Stratified Contingency Tables

Session 8

Module 1 Probability & Statistical Inference

The Summer Institutes
DEPARTMENT OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

Overview

1. **2 x 2 Tables**
   - Paired Binary Data
2. **Stratified Tables**
   - Confounding
   - Effect Modification

EXAMPLE

2 x 2 Tables

Epidemiological Applications: Matched Case Control Study

213 subjects with a history of acute myocardial infarction (AMI) were **matched** by age and sex with one of their siblings who did not have a history of AMI. The prevalence of a particular polymorphism was compared between the siblings.

Question 1 Is there an association between the polymorphism prevalence and AMI?

Question 2 If there is an association then what is the magnitude of the effect?

EXAMPLE

2 x 2 Tables

Epidemiological Applications: Matched Case Control Study

Q: Can't we simply use Pearson's χ^2 Test to assess whether this is evidence for an increase in knowledge?

A: NO!!! Pearson's χ^2 test assumes that the columns are **independent** samples. In this design the 213 with AMI are genetically related to the 213 w/o AMI. This is an example of **paired binary data**.

Disease Status

Exposure Status	AMI	no AMI	TOTAL
Carrier	96	87	183
Noncarrier	117	126	243
TOTAL	213	213	426

EXAMPLE

2 x 2 Tables

Epidemiological Applications: Paired Binary Data

For **paired binary data** we display the results as shown in the table.

This analysis explicitly recognizes the heterogeneity of subjects.

The **concordant pairs** (73 and 103) provide no information about the association between AMI and the polymorphism.

🔑 The information regarding the association is in the **discordant pairs,** 14 and 23.

no AMI \ AMI	carrier	non-carrier	TOTAL
carrier	73	14	87
noncarrier	23	103	126
TOTAL	96	117	213

DEFINITION

2 x 2 Tables

Epidemiological Applications: Paired Binary Data

For **paired binary data** we display the results as shown in the table.

This analysis explicitly recognizes the heterogeneity of subjects.

$$p_1 = P(\text{carrier} \mid \text{AMI}) = p_{11} + p_{01}$$

$$p_0 = P(\text{carrier} \mid \text{No AMI}) = p_{11} + p_{10}$$

$$H_0 : p_1 = p_0$$

$$H_A : p_1 \neq p_0$$

🔑 The information for testing these hypotheses is contained in the **discordant pairs** (0,1) and (1,0).

		AMI 1	AMI 0	TOTAL
no AMI	1	n_{11}	n_{10}	
	0	n_{01}	n_{00}	
TOTAL				n

$n_{11}/n =$ $\mathbf{p_{11}}$	$n_{10}/n =$ $\mathbf{p_{10}}$
$n_{01}/n =$ $\mathbf{p_{01}}$	$n_{00}/n =$ $\mathbf{p_{00}}$

DEFINITION

2 x 2 Tables

Epidemiological Applications: McNemar's Test for Paired Binary Data

Under the null hypothesis we expect equal numbers of (0,1) pairs and (1,0) pairs. We can evaluate this hypothesis using or **McNemar's Test for Paired Binary Data**. The **McNemar's chi-squared statistic** is

$$X^2 = \frac{\left(|n_{10} - n_{01}| - 1\right)^2}{n_{10} - n_{01}} \sim \chi^2(1)$$

The **odds ratio** comparing the odds of carrier in those with AMI to odds of carrier in those w/o AMI is estimated by: $\widehat{OR} = \frac{n_{01}}{n_{10}}$

Confidence intervals can be obtained as described in Breslow and Day (1981), section 5.2, or in Armitage and Berry (1987), chapter 16.

Break #1

**Pause the video,
take a break, stretch,
then review relevant exercises
from worksheet.**

Afterwards, continue on!

Image Credit: indg0.com

EXAMPLE

Effect Modification

Stratified Tables

Often, **a third variable** influences the relationship between the two primary measures (e.g., disease and exposure).

Example (right):
Effect of seat belt use on car accident fatality

Seat Belt

Driver	Worn	Not Worn
Dead	10	20
Alive	40	30
TOTAL	50	50
Fatality Rate	10/50 (20%)	20/50 (40%)

DEFINITION

Effect Modification

Stratified Tables

But, suppose we also consider *impact speed*.

How does this affect your inference?

💡 **This is an example of effect modification or interaction.**

- Effects are different in subgroups of a third variable, and the overall effect is intermediate.

