \text{ CI: } (.39, .59)$$

$$\text{odds(female admitted)} = 219/399 = .55$$

$$\text{odds(male admitted)} = 473/412 = 1.15$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional

Independence:  $CA\mathcal{L}$

●  $CA\mathcal{L}$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables  
(continued)

# CAL Admissions Data by Department

## Department B:

Gender	Admitted		
	Yes	No	
Female	17	8	25
Male	353	207	560
	370	215	585

$$\hat{\theta} = 1.25$$

$$95\% \text{ CI: } (.53, 2.94)$$

## Department C:

Gender	Admitted		
	Yes	No	
Female	202	391	593
Male	120	205	325
	322	215	918

$$\hat{\theta} = .88$$

$$95\% \text{ CI: } (.67, 1.17)$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

- Conditional or "Partial" Odds Ratios

- Marginal Odds Ratios

- Example of Marginal vs Partial Odds Ratios

- Example: Partial Tables

- Marginal and Conditional Associations

- Four Situations

- Marginal

Independence/Conditional Dependence

- Marginal

Dependence/Conditional Independence

- Conditional Independence

- Example of Conditional Independence: CAL

- CAL Admissions Data by Department

- 3rd Example of Conditional Independence

- 3rd Example: Partial Tables

- Simpson's Paradox

- (Hypothetical) Example of Simpson's Paradox

- Picture of Simpson's Paradox

- Homogeneous Association

- Homogeneous Association

Three-Way Tables  
(continued)

# 3rd Example of Conditional Independence

... Maybe conditional independence... Job satisfaction (Andersen, 1985). These data are from a large scale investigation of blue collar workers in Denmark (1968).

Three variables:

- Worker job satisfaction (Low, High).
- Supervisor job satisfaction (Low, High).
- Quality of Management (Bad, Good).

The Worker  $\times$  Supervisor Job Satisfaction (Marginal Table):

Supervisor satisfaction	Worker satisfaction		
	Low	High	
Low	162	196	358
High	110	247	357
	272	443	715

$$\hat{\theta} = 1.86, \quad 95\% \text{ CI } (1.37, 2.52)$$

Statistics	df	Value	Prob
$X^2$	1	17.00	< .001
$G^2$	1	17.19	< .001

- Overview

---

- Marginal and Partial Tables

---

- Conditional and Marginal Odds Ratios

  - Conditional or "Partial" Odds Ratios
  - Marginal Odds Ratios
  - Example of Marginal vs Partial Odds Ratios
  - Example: Partial Tables
  - Marginal and Conditional Associations
  - Four Situations
  - Marginal

- Independence/Conditional Dependence

  - Marginal

- Dependence/Conditional Independence

  - Conditional Independence
  - Example of Conditional Independence:  $CAL$
  - $CAL$  Admissions Data by Department
  - 3rd Example of Conditional Independence
  - 3rd Example: Partial Tables
  - Simpson's Paradox
  - (Hypothetical) Example of Simpson's Paradox
  - Picture of Simpson's Paradox

- Homogeneous Association

  - Homogeneous Association

- Three-Way Tables

(continued)

# 3rd Example: Partial Tables

Job satisfaction conditional on management quality

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

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

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

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

We'll come back to this example....

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence: *CAL*

● *CAL* Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# Simpson's Paradox

The marginal association is in the opposite direction as the conditional (or partial) association.

Consider 3 dichotomous variables:  $X$ ,  $Y$ , and  $Z$  where

- $P(Y = 1|X = 1) =$  conditional probability  $Y = 1$  given  $X = 1$ ,
- $P(Y = 1|X = 1, Z = 1) =$  conditional probability  $Y = 1$  given  $X = 1$  and  $Z = 1$ .

- Simpson's Paradox:

$$\text{Marginal: } P(Y = 1|X = 1) < P(Y = 1|X = 2)$$

$$\text{Conditionals: } P(Y = 1|X = 1, Z = 1) > P(Y = 1|X = 2, Z = 1)$$

$$P(Y = 1|X = 1, Z = 2) > P(Y = 1|X = 2, Z = 2)$$

- In terms of odds ratios, it is possible to observed the following pattern of marginal and partial associations:

Marginal odds:  $\theta_{XY} < 1$ ; however, Partial odds:  $\theta_{XY(1)} > 1$  and  $\theta_{XY(2)} > 1$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional

Independence:  $\mathcal{CAL}$

●  $\mathcal{CAL}$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# (Hypothetical) Example of Simpson's Paradox

$Z = 1$			$Z = 2$			
	$Y = 1$	$Y = 2$		$Y = 1$	$Y = 2$	
$X = 1$	50	900	950	500	5	505
$X = 2$	1	100	101	500	95	595
	51	1000	1051	1000	100	1100

$$\theta_{XY(z=1)} = 5.56 \quad \text{and} \quad \theta_{XY(z=2)} = 19.0$$

$$\pi_1(x=1, z=1) = 50/950 = .05 \quad \text{and} \quad \pi_1(x=1, z=2) = 500/505 = .99$$

$$\pi_2(x=2, z=1) = 1/101 = .01 \quad \text{and} \quad \pi_2(x=2, z=2) = 500/595 = .84$$

The  $XY$  margin:

	$Y = 1$	$Y = 2$	
$X = 1$	550	905	1455
$X = 2$	501	195	696
	1051	1100	2151

$$\theta_{XY} = .237$$

$$\pi_1 = 550/1455 = .38$$

$$\pi_2 = 501/696 = .72$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence:  $\mathcal{CAL}$

