

What Can You Do with Multivariate Data?

- ✓

Video:

 Looking at Associations with Multivariate Categorical Data
9 min
- ✓

Video:

 Looking at Associations with Multivariate Quantitative Data
7 min
- ✓

Practice Quiz:

 Practice Quiz: Multivariate Data
10 questions

Try it Out - What Can YOU Do with Data?

- ✓

Video:

 Demo: Interactive Scatterplot
2 min
- 📖

Reading:

 Pitfall: Simpson's Paradox
10 min
- 💬

Discussion Prompt:

 Discussion: Find Your Own Example
15 min
- 📖

Reading:

 Modern Ways to Visualize Data
10 min
- ▶

Video:

 Introduction to Pizza Assignment
2 min
- 📖

Reading:

 Pizza Study Design Assignment Instructions
10 min
- 🔒

Peer-graded Assignment:

 Pizza Study Design Assignment
1h
- 🔒

Review Your Peers:

 Pizza Study Design Assignment

Using Python for analysis of multivariate data

Week 3 Python Assignment

Pitfall: Simpson's Paradox

In this section, we're going to introduce a Pitfall within Statistics: **Simpson's Paradox**.

What if I told you that the rate of completing Course 2 after you've completed Course 1 of a Specialization is higher for males than for females? What would you think?

Now, what if I told you that more males are enrolled in a full Specialization, while more females are interested in taking each course as a freestanding course? Does that change how you think about the situation?

What might actually be happening is that there are different rates of Course 2 completion for those in the Specialization versus those taking freestanding courses, and that the gender ratio may also be different for those two situations.

Let's take a look at the numbers:

	Specialization	Course	Total
Men	55/70 (79%)	10/20 (50%)	65/90 (72%)
Women	25/30 (83%)	17/30 (57%)	42/60 (70%)
Total	80/100 (80%)	27/50 (54%)	

You may notice that the completion rate of Course 2 is higher for women within each of the methods of taking this course, but overall it appears that the completion rate of Course 2 is higher for men. The way of taking courses hides the true direction of the relationship between gender and completion rate of Course 2.

The method of taking the course can be considered a **confounding variable**. A confounding variable is an outside influence that changes the relationship between the independent and the dependent variable. It oftentimes works by affecting the causal relationship between the primary independent variable and the dependent variable. This confounding variable confuses the relationship between two other variables; it may act by hiding, obscuring, or enhancing the existing relationship.

For example, suppose that you are interested in examining how activity level affects weight change. Other factors, like diet, age, and gender may also affect weight change and come into play when looking at the relationship between activity level and weight change. If one doesn't control for these factors, the relationship between activity level and weight change can be distorted.

This also serves as a reminder that *Correlation* does not imply *Causation*. There may be other factors involved with a causal relationship. For this reason, you shouldn't place too much stock in correlation alone.

There are many other examples of Simpson's Paradox, including [this article on the relationship between gender and Nobel prize awards](#). In the next discussion, you'll have a chance to think about some examples of Simnson's Paradox in your daily life.