Section 4.4

Contingency Tables and Association

Definitions

- Contingency Table (Two-Way Table): Relates two categories of data
- Row Variable: Each row in a table describes a specific topic, the title of the topics is the row variable
- Column Variable: Each column in a table describes a specific topic, the title of the topics is the column variable
- Cell: Each box inside the table is called a cell

Example

Hair Color	Education		
	High School	College	Graduate
Red	45	33	2
Brunette	90	42	7

Education is our column variable and Hair Color is our row variable. The high school graduates with red hair would be in the cell in the first row, first column

Definitions

- Marginal Distribution of a Variable : a frequency or relative frequency distribution of either the row or column variable in the contingency table

Creating a Frequency Marginal Distribution

1. Total all the columns and all the rows
2. Put a grand total in the bottom right

Example

Hair Color	Education			
	High School	College	Graduate	Totals
Red	45	33	2	80
Brunette	90	42	7	139
Totals	135	75	9	219

Creating a Relative Frequency Marginal Distribution

1. Total all the columns and all the rows
2. Put a grand total in the bottom right
3. Create a new row and column: divide each total by the grand total

Example

Hair Color	Education High School	College	Graduate	Totals	Rel. Freq.
Red	45	33	2	80	80/219 =0.365
Brunette	90	42	7	139	139/219 =0.635
Totals	135	75	9	219	
Rel. Freq.	135/219 =0.616	75/219 =0.342	9/219 =0.041		

Definition

- Conditional Distribution: Lists the relative frequency of each category of the response variable, given a specific value of the explanatory variable in the contingency table.

Creating a Conditional Distribution

1. Total all the columns and all the rows
2. Decide if you are studying it by column variable or row variable
3. Divide each number inside the table by the appropriate row or column total that it resides in

Example (by Education)

Hair Color	Education High School	College	Graduate	Totals
Red	45	33	2	80
Brunette	90	42	7	139
Totals	135	75	9	

Hair Color	Education High School	College	Graduate	Totals
Red	45/135 =0.333	33/75 =0.44	2/9 =0.222	80
Brunette	90/135 =0.667	42/75 =0.56	7/9 =0.778	139
Totals	135	75	9	

Simpson's Paradox

- Describes a situation in which an association between two variables inverts or goes away when a third variable is introduced to the analysis.
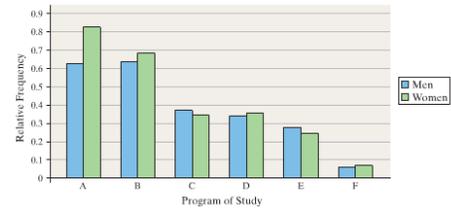
Sex Bias in Graduate Admissions (University of California)

TABLE 13			
	Accepted (A)	**Not Accepted (NA)**	**Total**
Men	1191	1399	**2590**
Women	557	1278	**1835**
Total	**1748**	**2677**	**4425**

Sex Bias in Graduate Admissions (University of California)

Conditional Distribution for Men and Women Admitted by Program of Study

Conclusion

- According to Sullivan: "The initial analysis did not account for the lurking variable, program of study. There were many more male applicants in programs A and B than female applicants, and these two programs happen to have higher acceptance rate. The higher acceptance rates in these programs led to the false conclusion that the Univ. of Calif., Berkeley, was biased against gender in its admissions"