●  $\mathcal{CAL}$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

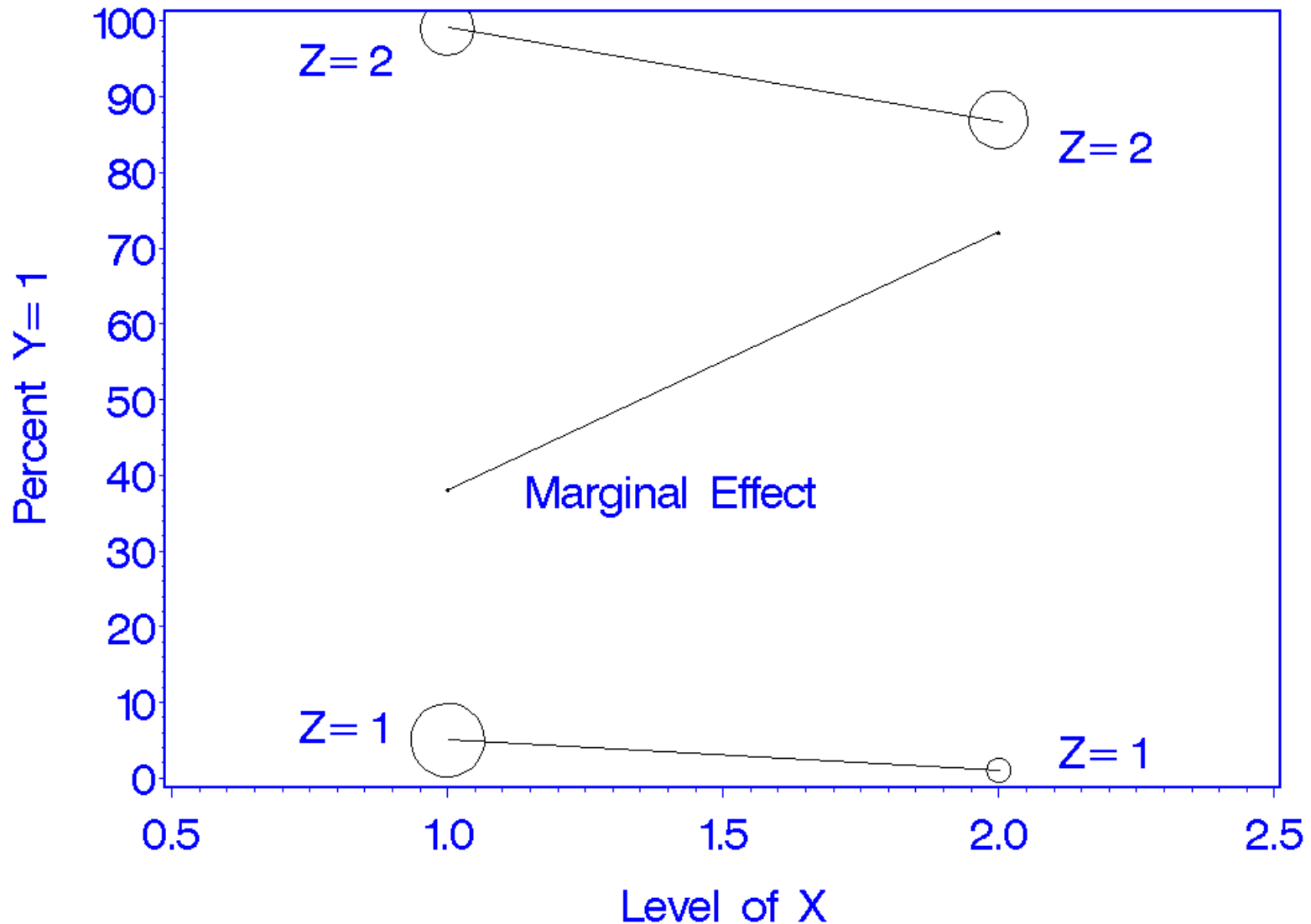
● Homogeneous Association Three-Way Tables

(continued)

# Picture of Simpson's Paradox

## Example of Simpson's Paradox

Area = number of observations with  $X=i$  &  $Z=k$



Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence:  $CAL$

●  $CAL$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# Homogeneous Association

Definition: The association between variables  $X$ ,  $Y$ , and  $Z$  is “homogeneous” if the following three conditions hold:

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

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

$$\theta_{YZ(1)} = \dots = \theta_{YZ(i)} = \dots = \theta_{YZ(I)}$$

- There is “no interaction between any 2 variables in their effects on the third variable”.
- There is “no 3-way interaction” among the variables.
- If one of the above holds, then the other two will also hold.
- Conditional independence is a special case of this.  
For example,

$$\theta_{YZ(1)} = \dots = \theta_{YZ(i)} = \dots = \theta_{YZ(I)} = 1$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or “Partial” Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional Independence:  $\mathcal{CAL}$

●  $\mathcal{CAL}$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

# Homogeneous Association (continued)

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

● Conditional or "Partial" Odds Ratios

● Marginal Odds Ratios

● Example of Marginal vs Partial Odds Ratios

● Example: Partial Tables

● Marginal and Conditional Associations

● Four Situations

● Marginal

Independence/Conditional Dependence

● Marginal

Dependence/Conditional Independence

● Conditional Independence

● Example of Conditional

Independence:  $CAL$

●  $CAL$  Admissions Data by Department

● 3rd Example of Conditional Independence

● 3rd Example: Partial Tables

● Simpson's Paradox

● (Hypothetical) Example of Simpson's Paradox

● Picture of Simpson's Paradox

● Homogeneous Association

● Homogeneous Association

Three-Way Tables

(continued)

- There are even simpler independence conditions that are special cases of homogeneous association, but this is a topic for another day.
- When these three conditions (equations) do **not** hold, then the conditional odds ratios for any pair of variables are not equal. Conditional odds ratios differ/depend on the level of the third variable.
- Example of 3-way Interaction — the Age  $\times$  Smoking  $\times$  Breath test results example.

# Example of Homogeneous Association

Attitude Toward Media (Fienberg, 1980). “Are radio and TV networks doing a good, fair, or poor job?”

Year	Race	Response		
		Good	Fair	Poor
1959	Black	81	23	4
	White	325	243	54
1971	Black	224	144	24
	White	600	636	158

$$\hat{\theta}_{RQ1(1959)} = (81)(243)/(325)(23) = 2.63$$

$$\hat{\theta}_{RQ1(1971)} = (224)(636)/(600)(144) = 1.6$$

$$\hat{\theta}_{RQ2(1959)} = (23)(54)/(243)(4) = 1.28$$

$$\hat{\theta}_{RQ2(1971)} = (144)(158)/(636)(24) = 1.49$$

$$\hat{\theta}_{YR(good)} = (81)(600)/(325)(224) = .68$$

$$\hat{\theta}_{YR(fair)} = (23)(636)/(243)(144) = .42$$

$$\hat{\theta}_{YR(poor)} = (4)(158)/(54)(24) = .48$$

$$\hat{\theta}_{YQ1(black)} = (81)(144)/(23)(224) = 2.26$$

$$\hat{\theta}_{YQ1(white)} = (325)(636)/(600)(243) = 1.4$$

$$\hat{\theta}_{YQ2(black)} = (23)(24)/(4)(144) = .96$$

$$\hat{\theta}_{YQ2(white)} = (243)(158)/(54)(646) = 1.10$$

## Overview

### Marginal and Partial Tables

### Conditional and Marginal Odds Ratios

- Conditional or “Partial” Odds Ratios

### ● Marginal Odds Ratios

- Example of Marginal vs Partial Odds Ratios

- Example: Partial Tables
- Marginal and Conditional Associations

- Four Situations
- Marginal

### Independence/Conditional Dependence

- Marginal

### Dependence/Conditional Independence

- Conditional Independence
- Example of Conditional Independence:  $CAL$

- $CAL$  Admissions Data by Department

- 3rd Example of Conditional Independence

- 3rd Example: Partial Tables

- Simpson's Paradox

- (Hypothetical) Example of Simpson's Paradox

- Picture of Simpson's Paradox

- Homogeneous Association

- Homogeneous Association Three-Way Tables

(continued)

# Statistical Inference & 3–Way Tables

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3–Way Tables

● Statistical Inference & 3–Way Tables

● Sampling Models for 3–Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

(Large samples)

We'll focus methods for  $2 \times 2 \times K$  tables.

- Sampling Models for 3–Way tables.
- Test of conditional independence.
- Estimating common odds ratio.
- Test of homogeneous association.
- Further Comments

# Sampling Models for 3–Way Tables

Generalizations of the ones for 2–way tables, but there are now more possibilities.

Possible Sampling Models for 3–Way tables:

- **Independent Poisson** variates — nothing fixed, each cell is Poisson.
- **Multinomial counts** with only the overall total  $n$  is fixed.
- Multinomial counts with fixed sample size for each partial table.

For example, the partial tables of  $X \times Y$  for each level of  $Z$ , only the total of each of the  $K$  tables is fixed,  $n_{++k}$ .

- **Independent binomial (or multinomial) samples** within each partial table.

For example, if  $n_{1+k}$  and  $n_{2+k}$  are fixed in each  $2 \times 2$  partial table of  $X$  crossed with  $Y$  for  $k = 1, \dots, K$  levels of  $Z$ , then we have independent binomial samples within each partial table.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3–Way Tables

● Statistical Inference & 3–Way Tables  
● Sampling Models for 3–Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

# Tests of Conditional Independence

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence
- Cochran-Mantel-Haenszel Test
- Idea Behind the CMH Test
- Statistical Hypotheses
- CMH Test Statistic
- Properties of the CMH Test Statistic
- More Properties of the CMH Test Statistic
- Age × Smoking × Breath test results
- CMH Statistic for Age × Smoking × Breath
- CMH Example: CAL graduate admissions
- Example: Table × Group × Response
- Management × Supervisor × Worker
- Management × Supervisor × Worker

Two methods:

- Sum of test statistics for independence in each of the partial tables to get an overall chi-squared statistic for “conditional independence” — this is the equivalent to a model based test discussed later in course.
- Cochran-Mantel-Haenszel Test — we’ll talk about this one first.

(continued)

# Cochran-Mantel-Haenszel Test

Example: Cal graduate admission data

- $X$ : Gender (female, male).
- $Y$ : Admission to graduate school (admitted, denied).
- $Z$ : Department to which person applied (6 largest ones, A–F).

So we have a  $2 \times 2 \times 6$  table of Gender by Admission by Department.

For each Gender by Admission partial table (condition on department), if we take the row totals ( $n_{1+k}$  and  $n_{2+k}$ ) and the column totals ( $n_{+1k}$  and  $n_{+2k}$ ) as fixed, then once we know the value of a single cell within the table, we can fill in the rest of the table. For department  $A$ :

Gender	Admitted?		
	Yes	No	
Female	89	(19)	108
Male	(512)	(313)	825
	601	332	933

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

● Tests of Conditional Independence

● Cochran-Mantel-Haenszel Test

● Idea Behind the CMH Test

● Statistical Hypotheses

● CMH Test Statistic

● Properties of the CMH Test Statistic

● More Properties of the CMH Test Statistic

● Age  $\times$  Smoking  $\times$  Breath test results

● CMH Statistic for Age  $\times$  Smoking  $\times$  Breath

● CMH Example: CAL graduate admissions

● Example: Table  $\times$  Group  $\times$  Response

● Management  $\times$  Supervisor  $\times$  Worker

● Management  $\times$  Supervisor  $\times$  Worker

(continued)

# Idea Behind the CMH Test

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

● Tests of Conditional Independence

● Cochran-Mantel-Haenszel Test

● Idea Behind the CMH Test

● Statistical Hypotheses

● CMH Test Statistic

● Properties of the CMH Test Statistic

● More Properties of the CMH Test Statistic

● Age  $\times$  Smoking  $\times$  Breath test results

● CMH Statistic for Age  $\times$  Smoking  $\times$  Breath

● CMH Example: CAL graduate admissions

● Example: Table  $\times$  Group  $\times$  Response

● Management  $\times$  Supervisor  $\times$  Worker

● Management  $\times$  Supervisor  $\times$  Worker

(continued)

- From discussion of Fisher's exact test, we know that the distribution of  $2 \times 2$  tables with fixed margins is hypergeometric.
- Regardless of sampling scheme, if we consider row and column totals of partial tables as fixed, we can use hypergeometric distribution to compute probabilities.
- The test for conditional association uses one cell from each partial table.
- Historical Note: In developing this test, Mantel and Haenszel were concerned with analyzing retrospective studies of diseases ( $Y$ ). They wanted to compare two groups ( $X$ ) and adjust for a control variable ( $Z$ ). Even though only 1 margin of the data (disease margin,  $Y$ ) is fixed, they analyzed data by conditioning on both the outcome ( $Y$ ) and group margins ( $X$ ) for each level of the control variable ( $Z$ ).

# Statistical Hypotheses

If the null hypothesis of conditional independence is true, i.e.,

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

Then the mean of the (1,1) cell of  $k$ th partial table is

$$\mu_{11k} = E(n_{11k}) = \hat{\mu}_{11k} = n_{++k} \hat{\pi}_{1+k} \hat{\pi}_{+1k} = \frac{n_{1+k} n_{+1k}}{n_{++k}}$$

and the variance of the (1,1) cell of the  $k$ th partial table is

$$\widehat{\text{Var}}(n_{11k}) = \frac{n_{1+k} n_{2+k} n_{+1k} n_{+2k}}{n_{++k}^2 (n_{++k} - 1)}$$

If the null is false, then we expect that for tables where

- $\theta_{XY(k)} > 1 \implies (n_{11k} - \mu_{11k}) > 0$
- $\theta_{XY(k)} < 1 \implies (n_{11k} - \mu_{11k}) < 0$
- $\theta_{XY(k)} = 1 \implies (n_{11k} - \mu_{11k}) \approx 0$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

● Tests of Conditional Independence

● Cochran-Mantel-Haenszel Test

● Idea Behind the CMH Test

● Statistical Hypotheses

● CMH Test Statistic

● Properties of the CMH Test Statistic

● More Properties of the CMH Test Statistic

● Age × Smoking × Breath

test results

● CMH Statistic for Age × Smoking × Breath

● CMH Example: CAL graduate admissions

● Example: Table × Group × Response

● Management × Supervisor × Worker

● Management × Supervisor × Worker

(continued)

# CMH Test Statistic

Mantel & Haenszel (1959) proposed the following statistic

$$M^2 = \frac{(\sum_k |n_{11k} - \mu_{11k}| - \frac{1}{2})^2}{\sum_k \text{Var}(n_{11k})}$$

If  $H_o$  is true, then  $M^2$  is approximately chi-squared with  $df = 1$ .

Cochran (1954) proposed a similar statistic, except that

- He did not include the continuity correction, “ $-1/2$ ”.
- He used a different  $\text{Var}(n_{11k})$ .

The statistic the we will use is a combination of these two proposed statistics, the “Cochran-Mantel-Haenszel” statistic

$$CMH = \frac{[\sum_k (n_{11k} - \hat{\mu}_{11k})]^2}{\sum_k \widehat{\text{Var}}(n_{11k})}$$

where

- $\hat{\mu}_{11k} = n_{1+k}n_{+1k}/n_{++k}$
- $\widehat{\text{Var}}(n_{11k}) = n_{1+k}n_{2+k}n_{+1k}n_{+2k}/n_{++k}^2(n_{++k} - 1)$

- Overview

---

- Marginal and Partial Tables

---

- Conditional and Marginal Odds Ratios

---

- Statistical Inference & 3-Way Tables

---

- Tests of Conditional Independence
  - Tests of Conditional Independence
  - Cochran-Mantel-Haenszel Test
  - Idea Behind the CMH Test
  - Statistical Hypotheses
  - CMH Test Statistic
  - Properties of the CMH Test Statistic
  - More Properties of the CMH Test Statistic
  - Age × Smoking × Breath test results
  - CMH Statistic for Age × Smoking × Breath
  - CMH Example: CAL graduate admissions
  - Example: Table × Group × Response
  - Management × Supervisor × Worker
  - Management × Supervisor × Worker

(continued)

# Properties of the CMH Test Statistic

$$CMH = \frac{(\sum_k (n_{11k} - \mu_{11k}))^2}{\sum_k \text{Var}(n_{11k})}$$

- For large samples, when  $H_o$  is true, CMH has a chi-squared distribution with  $df = 1$ .

- If all  $\theta_{XY(k)} = 1$ , then CMH is small (close to 0).

Example: SES  $\times$  Boy Scout  $\times$  Delinquent. Since  $\hat{\theta} = 1$  for each partial table, if we compute  $CMH$ , it would equal 0 and  $p\text{-value}=1.00$ .

- If some/all  $\theta_{XY(k)} > 1$ , then CMH is large.

Example: Age  $\times$  Smoking  $\times$  Breath Test.

Example: CAL graduate admissions data, Departments (6 versus 5)  $\times$  Gender  $\times$  Admission.

