

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > EDA: Examining Relationships > Causation > Causation and Lurking Variables: Confounds

🔖 Bookmark this page

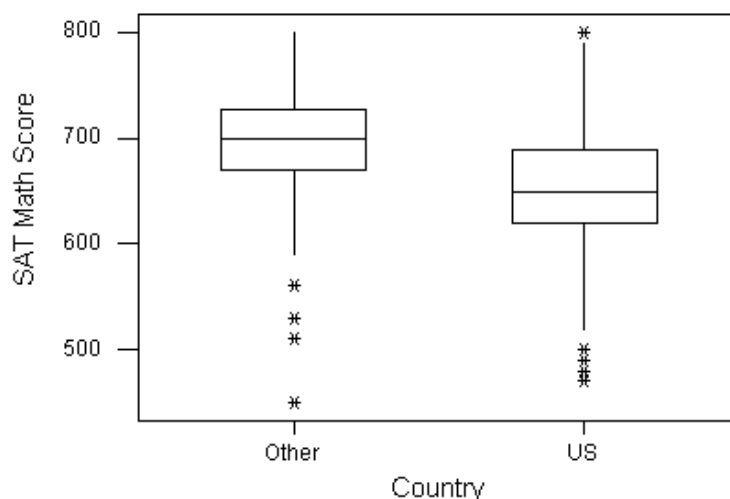
Causation and Lurking Variables: Confounds

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

The next example will illustrate another way in which a lurking variable might interfere and prevent us from reaching any causal conclusions.

Example: SAT Test

For U.S. colleges and universities, a standard entrance examination is the SAT test. The side-by-side boxplots below provide evidence of a relationship between the student's country of origin (the United States or another country) and the student's SAT Math score.

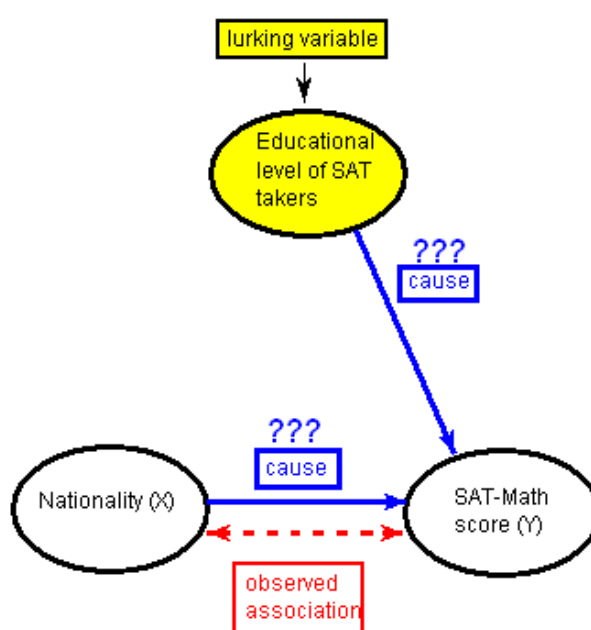


The distribution of international students' scores is higher than that of U.S. students. The international students' median score (about 700) exceeds the third quartile of U.S. students' scores. Can we conclude that the country of origin is the **cause** of the difference in SAT Math scores, and that students

in the United States are weaker at math than students in other countries?

No, not necessarily. While it *might* be true that U.S. students differ in math ability from other students—i.e. due to differences in educational systems—we can't conclude that a student's country of origin is the cause of the disparity. One important **lurking variable** that might explain the observed relationship is the educational level of the two populations taking the SAT Math test. In the United States, the SAT is a standard test, and therefore a broad cross-section of all U.S. students (in terms of educational level) take this test. Among all international students, on the other hand, only those who plan on coming to the U.S. to study, which is usually a more selected subgroup, take the test.

The following figure will help you visualize this explanation:



Here, the explanatory variable (X) **may** have a causal relationship with the response variable (Y), but the lurking variable might be a contributing factor as well, which makes it very hard to isolate the effect of the explanatory variable and prove that it has a causal link with the response variable. In this case, we say that the lurking variable is **confounded** with the explanatory variable, since their effects on the response variable cannot be distinguished from each other.

Note that in each of the above two examples, the lurking variable interacts differently with the variables studied. In example 1, the lurking variable has an effect on both the explanatory and the response variables, creating the illusion that there is a causal link between them. In example two, the lurking variable is confounded with the explanatory variable, making it hard to assess the isolated effect of the explanatory variable on the response variable.

The distinction between these two types of interactions is not as important as the fact that in either case, the observed association can be at least partially explained by the lurking variable. The most important message from these two examples is therefore: **An observed association between two variables is not enough evidence that there is a causal relationship between them.**

In other words ...

Principle

Association **does not** imply causation!

Learn By Doing

1/1 point (graded)

Research has suggested that there is a correlation between Internet usage and obesity. Using this information, a fictional study found a correlation of $r = 0.79$ between the number of hours of daily Internet usage and Body Mass Index, a measure of body weight.

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

- ☐ There is a pretty strong positive linear relationship between the number of hours spent on the Internet and the Body Mass Index scores.
- ☐ Using the Internet causes people to gain weight.
- ☒ Physical activity is a possible “lurking” variable behind this relationship, since people who spend a large amount of time using the Internet may be more likely to be less physically active.



Answer

Correct: It is possible that a lurking variable, such as physical activity, is behind the relationship.

Submit

Learn By Doing

1/1 point (graded)

The tourism industry hires a marketing company to promote overseas vacations. The marketing company's research shows a $+0.72$ correlation between the number of overseas vacations booked and amount of free time, in days, people have each year. Therefore, the marketing company decides to develop an advertising campaign aimed specifically at people who have lots of free time.

Which of the following is true about this decision?

- ☒ The decision may not be a good one, since it is possible that there is a “lurking” variable, such as income, behind both the number of booked overseas vacations and the number of free days people have each year. ✓

- ☐ The decision is a good one, since we know that the number of free days people have causes them to book more vacations.
- ☐ The decision is a good one, since there is a strong positive relationship between number of booked vacations and the amount of free days people have each year.

Answer

Correct: It is possible that there is lurking variable, such as income, which is behind the relationship.

Submit

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved