Relationships Between Two Variables: Cross-Tabulation

Introduction

- **Bivariate Analysis:** A statistical method designed to detect and describe the relationship between two variables.
- **Cross-Tabulation:** A technique for analyzing the relationship between two variables that have been organized in a table.

Bivariate Analysis

- Bivariate analysis is a statistical method designed to detect and describe the relationship between two variables.
- What is the relationship?

  When the temperature rises there is more criminal activity.

Cross-Tabulation

- Cross-tabulation (cross-tabs) is the most common technique used in the analysis of relationships between two variables.
- The variables are organized on a table to demonstrate the strength of the association and when appropriate its direction.

Home Ownership by Race

	Race		
Home Ownership	African-America	White	
Own	2	5	7
Rent	4	4	8
	6	9	

Independent and Dependent Variables

- What is the independent variable and the dependent variable in home ownership by race?
- What is the independent variable and dependent variable in English proficiency by nativity?

Independent and Dependent Variables

- For some variables, whether it is the independent or dependent variable depends on the research question.
- Example:
  - ❖ People who live in poverty have a larger problem with drug use.
  - ❖ People who use drugs tend to live in poverty.

How to Construct a Bivariate Table:

- A bivariate table displays the distribution of one variable across the categories of another variable.
- We must classify the cases based on their joint scores for the two variables.
- Think of it as a series of frequency distributions joined to make one table.

Home Ownership by Race

Race

Home Ownership	African-America	White	
Own	2	5	7
Rent	4	4	8
	6	9	

Row Marginals or Row totals

Column Marginals or Column Totals

Bivariate Tables Should Have the Following Features:

- A title that is descriptive.
- It should have two dimensions, one for each variable, (usually the independent variable is the column variable and the dependent variable is the row variable).
- The intersection of a row and a column is called a cell.

Bivariate Tables Should Have the Following Features:

- The column and row totals are the frequency distribution for each variable. The column total is the frequency distribution for independent variable and the row total is the frequency distribution for the dependent variable.
- We refer to tables as r$_x$c. So the table on homeownership by race is a 2x2 table.

Bivariate Tables Should Have the Following Features:

- The source of the data should always be clearly noted so that we are giving credit to the person who came up with the information.
- You should now know what a bivariate table is, a column variable, a row variable, a cell, and marginals.

How to Compute Percentages in a Bivariate Table

- When we use percentages it is easier for the reader to see the results.
- To calculate percentages you must calculate the percentages within each category of the independent variable.
- Interpret the table by comparing the percentage point difference for different categories of the independent variable.

Calculating Percentages within each Category

- The first rule means that we have to calculate percentages within each category of the variable that the investigator defines as the independent variable.
- So in the table “Home Ownership by Race” we would divide the number of black homeowners by the total number of blacks in the sample.

The Way People Voted in the 1992 Presidential Election by Sex

	Respondents Sex		
Voted for	Male	Female	Total
Bush	39%	33%	36%
Perot	19%	12%	15%
Clinton	42%	55%	49%
Total	100%	100%	100%

The Way People Voted in the 2000 Presidential Election by Sex

Voted for	Respondents Sex		
	Male	Female	Total
Gore	43%	54%	48%
Bush	54%	44%	48%
Buchanan	>1%	1%	1%
Nader	3%	2%	3%

The Way People Voted in the 2004 Presidential Election by Sex

Voted for	Respondents Sex		
	Male	Female	Total
Bush	55%	48%	51%
Kerry	44%	51%	48%

The Way People Voted in the 2008 Presidential Election by Sex

Voted for	Respondents Sex		
	Male	Female	Total
Obama	49%	56%	53%
McCain	56%	43%	46%

Comparing the Percentages

- The second rule says you must compare the percentages across different categories of the independent variable.
- In a 2x2 table there is only one comparison that needs to be made. For larger tables, more than one comparison needs to be made.

Properties of a Bivariate Relationship

Three questions we can ask when examining a bivariate relationship:

1. Does there appear to be a relationship?
2. How strong is the relationship?
3. What is the direction of the relationship?

The Existence of the Relationship

- When we calculate the percentages for the categories we can then compare the percentages to see if there is a relationship between the variables.
- Compare the next two slides and tell me whether or not they have a relationship.

Support for Abortion by Trauma

	Number of Traumas			
Abortion	0	1	2+	Total
Yes	19%	41%	76%	46%
No	81%	59%	24%	54%
Total	100%	100%	100%	100%
(N)	(27)	(44)	(33)	(104)

Support for Abortion by Trauma

	Number of Traumas			
Abortion	0	1	2+	Total
Yes	46%	46%	46%	46%
No	54%	54%	54%	54%
Total	100%	100%	100%	100%
(N)	(27)	(44)	(33)	(104)

The Strength of the Relationship

- Once you have established that there is a relationship, the next step is to see the strength of the relationship.
- The easiest way is to examine the percentage difference across the different categories of the independent variable. The larger the difference across the categories, the stronger the association.