- If some/all  $\theta_{XY(k)} < 1$ , then CMH is large.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence

- Cochran-Mantel-Haenszel Test

- Idea Behind the CMH Test

- Statistical Hypotheses

- CMH Test Statistic

- Properties of the CMH Test Statistic

- More Properties of the CMH Test Statistic

- Age  $\times$  Smoking  $\times$  Breath test results

- CMH Statistic for Age  $\times$  Smoking  $\times$  Breath

- CMH Example: CAL graduate admissions

- Example: Table  $\times$  Group  $\times$  Response

- Management  $\times$  Supervisor  $\times$  Worker

- Management  $\times$  Supervisor  $\times$  Worker

(continued)

# More Properties of the CMH Test Statistic

$$CMH = \frac{(\sum_k (n_{11k} - \mu_{11k}))^2}{\sum_k \text{Var}(n_{11k})}$$

- If some  $\theta_{XY(k)} > 1$  and some  $\theta_{XY(k)} < 1$ , *CMH* test is not appropriate.

Example: Three tables of Group  $\times$  Response (hypothetical “DIF” case).

- The test works well and is more powerful when  $\theta_{XY(k)}$ 's are in the same direction and of comparable size.

Example: Management quality  $\times$  Worker satisfaction  $\times$  Supervisor's satisfaction.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence

- Cochran-Mantel-Haenszel Test

- Idea Behind the CMH Test

- Statistical Hypotheses

- CMH Test Statistic

- Properties of the CMH Test Statistic

- More Properties of the CMH Test Statistic

- Age  $\times$  Smoking  $\times$  Breath test results

- CMH Statistic for Age  $\times$  Smoking  $\times$  Breath

- CMH Example: CAL graduate admissions

- Example: Table  $\times$  Group  $\times$  Response

- Management  $\times$  Supervisor  $\times$  Worker

- Management  $\times$  Supervisor  $\times$  Worker

(continued)

# Age × Smoking × Breath test results

Example: These data are from a study reported by Forthofer & Lehnen (1981) (Agresti, 1990). Subjects were whites who work in certain industrial plants in Houston.

Partial Tables:

Smoking Status	Age < 40			Age 40–59		
	Normal	Not Normal		Normal	Not Normal	
Never	577	34	611	164	4	168
Current	682	57	739	245	74	319
	1259	91	1350	409	78	487

Statistical Hypotheses:

$$H_o : \theta_{SB(<40)} = \theta_{SB(40-50)} = 1$$

$H_A$  : Smoking and test results are conditionally dependent.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence

- Cochran-Mantel-Haenszel Test

- Idea Behind the CMH Test

- Statistical Hypotheses

- CMH Test Statistic

- Properties of the CMH Test Statistic

- More Properties of the CMH Test Statistic

- Age × Smoking × Breath test results

- CMH Statistic for Age × Smoking × Breath

- CMH Example: CAL graduate admissions

- Example: Table × Group × Response

- Management × Supervisor × Worker

- Management × Supervisor × Worker

(continued)

# CMH Statistic for Age × Smoking × Breath

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

● Tests of Conditional Independence

● Cochran-Mantel-Haenszel Test

● Idea Behind the CMH Test

● Statistical Hypotheses

● CMH Test Statistic

● Properties of the CMH Test Statistic

● More Properties of the CMH Test Statistic

● Age × Smoking × Breath

test results

● CMH Statistic for Age × Smoking × Breath

● CMH Example: CAL graduate admissions

● Example: Table × Group × Response

● Management × Supervisor × Worker

● Management × Supervisor × Worker

(continued)

Age < 40

$$\hat{\theta}_1 = 1.418$$

$$\hat{\mu}_{111} = (611)(1259)/1350 = 569.81$$

$$n_{111} - \hat{\mu}_{111} = 577 - 569.81 = 7.19$$

$$\widehat{\text{var}}(n_{111}) = \frac{(611)(739)(1259)(91)}{1350^2(1350-1)} = 21.04$$

Age 40–59

$$\hat{\theta}_2 = 12.38$$

$$\hat{\mu}_{112} = (168)(409)/487 = 141.09$$

$$n_{112} - \hat{\mu}_{112} = 164 - 141.09 = 22.91$$

$$\widehat{\text{var}}(n_{112}) = \frac{(168)(319)(409)(78)}{487^2(487-1)} = 14.83$$

$$\begin{aligned} CMH &= \frac{(7.19 + 22.91)^2}{21.04 + 14.83} \\ &= 24.24 \end{aligned}$$

with  $df = 1$  has  $p$ -value  $< .001$ .

# CMH Example: CAL graduate admissions

The null hypothesis of no sex discrimination is

$$\theta_{GA(1)} = \theta_{GA(2)} = \theta_{GA(3)} = \theta_{GA(4)} = \theta_{GA(5)} = \theta_{GA(6)} = 1$$

Department A				Department B			Department C		
Gender	admit	deny		admit	deny		admit	deny	
female	89	19	108	17	8	25	202	391	593
male	512	313	825	353	207	560	120	205	325
	601	332	933	370	215	585	322	596	918

Department D				Department E			Department F		
Gender	admit	deny		admit	deny		admit	deny	
female	131	244	375	94	299	393	24	317	341
male	138	279	417	53	138	191	22	351	373
	269	523	792	147	437	584	46	668	714

$$\begin{aligned}
 CMH &= \frac{(19.42 + 1.19 - 6.00 + 3.63 - 4.92 + 2.03)^2}{21.25 + 5.57 + 47.86 + 44.34 + 24.25 + 10.75} \\
 &= (15.36)^2 / 154.02 \\
 &= 1.53 \qquad (p\text{-value} = .217)
 \end{aligned}$$

Department A:  $\hat{\theta}_A = 2.86$ ,  $G^2 = 17.248$ ,  $df = 1$ ,  $p\text{-value} < .001$ .

