JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Statistical Interaction/Effect Modification

Effect Modification/Interaction

- Let's look at the results from a fictitious data set comparing two treatments for a fatal disease as to the impact of each on reducing deaths: there are 600 "younger" patients and 600 "older" patients in this random sample of 1,200 patients
- Here is the data on all patients

	All patients		
	Surgery	**Drug**	**Total**
Died	300	300	600
Survived	300	300	600
Total	600	600	

$$\hat{RR}_{death} = \frac{\hat{P}_{surgery}}{\hat{p}_{drug}} = \frac{300/600}{300/600} = 1; \hat{OR}_{death} = 1$$

- Surgery and drug groups have identical proportions dying (50% in each group)!

2

Effect Modification/Interaction

- Here is the data on only the 600 younger patients

	Younger patients		
	Surgery	**Drug**	**Total**
Died	100	200	300
Survived	200	100	300
Total	300	300	

$$\hat{RR}_{death\ younger} = \frac{\hat{P}_{surgery}}{\hat{p}_{drug}} = \frac{100/300}{200/300} = 0.5; \hat{OR}_{death\ younger} = 0.25$$

3

Effect Modification/Interaction

- Here is the data on only the 600 older patients

	Older patients		
	Surgery	**Drug**	**Total**
Died	200	100	300
Survived	100	200	300
Total	300	300	

$$\hat{RR}_{death\ olderr} = \frac{\hat{P}_{surgery}}{\hat{p}_{drug}} = \frac{200/300}{100/300} = 2.0; \hat{OR}_{death\ older} = 4.0$$

4

Effect Modification/Interaction

- A recap of the overall and age specific results

5

Effect Modification/Interaction

- Age is an effect modifier:
  - Age modifies the association between death and treatment (statistical interaction between age and treatment)!
- The association between death and treatment depends on age
  - Surgery is better for younger patients, drug therapy is better for the older patients

6

Effect Modification/Interaction

- The association between death and one variable (**treatment**) depends on the level of another variable (**age**)
- Here, it would not make sense to estimate one composite, overall measure of the association between death and treatment
- Best way to look at this data is to just look at the two tables (**young and old**) separately and estimate two separate death/treatment associations

7

Example: Tree Damage and Elevation

- Data on elevation and percentage of dead or badly damaged trees, from 64 Appalachian sites (reported by Committee on Monitoring and Assessment of Trends in Acid Deposition, 1986)
- Eight of the 64 sites are in Southern states
- Elevation information—whether the site was above or below 1,100 meters
- This is an observational study (why?)

8

Example: Tree Damage and Elevation

- Data for the first ten sites

```
list  damage elev_group region in 1/10

     +---------------------------------+
     | damage       elev_group   region |
     |---------------------------------|
  1. |      5   >=1,100 meters    South |
  2. |     13   >=1,100 meters    South |
  3. |      6   >=1,100 meters    South |
  4. |     21   >=1,100 meters    South |
  5. |      4   >=1,100 meters    South |
     |---------------------------------|
  6. |     20   < 1,100 meters    South |
  7. |     17   >=1,100 meters    South |
  8. |     31   < 1,100 meters    South |
  9. |     10   < 1,100 meters    North |
 10. |     28   < 1,100 meters    North |
     +---------------------------------+
```

9

Example: Tree Damage and Elevation

- Let's try to assess the relationship between the percentage of damaged trees and elevation—here is a boxplot of the percentage of damaged trees by elevation

10

Example: Tree Damage and Elevation

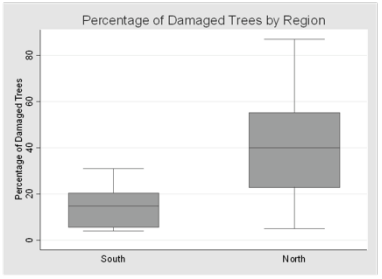
- Mean percentage of damaged trees by elevation group

Elevation	n	Mean tree damage (%)	SD
< 1,100 m	51	37.5	18.3
≥ 1,100 m	13	37.7	30.6

11

Example: Tree Damage and Elevation

- What about region though?
  - Boxplot percentage of damaged trees by region



12

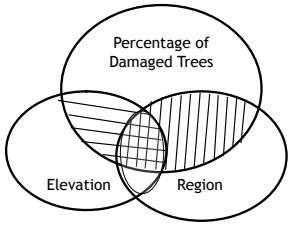
Example: Tree Damage and Elevation

- So sites in the South have less damage on average (i.e., not only is the percentage of damaged trees related to elevation, but it is also related to region)
- If region is related to elevation, then it's possible that part of the relationship we saw (or didn't see) between damage and elevation is because of the region-damage-elevation relationship

13

Example: Tree Damage and Elevation

- Possible diagram of this scenario



14

Example: Tree Damage and Elevation

- Recall the original finding—sites with lower elevation had a marginally lower percentage of damaged trees on average: 0.2% less damaged than sites at higher elevation
- To adjust for regional differences in the elevation groups, and the damage/region relationship, let's stratify by region, and estimate the damage/elevation association in each region

15

Example: Tree Damage and Elevation

- Mean percentage of damage tree by elevation in southern sites

	Southern sites only		
Elevation	**n**	**Mean tree damage (%)**	**SD**
< 1,100 m	2	25.5	7.7
≥ 1,100 m	6	11.0	7.1

  - Southern sites at higher elevation have 14.5% less damaged trees on average than southern sites at lower elevation (95% CI: 48.0 % lower—19.0 % higher)

16

Example: Tree Damage and Elevation

- Mean percentage of damage tree by elevation in northern sites

	Northern sites only		
Elevation	**n**	**Mean tree damage (%)**	**SD**
< 1,100 m	49	38.0	18.5
≥ 1,100 m	7	61.0	22.6

  - Northern sites at higher elevation have 23% greater damaged trees on average than northern sites at lower elevation (95% CI: 2 % higher—44 % higher)

17

Example: Tree Damage and Elevation

- A recap
  - Ignoring region, sites with lower PCV and sites with lower elevation had marginally lower percentage of damaged trees on average—0.2 % less damaged than sites at higher elevation
- When stratified by region . . .
  - Northern sites showed positive, statistically-significant association between damage and elevation
  - Southern sites showed negative, non-statistically-significant association between damage and elevation

18

Example: Tree Damage and Elevation

- So, it appears as though the association between tree damage and elevation is different, both in magnitude and direction depending on region
- We have a small dataset, so lack of statistical significance of negative damage/elevation associations in the south may be because of low power

19

Example: Tree Damage and Elevation

- One approach—take a weighted average of the average damage differences between sites at low and high elevations within each region, weighted by number of observations in each region
- However, does not necessarily make sense here—why combine estimates that differ in direction into one overall estimate?
- Better approach—to report two mean differences in damage between low and high elevation sites (one estimate for northern sites, one estimate for southern sites)
- Better approach—multiple regression methods (forthcoming!)

20

Lecture Evaluation

Please take a moment to evaluate this lecture. Your feedback is very important and will be used for future revisions. The Evaluation link is available on the lecture page.

21