The Strength of the Relationship

- In the hypothetical example of no relationship between trauma and attitude toward abortion there is a 0 percent difference between the columns.
- At the other extreme, if all women who suffered 1 or more traumas were pro-choice, and none of the women with 0 traumas was pro-choice there is a 100% difference between the columns.

The Direction of the Relationship

- When both the independent and dependent variable in the bivariate table are measured at the ordinal level we can look at the direction of the relationship.
- There are two types of relationships:
  - Positive relationship
  - Negative relationship

Positive relationship

- In a positive bivariate relationship the two variables vary in the same direction.
- Basically, what I mean is when the independent variable increases so does the dependent variable.
- Example

Health Condition by Social Class

	Class		
Health	Low	Middle	High
Poor	39%	12%	9%
Fair	36%	45%	28%
Good	25%	43%	63%
Total	100%	100%	100%
(N)	(39)	(254)	(202)

Negative Relationship

- In a negative bivariate relationship the two variables vary in opposite directions.
- When one variable increases the other variable will decrease.
- Example

Frequency of Trauma by Social Class

	Class		
Trauma	Low	Middle	High
0	15%	42%	48%
1	38%	36%	32%
2+	47%	22%	20%
Total	100%	100%	100%
(N)	(48)	(220)	(180)

Elaboration

- Elaboration is a process designed to further explore a bivariate relationship.
- We look at other variables called control variables to see if they explain the bivariate relationship under consideration.

Three Goals of Elaboration

1. Elaboration allows us to test for nonspuriousness.
2. Elaboration clarifies the causal sequence of bivariate relationships by introducing variables hypothesized to intervene between the IV and DV.
3. Elaboration specifies the different conditions under which the original bivariate relationship might hold.

Why Would we Introduce a Control Variable?

1. Elaboration allows us to test for nonspuriousness.
   - This shows us if there is a cause and effect relationship between the independent and dependent variables in the bivariate relationship.
   - It also establishes the time order between them.
   - Also, we want to know if there is another variable that better fits the relationship.

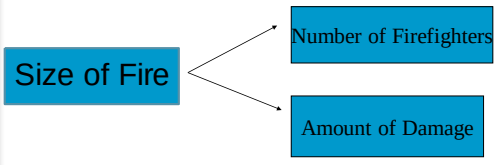
Testing for Nonspuriousness

- **Direct causal relationship:** a bivariate relationship that cannot be accounted for by other theoretically relevant variables.
- **Spurious relationship:** a relationship in which both the IV and DV are influenced by a causally prior control variable and there is no causal link between them. The relationship between the IV and DV is said to be "explained away" by the control variable.

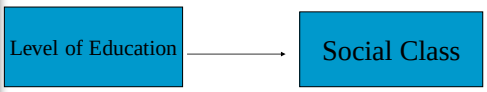
The Bivariate Relationship Between Number of Firefighters and Property Damage

Number of Firefighters → Property Damage
(IV) (DV)

A Spurious Relationship



A Direct Causal Relationship



Why Would we Introduce a Control Variable?

2. Elaboration clarifies the causal sequence.
   - It does this by introducing variables hypothesized to intervene between the independent and dependent variables.
   - Example: What else might cause the dependent variable besides the independent variable?

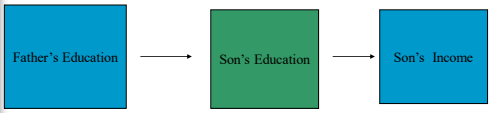
Intervening Relationship

- ***Intervening variable:*** a control variable that follows an independent variable but precedes the dependent variable in a causal sequence.
- ***Intervening relationship:*** a relationship in which the control variable intervenes between the independent and dependent variables.

Intervening Variable

- Is there a direct casual link between the two variables. Sometimes the variables are linked but, there is an intervening variable that explains the relationship even more.

Intervening Variable



Why Would we Introduce a Control Variable?

3. Elaboration specifies conditional relationships.
   - A conditional relationship is a relationship between the independent and dependent variables that differs for different conditions of the control variable.
   - What other variables may explain the relationship between father's education and son's income?

Limitations of Elaboration

- Elaboration really is just a stab in the dark, you are trying to see if other variables (control variables) will explain the relationship better.
- The problem arises when you miss an important variable that may help explain the relationship.