Without Department A:  $CMH = .125$ ,  $p\text{-value} = .724$ .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence
- Cochran-Mantel-Haenszel Test
- Idea Behind the CMH Test
- Statistical Hypotheses
- CMH Test Statistic
- Properties of the CMH Test Statistic
- More Properties of the CMH Test Statistic
- Age × Smoking × Breath test results
- CMH Statistic for Age × Smoking × Breath
- CMH Example: CAL graduate admissions
- Example: Table × Group × Response
- Management × Supervisor × Worker
- Management × Supervisor × Worker

(continued)

# Example: Table × Group × Response

(Hypothetical DIF data)

$Z = 1$

Group	yes	no	
A	5	15	20
B	15	5	20
	20	20	40

$\theta = 0.11$

$Z = 2$

Group	yes	no	
A	10	10	20
B	10	10	20
	20	20	40

$\theta = 1.00$

$Z = 3$

Group	yes	no	
A	15	5	20
B	5	15	20
	20	20	40

$\theta = 9.00$

$$\begin{aligned}
 CMH &= \frac{((5 - 10) + (10 - 10) + (15 - 10))^2}{\sum_{k=1}^3 \text{Var}(n_{11k})} \\
 &= \frac{(-5 + 0 + 5)^2}{\sum_{k=1}^3 \text{Var}(n_{11k})} \\
 &= 0
 \end{aligned}$$

Why is this test a bad thing to do here?

- Overview
- Marginal and Partial Tables
- Conditional and Marginal Odds Ratios
- Statistical Inference & 3-Way Tables
- Tests of Conditional Independence
  - Tests of Conditional Independence
  - Cochran-Mantel-Haenszel Test
  - Idea Behind the CMH Test
  - Statistical Hypotheses
  - CMH Test Statistic
  - Properties of the CMH Test Statistic
  - More Properties of the CMH Test Statistic
  - Age × Smoking × Breath test results
  - CMH Statistic for Age × Smoking × Breath
  - CMH Example: CAL graduate admissions
  - Example: Table × Group × Response
  - Management × Supervisor × Worker

- Management × Supervisor × Worker

(continued)

# Management × Supervisor × Worker

Bad Management				Good Management			
Supervisor Satisfaction	Worker Job			Supervisor Satisfaction	Worker Job		
	Low	High			Low	High	
Low	103	87	190	Low	59	109	168
High	32	42	74	High	78	205	283
	135	129	264		137	314	448

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

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

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

Note:  $G^2 = 2.57 + 2.82 = 5.39$  with  $df = 2$  has  $p\text{-value} = .068$ .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

- Tests of Conditional Independence

- Cochran-Mantel-Haenszel Test

- Idea Behind the CMH Test

- Statistical Hypotheses

- CMH Test Statistic

- Properties of the CMH Test Statistic

- More Properties of the CMH Test Statistic

- Age × Smoking × Breath

- test results

- CMH Statistic for Age × Smoking × Breath

- CMH Example: CAL graduate admissions

- Example: Table × Group × Response

- Management × Supervisor × Worker

- Management × Supervisor × Worker

(continued)

# Management × Supervisor × Worker (continued)

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

● Tests of Conditional Independence

● Cochran-Mantel-Haenszel Test

● Idea Behind the CMH Test

● Statistical Hypotheses

● CMH Test Statistic

● Properties of the CMH Test Statistic

● More Properties of the CMH Test Statistic

● Age × Smoking × Breath test results

● CMH Statistic for Age × Smoking × Breath

● CMH Example: CAL graduate admissions

● Example: Table × Group × Response

● Management × Supervisor × Worker

● Management × Supervisor × Worker

(continued)

- Combining the results from these two tables to test conditional independence yields  $G^2 = 2.57 + 2.82 = 5.39$  with  $df = 2$  has  $p\text{-value} = .068$ .

- Conclusion:

$H_0$  : Conditional independence,  $\theta_{SW(bad)} = \theta_{SW(good)} = 1$ , is a tenable hypothesis.

- Since  $\hat{\theta}_{bad} \approx \hat{\theta}_{good}$ , CMH should be more powerful.

$$CMH = 5.43$$

$$p\text{-value} = .021$$

- Next steps:

- ◆ Estimate the common odds ratio.
- ◆ Test for homogeneous association.

# Estimating Common Odds Ratio

For a  $2 \times 2$  table where  $\theta_{XY(1)} = \dots = \theta_{XY(K)}$ , the “*Mantel-Haenszel Estimator*” of a common value of the odds ratio is

$$\hat{\theta}_{MH} = \frac{\sum_k (n_{11k}n_{22k}/n_{++k})}{\sum_k (n_{12k}n_{21k}/n_{++k})}$$

For the blue-collar worker example, this value is

$$\begin{aligned}\hat{\theta}_{MH} &= \frac{(103)(42)/264 + (59)(205)/448}{(32)(87)/264 + (78)(109)/448} \\ &= \frac{16.39 + 27.12}{10.55 + 18.98} \\ &= 43.51/29.52 = 1.47\end{aligned}$$

Which is in between the two estimates from the two partial tables:

$$\hat{\theta}_{bad} = 1.55 \quad \text{and} \quad \hat{\theta}_{good} = 1.42$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

- Estimating Common Odds Ratio
- SE for Common Odds Ratio Estimate
- SAS input & Common Odds Ratio Estimate
- Notes Regarding CMH

Testing Homogeneity of Odds Ratios

Concluding comments

# SE for Common Odds Ratio Estimate

For our example,

95% confidence interval for  $\theta \longrightarrow (1.06, 2.04)$

The standard error for  $\hat{\theta}_{MH}$  is complex, so we will rely on SAS/FREQ to get this. When you supply the “CMH” option to the TABLES command, you will get both CMH test statistic and  $\hat{\theta}_{MH}$  along with a 95% confidence interval for  $\theta$ .