		< 40 mph Seat Belt		> 40 mph Seat Belt	
		Worn	Not Worn	Worn	Not Worn
Driver	Dead	3	2	7	18
	Alive	27	18	13	12
	TOTAL	30	20	20	30
	Fatality Rate	3/30 (10%)	2/20 (10%)	7/20 (35%)	18/30 (60%)

EXAMPLE

Effect Modification

Dependence on the effect measure used

🚨 Effect modification depends on the effect measure used!

Table Rate of fractures over 5 years by age and calcium level in drinking water.

> There's evidence of effect modification on the risk ratio scale.
> There's no evidence of effect modification on the risk difference scale.

Calcium Level	Age: 20-35 yrs	Age: 55-80 yrs	Overall (pooled)
High	1.1%	11.0%	7.8%
Low	3.3%	13.2%	10.0%
Risk Ratio	0.33	0.83	0.78
Risk Difference	-2.2%	-2.2%	-2.2%

EXAMPLE

Confounding

Suppose we are interested in the relationship between

lung cancer incidence

and

heavy drinking (defined as ≥ 2 drinks per day)

We conduct a **prospective cohort study** where drinking status is determined at baseline and the cohort is followed for 10 years to determine cancer endpoints.

We also measure **smoking status** at baseline.

EXAMPLE

Confounding

1) Pooled data, not controlling for smoking

Lung Cancer Status	Heavy Drinker: Yes	Heavy Drinker: No	TOTAL
Yes	33	27	60
No	1667	2273	3940
TOTAL	1700	2300	4000

EXAMPLE

Confounding

- A higher proportion of heavy drinkers are smokers (800/1700 vs 200/2300)
- A higher proportion of lung cancer cases are smokers (30/1000 vs 30/3000)
- The comparison of heavy drinkers to not-heavy drinkers is really a comparison of smokers to nonsmokers

2) Stratify by smoking status at baseline

Smokers

Lung Cancer Status	Heavy Drinker: Yes	Heavy Drinker: No	TOTAL
Yes	24	6	30
No	776	194	970
TOTAL	800	200	1000

OR = 1

Nonsmokers

Lung Cancer Status	Heavy Drinker: Yes	Heavy Drinker: No	TOTAL
Yes	9	21	30
No	891	2079	2970
TOTAL	900	2100	3000

OR = 1

DEFINITION

Confounding

🔑 **A confounder is associated with both the disease and exposure and is not in the causal path between disease and exposure**

$$
\begin{tikzcd}
 & C \arrow[dl] \arrow[dr] & \\
E \arrow[rr, dashed] & & D
\end{tikzcd}
$$

An apparent association between E and D is completely explained by C.
C is a confounder.

- The implicit assumption is that we want to know if E "causes" D
- A simple, common example from genetics is the linked gene: we discover a gene which appears to be associated with disease … does it cause the disease or is it merely linked to the true causal gene?

Break #2

Pause the video,
take a break, stretch,
then review relevant exercises from worksheet.

Afterwards, continue on!

