Causation and Lurking Variables: Summary

Learning Objective: Recognize the distinction between association and causation, and identify potential lurking variables for explaining an observed relationship.

Learning Objective: Recognize and explain the phenomenon of Simpson's paradox as it relates to interpreting the relationship between two variables.

It is *not* always the case that including a lurking variable makes us rethink the direction of the association. In the next example we will see how including a lurking variable just helps us gain a deeper understanding of the observed relationship.

Example: College Entrance Exams

As discussed earlier, in the United States, the SAT is a widely used college entrance examination, required by the most prestigious schools. In some states, a different college entrance examination is prevalent, the ACT.

Including a Lurking Variable

Start of transcript. Skip to the end.

The following scatterplot displays the relationship between the percentage of

students taking the SAT and the median SAT Math scores in each of the fifty states.

Note that the explanatory variable is the percent taking the SAT in each of the 50 states

and the response variable is the SAT Math median score in each of the

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

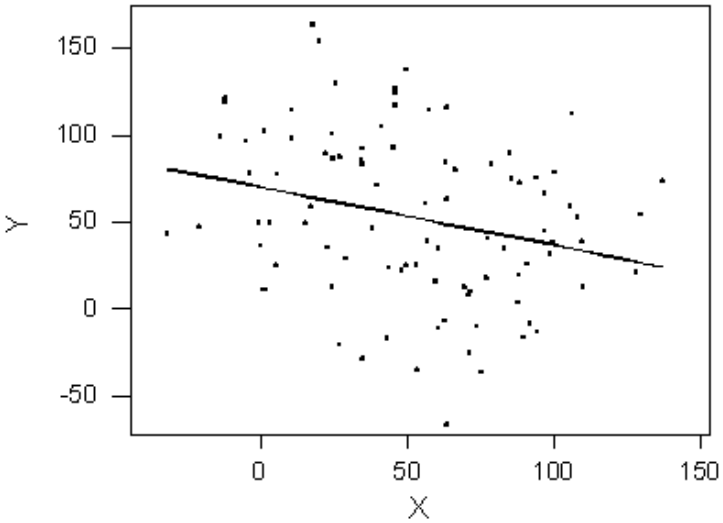
The last two examples showed us that including a lurking variable in our exploration may:

- lead us to rethink the direction of an association (as in the Hospital/Death Rate example) or,
- help us to gain a deeper understanding of the relationship between variables (as in the SAT/ACT example).

Scenario: Simpson's Paradox

Recall that an instance of Simpson's paradox occurs whenever including a lurking variable causes you to rethink the direction of an association. In this activity, we will reinforce our understanding of Simpson's paradox.

The following scatterplot displays the relationship between two quantitative variables, X and Y:



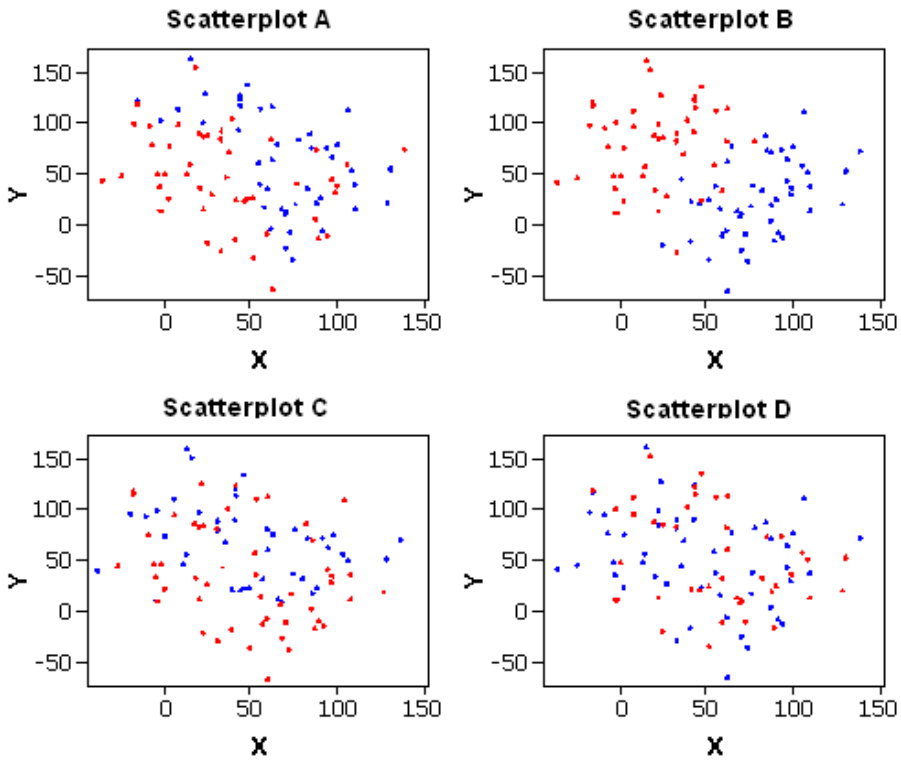
This graphical display indicates that the overall relationship between X and Y is **negative**.

In each of the four labeled scatterplots below, we included a different lurking variable, which separates the data points into two groups, blue points and red points. Your task is to look at all four displays and decide in which case including the lurking variable leads to an instance of Simpson's paradox.

Learn By Doing

1/1 point (graded)

Which scatterplot is an example of Simpson's paradox?



- Scatterplot A
- Scatterplot B ✔
- Scatterplot C
- Scatterplot D

Answer
Correct:
While the **overall relationship between X and Y is negative**, the relationship within the two groups created by this lurking variable (represented by the blue and red groups), **is positive**. Therefore, including this lurking variable in our exploration **does change** the apparent direction of the relationship between X and Y.

Submit

Did I Get This

1/1 point (graded)
A study was done on the graduation rates (graduate in 4 years versus did not graduate) from two universities: Mode College and Median University. Data were collected for the year 2012 in a variety of majors, and it was found that Mode College does better overall (i.e., has a higher percentage of students who graduate in 4 years). However, in each of the majors separately, Median University does better.

Which of the following is true? Check all that apply.

- [x] This is an example of Simpson's Paradox.
- [ ] This is an example of a negative association between variables.
- [ ] This situation is mathematically impossible.
- [x] “Type of major” could be a lurking variable in this example.

✔

Answer
Correct:
This is an example of Simpson’s Paradox, and "type of major" could be a lurking variable in this example.

Submit

Let's Summarize

- A **lurking variable** is a variable that was not included in your analysis, but that could substantially change your interpretation of the data if it were included.
- Because of the possibility of lurking variables, we adhere to the principle that **association does not imply causation**.
- Including a lurking variable in our exploration may:
  - help us to gain a deeper understanding of the relationship between variables, or
  - lead us to rethink the direction of an association.
- Whenever including a lurking variable causes us to rethink the direction of an association, this is an instance of **Simpson's paradox**.