STATS 111/202

Lecture 5: Confounding and effect modifiers, MH-test, BD-test

1/22/2025

Ana Maria Kenney

UC Irvine Department of Statistics

Three-way tables (cont'd)

2.7

Recall confounders vs effect modifiers

Confounders

- **Masks** the true relationship between the explanatory variable and the outcome.
- Wholly or partially accounts for the observed effect of the explanatory variable on the outcome. Can lead to inaccurate conclusions

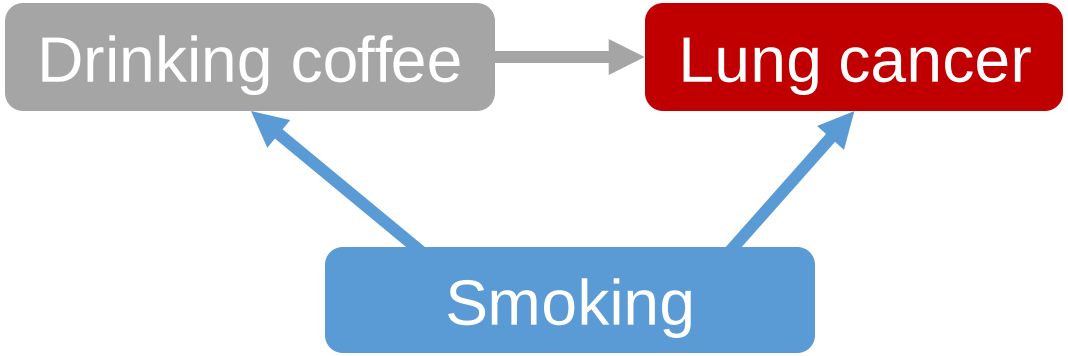
Effect modifiers

- **Alters** the relationship between the explanatory variable and the outcome.
- Natural behavior we want to understand. Different groups may have different outcomes when effect modification is present.

Recall confounders vs effect modifiers

Confounders

- **Masks** the true relationship between the explanatory variable and the outcome.
- Wholly or partially accounts for the observed effect of the explanatory variable on the outcome. Can lead to inaccurate conclusions

