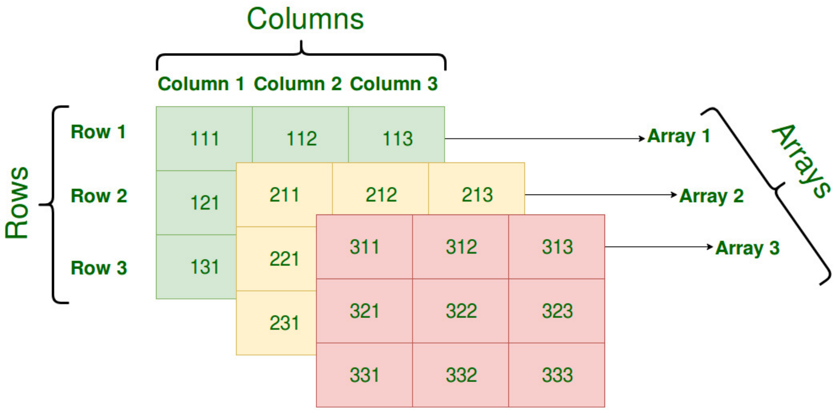
Contingency Tables

Three-way

Demetris Athienitis

Three-way contingency tables

- Three random variables X, Y, and Z
- There are new quantities of interest in this setting
  - Conditional associations
  - Odds ratio between X and Y at a particular value of Z
- Conditional associations can be different from marginal associations between X and Y
- Commonly interest is in the association between X and Y after controlling for (conditioning on) Z


Columns
Column 1
Column 2
Column 3
Rows
Row 1
Row 2
Row 3
111
112
113
121
131
211
212
213
221
231
311
312
313
321
322
323
331
332
333
Array 1
Array 2
Array 3
Arrays

Example

A $2 \times 2 \times 2$ table from data from Florida 1976-1987

Victim's Race	Defendant's Race	Death Penalty Yes	Death Penalty No	Percentage Yes
White	White	53	414	11.3
	Black	11	37	22.9
Black	White	0	16	0.0
	Black	4	139	2.8
Total	White	53	430	11.0
	Black	15	176	7.9

- Y be the response whether they receive death penalty
- X be the defendant's race
- Z be the victim's race

Example (continued)

- If we ignore Z and look at the marginal table see get

$$\widehat{\theta}_{XY} = \frac{53 \times 176}{15 \times 430} = 1.45$$

  which indicates that whites are more likely to get the death penalty

- Of interest is $\theta_{XY(1)}$ and $\theta_{XY(2)}$ the odds ratios between X and Y within the two levels of Z
  - $Z =$ white, $\widehat{\theta}_{XY(1)} = \frac{53 \times 37}{414 \times 11} = 0.43$ (0.42 after adding 0.5 to each cell)
  - $Z =$ black, $\widehat{\theta}_{XY(2)} = \frac{0 \times 139}{16 \times 4} = 0$ (0.94 after adding 0.5 to each cell)

- Conditional odds ratios indicate that whites are less likely to get the death penalty

Three-way contingency tables

Definition (Simpson's paradox)

When a marginal association can have different direction from the conditional associations this is called *Simpson's paradox*.

Definition (Conditional Independence)

Variables X and Y are conditionally independent given Z if they are independent in each conditional table.

In a $2 \times 2 \times K$ table this implies that

$$\theta_{XY(1)} = \cdots = \theta_{XY(k)} = 1$$

Example

Clinic	Treatment	Response: Success	Response: Failure	$\hat{\theta}$
1	A	18	12	1.0
	B	12	8	
2	A	2	8	1.0
	B	8	32	
Total	A	20	20	2.0
	B	20	40	

X and Y are conditionally independent, but marginally, are dependent

Definition (Homogeneous Association)

A *homogeneous association* exists if the conditional odds ratios between X and Y are identical at all levels of Z.

Cochran-Mantel-Haenszel Test

In a $2 \times 2 \times K$ table we wish to test

$$\mathrm{H}_0\colon \theta_{XY(1)} = \cdots = \theta_{XY(K)} = 1$$

- The null is that X and Y are conditionally independent given Z
- The alternative is that at least one of the conditional odds ratios is different from 1

Cochran-Mantel-Haenszel Test

The test statistic is

$$CMH = \frac{\left[\sum_{k=1}^{K} (n_{11k} - E(n_{11k}))\right]^2}{\sum_{k=1}^{K} V(n_{11k})} \underset{\text{approx.}}{\overset{H_0}{\sim}} \chi^2_1$$

where under independence,

$$E(n_{11k}) = \frac{n_{1+k} n_{+1k}}{n_{++k}}$$

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

CMH test statistic

- Procedure also provides an estimate of the common odds ratio among the K tables

$$\hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} (n_{11k} n_{22k} / n_{++k})}{\sum_{k=1}^{K} (n_{12k} n_{21k} / n_{++k})}$$

- Though this is only useful when all of the individual tables have similar odds ratios

Remark

The Breslow-Day Test also exists for testing homogeneity of odds ratios, not just for conditional independence.

Example

Consider a $2 \times 2 \times 5$ table

```
> apply(MIOC,3,OR)
       1        2        3        4        5
6.465600 8.859104 1.675303 3.786661 3.810890
```

Five sample odds ratios do not very “drastically”, proceed with CMH

```
> mantelhaen.test(MIOC)
X-squared = 32.793, df = 1, p-value = 1.025e-08

95 percent confidence interval:
 2.426983 6.493688
sample estimates:
common odds ratio
         3.969895
```

We learned

- Conditional odds-ratio
- CHM test for homogeneous association/odds ratio