Statistical Controls for Extraneous Variables

Basic Ideas

- We have a bivariate association between a dependent variable (DV or Y) and an independent variable (IV or X)
- We want to know whether this relationship has internal validity = it cannot be "explained away" by other factors
- The "other factors" are extraneous variables
- In experiments these are explained away by experimental control
- Now we will discuss statistical controls

An example

There is a positive correlation between the number of prostitutes in a city and the number of preachers (including all religious leaders of congregations) in a city

- Unit of analysis = cities
- Variables =
  - # of prostitutes
  - # of preachers

Which causes which?

Measures of Bivariate Association

1. Correlation: The correlation between X and Y is .56
2. Difference of Means

	Male	Female	Total
Mean on Y	3.8	2.8	3.5

3. Difference of Percentages

	Walked	Rode	Total
Male	35%	65%	100%
Female	20%	80%	100%

Elaboration Model

- The idea is: will that bivariate relationship stay the same or change when we control for another variable?
- In most real non-experimental research, we control for many independent variables at once
- But it is worth getting the basic ideas
- Z will be our symbol for the extraneous variable, the control variable

Possible Results of Elaboration

1. Replication: the bivariate relationship between X and Y (IV and DV) remains basically the same after control for Z
2. Explanation: the bivariate relationship between X and Y goes to zero after control for Z (or at least becomes weaker)
3. Suppression: the bivariate relationship between X and Y becomes stronger or changes sign after control for Z

Replication

1. **Correlations & regression: the partial correlation or partial regression coefficient in a multiple regression with other variables is the same as the bivariate correlation or regression coefficient**
2. **Difference of means: the difference of means on Y for categories of X is about the same for different categories of Z.**
3. **Percentage differences: the difference of percentages of categories of Y by categories of X is the same for different categories of Z**

Replication: Difference of Means

Mean Seconds to First Honk

	Sex of Driver		
Frustrator	**Male**	**Female**	**Total**
Low Status	**6.8**	**7.6**	**7.1**
High Status	**8.5**	**10.9**	**9.2**

The difference between low and high status frustrators is pretty much the same for both sexes.

Explanation

1. **Correlation/regression: the partial correlation or regression coefficient is much weaker (closer to zero) than the bivariate coefficient**
2. **Difference of means: the difference of means on Y for categories of X becomes much smaller when Z is controlled**
3. **Difference of percentages: the difference of %s for categories of Y by categories of X becomes much smaller when Z is controlled**

Explanation: Difference of Means

Mean Seconds to First Honk

	Sex of Driver		
Frustrator	**Male**	**Female**	**Total**
Low Status	**9.9**	**6.1**	**7.1**
High Status	**10.1**	**6.3**	**9.2**

Sex "explains" the apparent relation between low/high status and horn-honking.
The difference between low and high status frustrators is pretty much the same for both sexes. (This pattern could happen only if more women encountered the low status car and more men the high status car.)

Suppression

- **Statistical control reveals a "true" bivariate relation between X and Y that was hidden by Z**
  - **Bivariate relation between X and Y is close to zero, becomes strongly positive or negative with control for Z**
  - **Bivariate relation between X and Y changes sign with control for Z**
- **Many complex possibilities**

Example of Suppression

- **Hypothetical numbers based on real older research finding**
- **Black people had lower rates of community participation than Whites**

Community Participation Scale

	Total
Black	**3.1**
White	**4.3**

Example of Suppression

BUT control for class reveals that the TRUE relationship between race and community participation has been obscured in the bivariate table! (Hypothetical data but the research result did really happen.)

Community Participation Scale

	Economic Class			
	Low	Med	High	Total
Black	2.3	4.9	6.2	3.1
White	1.7	4.1	5.2	4.3

Reading multivariate tables

- **In a multiple regression, you see the simultaneous effects of several/many independent variables at once**
- **Every independent variable effect in a multivariate regression is automatically controlled for every other independent variable in the same equation**
- **We do not usually talk about replication, explanation, suppression etc. in this context**
- **We focus on the sign (+ vs -) and the significance of regression coefficients**