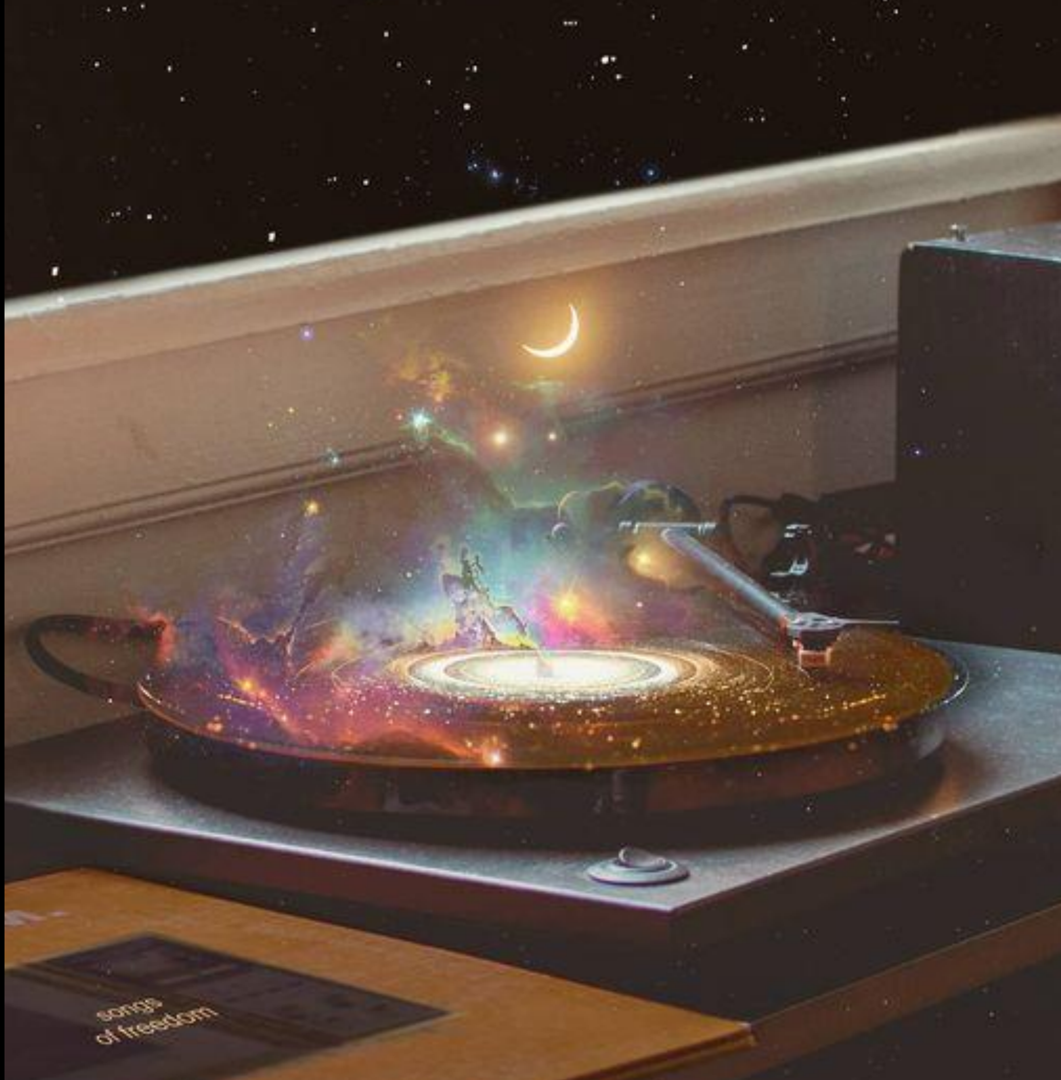

Image Credit: indg0.com

Adjusting the OR via Stratification

Basic idea

- Compute separate OR for each stratum
- Assess homogeneity of OR's across strata
  Is there EM?
- Pool OR's: used weighted average
  Adjust for confounding
- Global test of pooled OR = 1
  Is there association, after adjustment
- Different methods of pooling, testing have been proposed.
  We will focus on Mantel-Haenszel methods
- 👉 Same idea for RR and RD

EXAMPLE

Rosner §13.5

Mantel-Haenszel Methods

A 1985 study identified a group of **509 cancer cases and 489 controls** by mail questionnaire. The main purpose of the study was to look at the **effect of passive smoking on cancer risk**.

In the study **passive smoking** was defined as exposure to the cigarette smoke of a spouse who smoked at least one cigarette/day for at least 6 months.

One **potential confounding variable was smoking by the test subjects themselves** since personal smoking is related to both cancer risk and having a spouse that smokes.

Therefore, it was important to control for **personal smoking** before looking at the relationship between passive smoking and cancer risk.

EXAMPLE

Rosner §13.5

Mantel-Haenszel Methods

1) Pooled data, not controlling for personal smoking

1) Pooled data, not controlling for *personal smoking*

Passive Smoking

Cancer Status	Yes	No	TOTAL
Case	281	228	509
Control	210	279	489
TOTAL	491	507	998

```
. cci 281 228 210 279
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |       281         228  |        509      0.5521
        Controls |       210         279  |        489      0.4294
-----------------+------------------------+------------------------
           Total |       491         507  |        998      0.4920
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         1.637406       |    1.265013    2.119599
 Attr. frac. ex. |         .3892779       |    .2094943    .5282126
 Attr. frac. pop |         .2149059       |
                 +-------------------------------------------------
                               chi2(1) =    15.00  Pr>chi2 = 0.0001
```

For information on how to complete these calculations in R:
https://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/src/biomedicalstats.html

EXAMPLE

Rosner §13.5

Mantel-Haenszel Methods

2) Stratified by *personal smoking*

Personal Smoking: Smokers

Passive Smoking

Cancer Status	Yes	No	TOTAL
Case	161	117	278
Control	130	124	254
TOTAL	291	241	532