SAS output:

Estimates of the Common Relative Risk (Row1/Row2)				
Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	1.4697	1.0600	2.0377
	Logit	1.4692	1.0594	2.0374

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

- Estimating Common Odds Ratio
- SE for Common Odds Ratio Estimate
- SAS input & Common Odds Ratio Estimate
- Notes Regarding CMH

Testing Homogeneity of Odds Ratios

Concluding comments

# SAS input & Common Odds Ratio Estimate

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

- Estimating Common Odds Ratio
- SE for Common Odds Ratio Estimate
- SAS input & Common Odds Ratio Estimate
- Notes Regarding CMH

Testing Homogeneity of Odds Ratios

Concluding comments

DATA sat;

INPUT manager \$ super \$ worker \$ count;

LABEL manager='Quality of management'

super ='Supervisors Satisfaction'

worker='Blue Collar Workers Satisfaction';

DATALINES;

Bad Low Low 103

Bad Low High 87

⋮ ⋮ ⋮ ⋮

Good High Low 78

Good High High 205

PROC FREQ DATA=sat ORDER= data;

WEIGHT count;

TABLES manage\*super\*worker /nopercent norow nocol chisq cmh;

run;

# Notes Regarding CMH

- If we have homogeneous association, i.e.,

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

then  $\hat{\theta}_{MH}$  is useful as an estimate of the this common odds ratio.

- If the odds ratios are not the same but they are at least in the same direction, then  $\hat{\theta}_{MH}$  can be useful as a summary statistic of the  $K$  conditional (partial) associations.
- If there's a 3-way interaction, it is misleading to use an estimate of the common odds ratio. e.g., Age  $\times$  Smoking  $\times$  Breath test results, we get as a common estimate of the odds ratio

$$\hat{\theta}_{SB} = 2.57$$

But the ones from the separate tables are

$$\hat{\theta}_{SB(<40)} = 1.42 \quad \text{and} \quad \hat{\theta}_{SB(40-59)} = 12.38$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

- Estimating Common Odds Ratio
- SE for Common Odds Ratio Estimate
- SAS input & Common Odds Ratio Estimate
- Notes Regarding CMH

Testing Homogeneity of Odds Ratios

Concluding comments

# Testing Homogeneity of Odds Ratios

- For  $2 \times 2 \times K$  tables.
- Since  $\theta_{XY(1)} = \dots = \theta_{XY(K)}$  implies both

$$\theta_{YZ(1)} = \dots = \theta_{YZ(I)} \quad \text{and} \quad \theta_{XZ(1)} = \dots = \theta_{XZ(J)}$$

To test for homogeneous association we only need to test one of these, e.g.

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

- Given estimated expected frequencies assuming that  $H_O$  is true, the test statistic we use is the “**Breslow-Day**” statistic, which is like Pearson’s  $X^2$ :

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

- If  $H_O$  is true, then the Breslow-Day statistic has an approximate chi-squared distribution with  $df = K - 1$ .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments

Three-Way Tables

# Breslow-Day statistic

- We need  $\hat{\mu}_{ijk}$  for each table assuming that the null hypothesis of homogeneous association is true.
- $\{\hat{\mu}_{11k}, \hat{\mu}_{12k}, \hat{\mu}_{21k}, \hat{\mu}_{22k}\}$ , are found such that
- The margins of the table of estimated expected frequencies equal the observed margins; that is,

$\hat{\mu}_{11k}$	$\hat{\mu}_{12k}$	$(\hat{\mu}_{11k} + \hat{\mu}_{12k}) = n_{1+k}$
$\hat{\mu}_{21k}$	$\hat{\mu}_{22k}$	$(\hat{\mu}_{21k} + \hat{\mu}_{22k}) = n_{2+k}$
$n_{+1k}$	$n_{+2k}$	$n_{++k}$

- If the null hypothesis of homogeneous association is true, then  $\hat{\theta}_{MH}$  is a good estimate of the common odds ratio. When computing estimated expected frequencies, we want them such that the odds ratio computed on each of the  $K$  partial tables equals the Mantel-Haenszel estimate of the common odds ratio.

$$\hat{\theta}_{MH} = \frac{\hat{\mu}_{11k} \hat{\mu}_{22k}}{\hat{\mu}_{12k} \hat{\mu}_{21k}}$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments  
Three-Way Tables

# Breslow-Day statistic

- Computation of the estimated expected frequencies is a bit complex, so we will rely on SAS/FREQ to give us the Breslow-Day statistic. If you have a  $2 \times 2 \times K$  table and request “CMH” options with the TABLES command, you will automatically get the Breslow-Day statistic.

- SAS output for manager  $\times$  supervisor  $\times$  worker is

Breslow-Day Test for  
Homogeneity of the Odds Ratios

---

Chi-Square	0.0649
DF	1
Pr > ChiSq	0.7989

- For this test, your sample size should be relatively large, i.e.,

$$\hat{\mu}_{ijk} \geq 5 \quad \text{for at least 80\% of cells}$$

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments

Three-Way Tables

# Examples: Testing Homogeneity of Association

Worker  $\times$  Supervisor  $\times$  Management

■  $CMH = 5.34$  with  $p\text{-value} = .02 \implies$   
conditionally dependent.