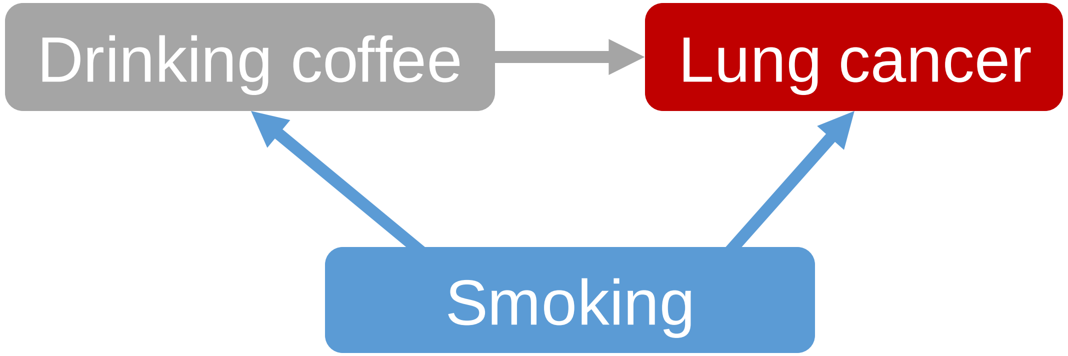


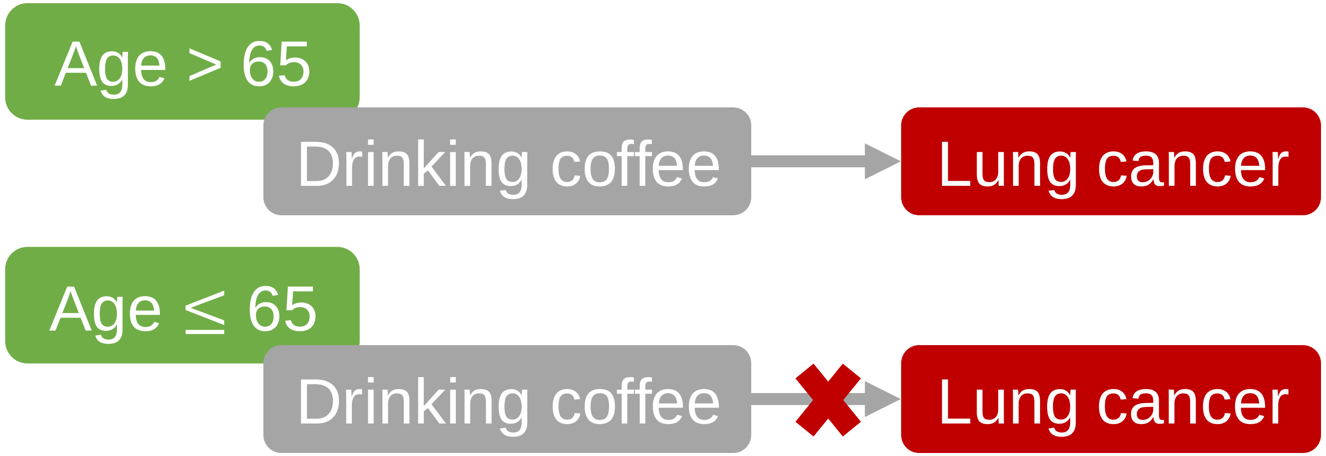
Effect modifiers

- **Alters** the relationship between the explanatory variable and the outcome.
- Natural behavior we want to understand. Different groups may have different outcomes when effect modification is present.

Recall confounders vs effect modifiers

Confounders	Effect modifiers
• **Masks** the true relationship between the explanatory variable and the outcome. • Wholly or partially accounts for the observed effect of the explanatory variable on the outcome. Can lead to inaccurate conclusions	• **Alters** the relationship between the explanatory variable and the outcome. • Natural behavior we want to understand. Different groups may have different outcomes when effect modification is present.





Recall confounders vs effect modifiers

Confounders

- A third variable W, may cause a spurious association between X and Y or even mask an effect taking place

Effect modifiers

- The association between X and Y varies at different levels of W

Possible confounder example

Possible effect modifier example

Overview

Assume we have a three-way table, which is really a two-way tables over levels of a third variable W.

- **Check confounding**
  - Combine tables to check for association between X and W
  - Combine tables to check for association between Y and W
  - Can use a chi-squared test for independence

Overview

Assume we have a three-way table, which is really a two-way tables over levels of a third variable W.

- **Check confounding**
  - Combine tables to check for association between X and W
  - Combine tables to check for association between Y and W
  - Can use a chi-squared test for independence

If we are worried about confounding, how can we see if there is an association between X and Y while controlling for W?

Overview

Assume we have a three-way table, which is really a two-way tables over levels of a third variable W.

- **Check confounding**
  - Combine tables to check for association between X and W
  - Combine tables to check for association between Y and W
  - Can use a chi-squared test for independence

- **Check for effect modification**
  - Use original tables to compute and test the association between X and Y over all levels of W

Overview

Assume we have a three-way table, which is really a two-way tables over levels of a third variable W.

- **Check confounding**
  - Combine tables to check for association between X and W
  - Combine tables to check for association between Y and W
  - Can use a chi-squared test for independence

- **Check for effect modification**
  - Use original tables to compute and test the association between X and Y over all levels of W

How can we test whether these associations are similar enough?

