STATS 111/202

Lecture 4: ordinal data cont'd and three-way tables

1/15/2025

Ana Maria Kenney

UC Irvine Department of Statistics

Ordinal data (cont’d)

2.5

CHD example #3

Three-way tables

2.7

Confounding

- **Confounding variables**
  - A variable that affects the response but is also related to the explanatory variable
  - In this case, its effect on the response cannot be separated from the effect on the explanatory variable
  - **Example: eating more icecream can lead to shark attacks**

- If a confounder W exists and is not controlled for, then we will obtain incorrect inference into how X is associated with Y

- The association between X and Y is confounded by W

- Ideally want to control for W either by design or including extra terms in the model

3-way tables

- Let’s assume X and Y are categorical variables and W is another categorical variable possibly confounding the relationship
- We can check the association between X and Y at each level of W
- Then we can check the association between X and Y combined over all levels of W
- Example: consider the effect of blood pressure on CHD. Other confounders to consider could be: sex, age group, diet, etc.

3-way tables

- **Effect modifier (interaction term)**
  - If the association between the explanatory variable X and the response Y depends on the level of W, then W is said to be an effect modifier
  - So the association between X and Y differs across different levels of W
  - There is an interaction that exists between X and W
  - The effect of X on Y depends on the level of W
  - Recall in regression, when we have a continuous X and Y with a W that has 2 levels, this meant that the slope of the line (coefficient on X) can vary between the two levels of W. **This was an interaction effect between W and X.**

Example

- Let Y be a categorical outcome of whether a sick patient was cured from a disease or not (0 not cured, 1 cured). Let X be a categorical explanatory variable of treatment (yes/no). Consider an effect modifier W which categorizes subjects into senior (above 65) or non-senior (below 65)
  - One can argue that the effect of X depends on the level of W
  - The effect of the treatment could be stronger for senior patients compared to non-senior patients

Example

- Recall odds are defined as $\frac{p}{1-p}$, ratio of probability of success vs failure
- Say the odds for non-seniors is $OR_{NS} = 1.4$
- Say the odds for seniors is $OR_S = 2$
- For non-seniors, there is a 40% higher odds of being cured if the treatment is used. For seniors, they have a 100% higher odds (2 times as much) of being cured if the take the treatment.

Three-way table

- So we can deal with these effect modifiers by stratifying on levels of W and seeing how these relationships change (or don't change)
- Examine the association between X and Y at every level of W
- In a regression setting, we would include an interaction terms between X and W (and then would test that term for significance)

Example

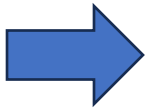
- Let's look at a 2x2x2 table (X, Y, and W each have two levels 0/1)

When W=0

	Y=0	Y=1
X=0	9	6
X=1	6	4

When W=1

	Y=0	Y=1
X=0	1	4
X=1	4	16



If we ignore levels of W

	Y=0	Y=1
X=0	10	10
X=1	10	20

Example

- Can study the relationships between X and W and Y and W

X and W

	W=0	W=1
X=0	15	5
X=1	10	20

Y and W

	W=0	W=1
Y=0	15	5
Y=1	10	20

- Now we can use our toolbox to study these 2x2 tables viewing W as the response and X or Y as explanatory variables.
- Chi-square test for independence
- Odds ratios
- If there is a relationship, then **W can be considered a potential confounder**

Announcements

- HW 1 due this Sunday at 11:59pm PT
- For additional practice:
  - See questions 1.1, 1.3, 1.7, 1.9, 1.15, 2.1a, 2.3, 2.5, 2.9, 2.11, 2.13, 2.15, 2.19a, 2.21 in the textbook (all have solutions in the back/online)