

Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

Chapter 16: Multiple 2×2 Contingency Tables

Jinko Graham

2018-10-06

Associations and extraneous variables

- ▶ Want to study the association between two dichotomous variables: an exposure variable and a disease outcome.
- ▶ However, a third, extraneous variable with g different values (levels) is also present
 - ▶ Does ignoring it lead to incorrect conclusions?
- ▶ When associating the disease outcome and the exposure variable, need to account for the possibility that the extraneous variable is either
 1. an **effect modifier** or
 2. a **confounder**.

Example: Aortic Stenosis (AS) Study

- ▶ What is the association between AS and smoking? How does sex affect this association?
 - ▶ AS: disease outcome
 - ▶ smoking: exposure variable
 - ▶ sex: extraneous variable with $g = 2$ levels.
- ▶ First few lines of the data:

##	smoke	AS	sex
## 1	1	1	1
## 2	1	1	1
## 3	1	1	1
## 4	1	1	1
## 5	1	1	1
## 6	1	1	1

- ▶ When associating AS and smoking, we should account for the possibility that sex is either
 1. an **effect modifier** or
 2. a **confounder**.

Effect Modification

Effect Modification

- ▶ When the ORs for different strata of the extraneous variable are different, the extraneous variable is an **effect modifier**.
- ▶ Inference (estimates, CIs, tests) of these different ORs must then be based on the stratified tables.
- ▶ That is, we estimate the ORs, perform tests, and construct CIs with each stratum-specific table, separately.

Testing for Effect Modification

- ▶ When the ORs for different strata of the extraneous variable are the same, they are said to be **homogeneous**.
- ▶ Can test for this with.
 - ▶ H_0 : homogeneity of ORs, versus
 - ▶ H_a : effect modification.
- ▶ If we reject H_0 , declare the extraneous variable to be an effect modifier.

Test of Homogeneity (Text, Section 16.2.1)

- ▶ Suppose the extraneous variable has g levels or **strata**
- ▶ Let \widehat{OR}_i be the estimated OR between the exposure and the disease outcome in the i th stratum of the extraneous variable.
- ▶ Let $y_i = \log_e \widehat{OR}_i$, with variance estimated by

$$se^2(\log_e \widehat{OR}_i) = 1/a + 1/b + 1/c + 1/d.$$

- ▶ Let $w_i = 1/se^2(\log_e \widehat{OR}_i)$.
 - ▶ The larger the $se(\log_e \widehat{OR}_i)$, the smaller the weight w_i accorded to stratum i of the extraneous variable.
- ▶ A weighted average of the y_i 's that diminishes the contribution of values with high standard errors is:

$$Y = \sum_{i=1}^g \left(\frac{w_i}{\sum_{j=1}^g w_j} \right) y_i = \frac{\sum_{i=1}^g w_i y_i}{\sum_{j=1}^g w_j}.$$

- ▶ The test statistic tallies up the deviations of the estimated log-ORs from this average:

$$X^2 = \sum_{i=1}^g w_i (y_i - Y)^2$$

Large deviations suggest heterogeneous ORs.

- ▶ Under H_0 : homogeneous ORs, the statistic X^2 has a chi-square distribution with $g - 1$ df.

Test of Homogeneity in the AS Example

- In the AS example, the gender-specific tables are:

```
## , , sex = 0
##
##      smoke
## AS    0  1
##    0 47 19
##    1 29 14
##
## , , sex = 1
##
##      smoke
## AS    0  1
##    0 20 24
##    1 25 37
```

- ▶ We test for homogeneity of the ORs for AS across the strata defined by sex at level 5%.
- ▶ We get (see R demo):

```
## $X2  
## [1] 0.003072521  
##  
## $pval  
## [1] 0.9557956
```

- ▶ Retain the null hypothesis of homogeneous ORs at 5% level.
Insufficient evidence that sex modifies the effect of smoking on AS.
 - ▶ See text, section 16.2.1 for another example.
- ▶ Next question (and topic): Sex doesn't appear to be an effect modifier but is it a confounder?

Confounding

What is a Confounder?

- ▶ A confounding variable is an extraneous variable that is associated with both the disease outcome and the exposure of interest.
- ▶ Why is this important?
- ▶ Ignoring a confounder leads to biased inference of the effect of the exposure on the disease outcome (i.e., incorrect estimate, test and CI) and potentially **wrong conclusions**.
- ▶ Confounding is thus a **big concern** in any observational study.
 - ▶ E.G. Confounding by health-care coverage confused researchers in the observational studies leading up to the WHI.
 - ▶ Women with better health-care coverage tended to have better disease outcomes and also to have better access to HRT. Hence HRT seemed to be associated with better disease outcomes.