Mantel-Haenszel test

Goal: test whether X and Y are independent at each level of W

- Let W have k = 1, 2,…,K many levels (K $\geq$ 2)
- Consider X and Y taking on 2 levels at each level of W
- Then we have K of these tables

	Y=0	Y=1	Total
X=0	n_{11k}	n_{12k}	$n_{1.k}$
X=1	n_{21k}	n_{22k}	$n_{2.k}$
Total	$n_{.1k}$	$n_{.2k}$	$n_{..k}$

K tables total

- Let's condition on the column and row totals (these are fixed)
- Then only a single cell in the 2x2 table can vary, say n_{11k}

H_0: X and Y are independent at each level of W
H_α: X and Y are dependent at at least one level of W

Mantel-Haenszel test

Goal: test whether X and Y are independent at each level of W

- Let W have k = 1, 2,…,K many levels (K $\geq$ 2)
- Consider X and Y taking on 2 levels at each level of W
- Then we have K of these tables

	Y=0	Y=1	Total
X=0	n_{11k}	n_{12k}	$\boldsymbol{n_{1.k}}$
X=1	n_{21k}	n_{22k}	$n_{2.k}$
Total	$\boldsymbol{n_{.1k}}$	$n_{.2k}$	$\boldsymbol{n_{..k}}$

K tables total

- Let's condition on the column and row totals (these are fixed)
- Then only a single cell in the 2x2 table can vary, say n_{11k}

H_0: X and Y are independent at each level of W
H_α: X and Y are dependent at at least one level of W

Under the null, n_{11k} follows a Hypergeometric distribution with parameters ($n_{..k}$, $n_{.1k}$, $n_{1.k}$)

Recall: Hypergeometric distribution

Say we have N objects/trials with K successes. What if we randomly select (without replacement) n of them?

- Three important parameters: N, K, and n

- Let X be the number of successes from the n draws
  - X can take on values (i.e., has support) $\{\max(0, n + K - N), \ldots, \min(n, K)\}$

- The probability mass function is:
  - $f(x) = P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$

Mantel-Haenszel test

Under the null, n_{11k} follows a Hypergeometric distribution with parameters $(n_{..k}, n_{.1k}, n_{1.k})$

- Then under the null where X and Y are independent across all levels of W
  - $E(n_{11k}) = \frac{n_{1.k}n_{.1k}}{n_{..k}}$
  - $Var(n_{11k}) = \frac{n_{1.k}n_{2.k}n_{.1k}n_{.2k}}{n_{..k}^2(n_{..k}-1)}$
- And the MH test statistic follows:

$$MH = \frac{\left(\sum_{k=1}^{K}(n_{11k} - E(n_{11k}))\right)^2}{\sum_{k=1}^{K} Var(n_{11k})}$$

	Y=0	Y=1	Total
X=0	n_{11k}	n_{12k}	$\mathbf{n_{1.k}}$
X=1	n_{21k}	n_{22k}	$n_{2.k}$
Total	$\mathbf{n_{.1k}}$	$n_{.2k}$	$\mathbf{n_{..k}}$

This follows an approximate chi-squared distribution with 1 degree of freedom

Mantel-Haenszel test

- Under the null, the odds ratio in each level of W will be equal (and each is equal to 1). We can combine across levels of W to form a **single common odds ratio**
- Then the MH estimator of a common odds ratio in a 2x2xK table is given by:

$$\widehat{OR}_{MH} = \frac{\sum_{k=1}^{K}(n_{22k}n_{11k}/n_{..k})}{\sum_{k=1}^{K}(n_{21k}n_{12k}/n_{..k})}$$

	Y=0	Y=1	Total
X=0	n_{11k}	n_{12k}	$\boldsymbol{n_{1.k}}$
X=1	n_{21k}	n_{22k}	$n_{2.k}$
Total	$\boldsymbol{n_{.1k}}$	$n_{.2k}$	$\boldsymbol{n_{..k}}$

- The MH common odds ratio is the odds ratio for Y = 1 comparing X = 1 to X = 0 **while accounting or controlling for W (a potential confounder)**

MH common odds ratio intuition

- What happened to our odds ratio?