OR = 1.31

p-value = 0.1192

Personal Smoking: Nonsmokers

Passive Smoker

Cancer Status	Yes	No	TOTAL
Case	120	111	231
Control	80	155	235
TOTAL	200	266	466

OR = 2.09

p-value = 0.0001

For information on how to complete these calculations in R:
https://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/src/biomedicalstats.html

EXAMPLE

Rosner §13.5

Mantel-Haenszel Methods

2) Stratified by *personal smoking*

Personal Smoking: Smokers

```
. cci 161 117 130 124
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      Exposed
-----------------+------------------------+------------------------
           Cases |       161         117  |        278       0.5791
        Controls |       130         124  |        254       0.5118
-----------------+------------------------+------------------------
           Total |       291         241  |        532       0.5470
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         1.312558       |    .9184614    1.875813
 Attr. frac. ex. |         .2381286       |   -.0887774    .4668978
 Attr. frac. pop |          .137909       |
                 +-------------------------------------------------
                               chi2(1) =     2.43  Pr>chi2 = 0.1192
```

Personal Smoking: Nonsmokers

```
. cci 120 111 80 155
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      Exposed
-----------------+------------------------+------------------------
           Cases |       120         111  |        231       0.5195
        Controls |        80         155  |        235       0.3404
-----------------+------------------------+------------------------
           Total |       200         266  |        466       0.4292
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         2.094595       |     1.41754    3.097165
 Attr. frac. ex. |         .5225806       |    .2945527    .6771241
 Attr. frac. pop |         .2714705       |
                 +-------------------------------------------------
                               chi2(1) =    15.24  Pr>chi2 = 0.0001
```

For information on how to complete these calculations in R:
https://a-little-book-of-r-for-biomedical-statistics.readthedocs.io/en/latest/src/biomedicalstats.html

DEFINITION

Stratified Contingency Tables

Mantel-Haenszel Methods

Q: How can we combine the information from both stratum-specific tables to obtain an overall test of significance that takes account of the stratification?

A: Mantel-Haenszel Methods – assesses association between disease and exposure after controlling for one or more confounding variables.

		Exposure		
		yes	no	TOTAL
Disease	yes	a_i	b_i	a_i+b_i
	no	c_i	d_i	c_i+d_i
	TOTAL	a_i+c_i	b_i+d_i	N_i

where i = 1,2,…,K is the number of strata.

DEFINITION

Stratified Contingency Tables

Mantel-Haenszel Methods

(1) Test of effect modification (heterogeneity, interaction)

H_0: $OR_1 = OR_2 = \ldots = OR_K$

H_A: not all stratum-specific ORs are equal

(2) Estimate the common odds ratio

The Mantel-Haenszel estimate of the odds ratio assumes there is a common odds ratio:

$$OR_{pool} = OR_1 = OR_2 = \ldots = OR_K$$

To estimate the common odds ratio we take a weighted average of the stratum-specific odds ratios:

MH estimate: $$\widehat{OR}_{pool} = \sum_{i=1}^{K} w_i \cdot \widehat{OR}_i$$

(3) Test of common odds ratio

H_0: common odds ratio is 1.0

H_A: common odds ratio $\neq$ 1.0

EXAMPLE

Rosner §13.5

Mantel-Haenszel Methods

```
     +------------------------------------+
     | case   passive   number   smoke |
     |------------------------------------|
  1. |    1         1      120       0 |
  2. |    1         0      111       0 |
  3. |    0         1       80       0 |
  4. |    0         0      155       0 |
  5. |    1         1      161       1 |
  6. |    1         0      117       1 |
  7. |    0         1      130       1 |
  8. |    0         0      124       1 |
     +------------------------------------+
```

Entering the stratum-specific data

```
. cc case passive [freq=number], by(smoke) bd

Personal Smoking |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   2.094595      1.41754   3.097165    19.05579 (exact)
               1 |   1.312558     .9184614   1.875813    28.59023 (exact)
-----------------+-------------------------------------------------
           Crude |   1.637406     1.265013   2.119599             (exact)
    M-H combined |   1.625329     1.263955   2.090024
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     3.27  Pr>chi2 = 0.0706
Test of homogeneity (B-D)      chi2(1) =     3.27  Pr>chi2 = 0.0704

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     14.42
                                                Pr>chi2 =    0.0001
```

Calculating the pooled OR and testing whether it is different from 1

Break #3

Pause the video,
take a break, stretch,
then review relevant exercises
from worksheet.

Afterwards, continue on!

Image Credit: indg0.com