Simpson's Paradox

- ▶ Confounding explains “Simpson's paradox”:
 - ▶ When the effect of an exposure on the odds of the disease outcome differs in the stratified and pooled analyses.
- ▶ In the AS example, the estimated ORs summarizing the association between AS and smoking, stratified by sex, are about the same: 1.19 for females and 1.23 for males.
 - ▶ These results suggest that, for both females and males, smoking *slightly* increases the odds of AS.
- ▶ However, the estimated OR from the pooled table across genders is 1.47.
 - ▶ This result suggests a *stronger* effect of smoking than we obtain from the stratified tables!

Effect Modifier vs. Confounder

- ▶ When considering the OR summarizing the association between the disease outcome and the exposure variable, the extraneous variable is:
 1. an *effect modifier* if the ORs for different strata of the extraneous variable are **different**.
 2. a *confounder* if the ORs for different strata of the extraneous variable are the **same** but different from the OR that we get when we ignore the extraneous variable altogether.

The Mantel-Haenszel (MH) estimator

- ▶ Need to adjust the OR for the effects of a confounding variable.
- ▶ The Mantel-Haenszel (MH) estimator of the OR adjusts for a confounding variable when the disease outcome and exposure variable are dichotomous (i.e. have 2 categories) and the confounding variable is categorical with 2 or more categories.

MH Summary OR (Section 16.2.2)

- ▶ The MH summary estimator is a weighted average of the stratum-specific OR estimators
 - ▶ The MH estimator was developed over 50 years ago, is common practice in Epidemiology, and has good statistical properties.
 - ▶ Hence it is worth knowing.
- ▶ The MH estimate of the OR is

$$\widehat{OR} = \frac{\sum_{i=1}^g a_i d_i / T_i}{\sum_{i=1}^g b_i c_i / T_i}$$

where a_i , b_i , c_i , d_i are from the table for the i th stratum of the extraneous variable, and T_i is the table total for this i th table.

- ▶ See the text, page 381 for a worked example of this calculation.

The MH Test

- ▶ Once we have the MH estimate of the common OR across the strata of the confounding variable, we may test the null hypothesis that this common OR is one.
 - ▶ i.e., that there is no association between the disease outcome and the exposure variable after adjusting for the confounding variable.
- ▶ The text, section 16.2.3, goes into the details of the derivation, if you are interested.
- ▶ Instead, we'll use R to apply this test (see R demo).

MH Test on AS Example

```
##  
## Mantel-Haenszel chi-squared test without continuity correction  
##  
## data:  smoke and AS and sex  
## Mantel-Haenszel X-squared = 0.44568, df = 1, p-value = 0.5044  
## alternative hypothesis: true common odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.6877312 2.1460496  
## sample estimates:  
## common odds ratio  
##           1.214868
```

- ▶ Referring to the p-value reported in the output, we see that, at the 5% level, there is no statistical evidence that smoking affects the odds of AS, once we adjust for sex.

Confidence Interval for the Common OR

- ▶ R returns a 95% CI of about (0.69, 2.1).
- ▶ Though not discussed, can extend the MH test to arbitrary values of the common OR.
 - ▶ We only covered the special case where the value hypothesized is 1; i.e. $H_0 : OR = 1$ (unassociated) vs. $H_a : OR \neq 1$ (associated).
- ▶ The 95% CI is obtained by finding the set of all OR values that would be retained in a test at the 5% level.
 - ▶ E.G., With the AS data, if we were to test $H_0 : OR = 0.69$ vs. $H_a : OR \neq 0.69$, we'd retain H_0 at 5% level.
 - ▶ Similarly, if we were to test $H_0 : OR = 2.1$ vs. $H_a : OR \neq 2.1$, we'd retain H_0 at 5% level.
 - ▶ And the same for all values in between.
 - ▶ Any values smaller than 0.69 or larger than 2.1 for H_0 would lead to rejection with these data.
- ▶ Thus the 95% CI obtained from the AS data is (0.69, 2.1).
 - ▶ This is called “inverting the test” and is how R gets the CI (but not the text, which we'll skip).

Summary

- ▶ Inference of an association between an exposure variable and disease outcome should consider extraneous variables that could be effect modifiers or confounders.
- ▶ We first look for effect modification by testing for homogeneous ORs across the strata of the extraneous variable.
- ▶ If an extraneous variable is not an effect modifier, we may adjust for it as a potential confounder by applying the MH estimator of the common OR.
 - ▶ The MH estimator automatically adjusts for the extraneous variable (but is only valid if there is no effect modification).
 - ▶ In this class, we'll omit the material on the approximate CI for the common OR given in section 16.2.2 of the text and instead use the output from R which “inverts” the hypothesis test as described on the previous page.