	Y=1	Y=0	Total
X=1	n_{11k}	n_{12k}	$n_{1.k}$
X=0	n_{21k}	n_{22k}	$n_{2.k}$
Total	$n_{.1k}$	$n_{.2k}$	$n_{..k}$

$$\widehat{OR}_{MH} = \frac{\sum_{k=1}^{K} (n_{22k} n_{11k} / n_{..k})}{\sum_{k=1}^{K} (n_{21k} n_{12k} / n_{..k})}.$$

$$OR = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} = \frac{\left(\frac{n_{11k}}{n_{1\cdot k}}\right) \Big/ \left(\frac{n_{12k}}{n_{1\cdot k}}\right)}{\left(\frac{n_{21k}}{n_{2\cdot k}}\right) \Big/ \left(\frac{n_{22k}}{n_{2\cdot k}}\right)}$$

$$= \frac{\left(\frac{n_{11k}}{n_{1\cdot k}}\right)\left(\frac{n_{22k}}{n_{2\cdot k}}\right)}{\left(\frac{n_{21k}}{n_{2\cdot k}}\right)\left(\frac{n_{12k}}{n_{1\cdot k}}\right)}$$

$$= \frac{\dfrac{n_{11k}\, n_{22k}}{\cancel{n_{1\cdot k}\, n_{2\cdot k}}}}{\dfrac{n_{21k}\, n_{12k}}{\cancel{n_{2\cdot k}\, n_{1\cdot k}}}}$$

$$= \frac{n_{11k}\, n_{22k}}{n_{21k}\, n_{12k}}$$

$$OR = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} = \frac{\left(\frac{n_{11k}}{n_{1\cdot k}}\right) \Big/ \left(\frac{n_{12k}}{n_{1\cdot k}}\right)}{\left(\frac{n_{21k}}{n_{2\cdot k}}\right) \Big/ \left(\frac{n_{22k}}{n_{2\cdot k}}\right)}$$

$$= \frac{\left(\frac{n_{11k}}{n_{1\cdot k}}\right)\left(\frac{n_{22k}}{n_{2\cdot k}}\right)}{\left(\frac{n_{21k}}{n_{2\cdot k}}\right)\left(\frac{n_{12k}}{n_{1\cdot k}}\right)}$$

$$= \frac{\cancel{\frac{n_{11k}\, n_{22k}}{n_{1\cdot k}\, n_{2\cdot k}}}}{\cancel{\frac{n_{21k}\, n_{12k}}{n_{2\cdot k}\, n_{1\cdot k}}}}$$

$$= \frac{n_{11k}\, n_{22k}}{n_{21k}\, n_{12k}}$$

$$\widehat{OR}_{MH} = \frac{\sum_{k=1}^{K} (n_{22k} n_{11k} / n_{..k})}{\sum_{k=1}^{K} (n_{21k} n_{12k} / n_{..k})}.$$

So we're just taking the weighted sum of the numerator and denominator (adjusting for the number of samples in each level of W)

Mantel-Haenszel test

- Then formally, the null hypothesis for the MH test is $H_0 : OR_{MH} = 1$

- The alternative can follow $H_\alpha: OR_{MH} \neq 1$ or $H_\alpha: OR_{MH} > 1$ or $H_\alpha: OR_{MH} < 1$

- This is testing for independence between X and Y while adjusting or controlling for W (a possible confounder)

Breslow-Day test

- Now if we want to directly test for a homogenous association, meaning the relationship between X and Y conditioned on W is the same across all levels of W, we would use the BD test

- $H_0: OR_1 = OR_2 = \cdots = OR_K \ vs \ H_a: OR_i \neq OR_j \ for \ at \ least \ one \ i \neq j$

- The B-D test statistic: $BD = \sum_{i,j,k} \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$
  - Where n_{ijk} is the observed count in the ij-th cell in the k-th table
  - $\hat{\mu}_{ijk}$ is the expected count in the ij-th cell in the k-th table under the null that all odds ratios are the same

- This test statistic follows a chi-squared distribution with K-1 degrees of freedom

MH and BD Test Example