■ The Mantel-Haenszel estimate of common odds ratio

$$\hat{\theta}_{MH} = 1.47$$

while the separate ones were

$$\hat{\theta}_{bad} = 1.55 \quad \text{and} \quad \hat{\theta}_{good} = 1.42$$

■ Now let's test the homogeneity of the odds ratios

$$H_0 : \theta_{WS(bad)} = \theta_{WS(good)}.$$

Breslow-Day statistic = .065,  $df = 1$ , and  $p\text{-value} = .80$ .

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments  
Three-Way Tables

# Cal Graduate Admissions data

Six of the largest departments:

- $CMH = 1.53, df = 1, p\text{-value} = .217 \implies$   
gender and admission are conditionally independent  
(given department).

- Mantel-Haenszel estimate of the common odds ratio

$$\hat{\theta}_{GA} = .91$$

and the 95% Confidence interval is

$$(.772, 1.061).$$

- Now let's test homogeneity of odds ratios

$$H_o : \theta_{GA(a)} = \theta_{GA(b)} = \theta_{GA(c)} = \theta_{GA(d)} = \theta_{GA(e)} = \theta_{GA(f)}$$

Breslow-Day statistic = 18.826,  $df = 5, p\text{-value} = .002.$

What's going on?

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments  
Three-Way Tables

# Cal Graduate Admissions data

Drop Department A, which is the only department for which the odds ratio appears to differ from 1.

- $CMH = .125, df = 1, p\text{-value} = .724 \implies$   
gender and admission are conditionally independent  
(given department)

- The Mantel-Haneszel estimate of the common odds ratio

$$\hat{\theta} = 1.031$$

and the 95% confidence interval for  $\theta_{GA}$  is

$$(.870, 1.211)$$

- The test of homogeneity of odds ratios

$$H_0 : \theta_{GA(b)} = \theta_{GA(c)} = \theta_{GA(d)} = \theta_{GA(e)} = \theta_{GA(f)}$$

Breslow-Day statistic = 2.558,  $df = 4, p\text{-value} = .63$ .

Conclusion?.

# Group × Response × Z

(Hypothetical DIF Example)

- $CMH = 0.00$ ,  $df = 1$ , and  $p\text{-value} = 1.00 \implies$   
Group and response are independent given  $Z$

- Mantel-Haenszel estimate of the common odds ratio

$$\hat{\theta}_{GR} = 1.00$$

- Test for homogeneity of the odds ratios yields  
Breslow-Day statistic = 20.00,  $df = 2$ , and  $p\text{-value} < .001$ .

- Conclusion?

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group × Response × Z

● Year × Race × Response to Question

● Year × Race × Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments

Three-Way Tables

# Year × Race × Response to Question

Response to question “Are radio and TV networks doing a good, fair, or poor job?”

		Response		
Year	Race	Good	Fair	Poor
1959	Black	81	23	4
	White	325	243	54
1971	Black	224	144	24
	White	600	636	158

We could test for conditional independence, but which variable should be condition on?

- Year and look at Race × Response to the Question?
- Race and look at Year × Response to the Question?
- Response to the Question and look at Year × Race?

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group × Response × Z

● Year × Race × Response to Question

● Year × Race × Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments

Three-Way Tables

# Year × Race × Response to Question

- Since the Breslow-Day statistic only works for  $2 \times 2 \times K$  tables, to test for homogeneous association we will set up the test for

$$H_0 : \theta_{YR(\text{good})} = \theta_{YR(\text{fair})} = \theta_{YR(\text{poor})}$$

even though we are more interested in the odds ratios between Year & Response and Race & Response.

- Breslow-Day statistic = 3.464,  $df = 2$ ,  $p\text{-value} = .18$ .

Note: There is a generalization of CMH for  $I \times J \times K$  tables and we can get an estimate of the common odds ratio between Year and Race (i.e.,  $\hat{\theta}_{MH} = .57$ ), what we'd really like are estimates of common odds ratios between Year and Question and between Race and Question.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association  
● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group × Response × Z

● Year × Race × Response to Question

● Year × Race × Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments  
Three-Way Tables

# One Last Example: Hurricane Katrina

Reference: <http://www.bls.gov/katrina/cpscesquestions.htm>  
The effects of hurricane Katrina on BLS employment and unemployment data collection.

- Employment status (employed, unemployed, not in labor force).
- Residence (same or different than in August).
- Month data from (October, November)

The data (in thousands):

	October		November	
	Same	Different	Same	Different
Employed	153	179	204	185
Unemployed	18	90	29	71
Not in labor	134	217	209	188

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

● Testing Homogeneity of Odds Ratios

● Breslow-Day statistic

● Breslow-Day statistic

● Examples: Testing

Homogeneity of Association

● Cal Graduate Admissions data

● Cal Graduate Admissions data

● Group  $\times$  Response  $\times$  Z

● Year  $\times$  Race  $\times$  Response to Question

● Year  $\times$  Race  $\times$  Response to Question

● One Last Example: Hurricane

Katrina

Concluding comments

Three-Way Tables

# Concluding comments on use & interpretation of

## CMH & Breslow-Day

- There is a generalization of CMH for  $I \times J \times K$  tables (which SAS/FREQ will perform).
- There is not such a generalization for the Breslow-Day statistic.
- Given that we can get a non-significant result using CMH when there is association in partial tables, you should check to see whether there is homogeneous association or a 3-way association.
- Breslow-Day statistic does not work well for small samples, while the Cochern-Mantel-Haenszel does pretty well.
- A modeling approach handles  $I \times J \times K$  tables and can test the same hypotheses.

Overview

Marginal and Partial Tables

Conditional and Marginal Odds Ratios

Statistical Inference & 3-Way Tables

Tests of Conditional Independence

Estimating Common Odds Ratio

Testing Homogeneity of Odds Ratios

Concluding comments

● Concluding comments on use & interpretation of