

Chapter 10: Analyzing the Association Between Categorical Variables

Section 10.1: What is Independence and What is Association?

Learning Objectives

1. Comparing Percentages
2. Independence vs. Dependence

Learning Objective 1:
Example: Is There an Association Between Happiness and Family Income?

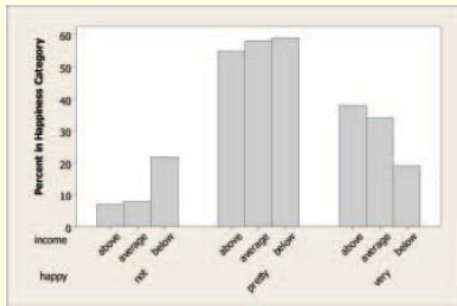
INCOME	Happiness			Total
	Not too Happy	Pretty Happy	Very Happy	
Above average	21	159	110	290
Average	53	372	221	646
Below average	94	249	83	426

Learning Objective 1:
Example: Is There an Association Between Happiness and Family Income?

- The percentages in a particular row of a table are called **conditional percentages**
- They form the **conditional distribution** for happiness, given a particular income level

INCOME	Happiness			TOTAL
	Not too Happy	Pretty Happy	Very Happy	
Above average	7%	55%	38%	290 (100%)
Average	8%	58%	34%	646 (100%)
Below average	22%	59%	19%	426 (100%)

Learning Objective 1:
Example: Is There an Association Between Happiness and Family Income?



5

Learning Objective 1:
Example: Is There an Association Between Happiness and Family Income?

- Guidelines when constructing tables with conditional distributions:
 - Make the response variable the column variable
 - Compute conditional proportions for the response variable within each row
 - Include the total sample sizes

6

Learning Objective 2:
Independence vs. Dependence

- For two variables to be *independent*, the population percentage in any category of one variable is the same for all categories of the other variable
- For two variables to be *dependent (or associated)*, the population percentages in the categories are not all the same

TABLE 11.3: Population Conditional Distributions
Showing Independence

The conditional distribution of happiness is the same for each gender, namely (10%, 58%, 32%).

GENDER	Happiness			TOTAL
	Not too Happy	Pretty Happy	Very Happy	
Females	10%	58%	32%	100%
Males	10%	58%	32%	100%
Overall	10%	58%	32%	100%

7

Learning Objective 2:
Independence vs. Dependence

- Are race and belief in life after death independent or dependent?
 - The conditional distributions in the table are similar but not exactly identical
 - It is tempting to conclude that the variables are dependent

TABLE 11.4: Sample Conditional Distributions for Belief in Life After Death, Given Race

RACE	Belief in Life After Death		TOTAL
	Yes	No	
White	1612 (83.2%)	326 (16.8%)	1938 (100%)
Black	312 (84.3%)	58 (15.7%)	370 (100%)
Other	253 (78.8%)	68 (21.2%)	321 (100%)
Overall	2177 (82.8%)	452 (17.2%)	2629 (100%)

Learning Objective 2:
Independence vs. Dependence

- Are race and belief in life after death independent or dependent?
 - The definition of independence between variables refers to a *population*
 - The table is a *sample*, not a *population*

9

Learning Objective 2:
Independence vs. Dependence

- Even if variables are independent, we would not expect the sample conditional distributions to be identical
- Because of sampling variability, each *sample* percentage typically differs somewhat from the *true population* percentage

10