Causation

Victor I. Piercey

October 28, 2009

What does a high correlation mean?

If you have high correlation, can you necessarily infer causation?
What issues can arise?

What does a high correlation mean?

If you have high correlation, can you necessarily infer causation? What issues can arise?

Consider the following examples where a relationship has been observed:

1. x is a mother's body mass index, and y is a daughter's body mass index.

2. x is the amount of saccharin in a rat's diet and y is the count of tumors in the rat's bladder.

Consider the following examples where a relationship has been observed:

1. x is a mother's body mass index, and y is a daughter's body mass index.

2. x is the amount of saccharin in a rat's diet and y is the count of tumors in the rat's bladder.

In the first example, from what we know about genetics we expect the relationship to involve direct causation.

However, in a certain study the correlation coefficient was $r = 0.506$ so that $r^2 = 0.256$. What gives?

Consider the second example. Should we avoid saccharin ourselves?

In the first example, from what we know about genetics we expect the relationship to involve direct causation.

However, in a certain study the correlation coefficient was $r = 0.506$ so that $r^2 = 0.256$. What gives?

Consider the second example. Should we avoid saccharin ourselves?

The first example shows that even when direct causation is present, the explanatory variable is rarely the only quantity which affects the response variable (beware of lurking variables!!)

The second example shows that one should be very careful in making generalizations.

The first example shows that even when direct causation is present, the explanatory variable is rarely the only quantity which affects the response variable (beware of lurking variables!!)

The second example shows that one should be very careful in making generalizations.

Consider the next two examples of associations:

1. x is a high school senior's SAT score and y is the student's first year college GPA.

2. x is the monthly flow of money into mutual funds and y is the monthly rate of return in the stock market.

Consider the next two examples of associations:

1. x is a high school senior's SAT score and y is the student's first year college GPA.

2. x is the monthly flow of money into mutual funds and y is the monthly rate of return in the stock market.

In the first example, students who study hard tend to have both high SAT scores and high college GPA's.

In the second example, people tend to put more money into the stock market when they are optimistic, and the stock market also goes up when investors are optimistic.

In both of these examples, we have "common response" to some other lurking variable.

In the first example, students who study hard tend to have both high SAT scores and high college GPA's.

In the second example, people tend to put more money into the stock market when they are optimistic, and the stock market also goes up when investors are optimistic.

In both of these examples, we have "common response" to some other lurking variable.

In the first example, students who study hard tend to have both high SAT scores and high college GPA's.

In the second example, people tend to put more money into the stock market when they are optimistic, and the stock market also goes up when investors are optimistic.

In both of these examples, we have "common response" to some other lurking variable.

Consider the next (and final) pair of examples of association:

1. x is whether a person regularly attends religious services and y is how long the person lives.

2. x is the number of years of education a worker has and y is the worker's income.

Consider the next (and final) pair of examples of association:

1. x is whether a person regularly attends religious services and y is how long the person lives.

2. x is the number of years of education a worker has and y is the worker's income.

In the first example, it could be that being part of a religious community raises one's spirit and elongates life, but it is also possible that people who regularly attend religious services also tend to take better care of themselves.

In the second example, it could be that higher paying jobs go to those workers with more education or it could be that people with more education come from better-off families.

In these examples, the explanatory variable is confounded with a lurking variable.

Two variables are **confounded** when their effects on a response variable cannot be distinguished from one another.

In the first example, it could be that being part of a religious community raises one's spirit and elongates life, but it is also possible that people who regularly attend religious services also tend to take better care of themselves.

In the second example, it could be that higher paying jobs go to those workers with more education or it could be that people with more education come from better-off families.

In these examples, the explanatory variable is confounded with a lurking variable.

Two variables are **confounded** when their effects on a response variable cannot be distinguished from one another.

In the first example, it could be that being part of a religious community raises one's spirit and elongates life, but it is also possible that people who regularly attend religious services also tend to take better care of themselves.

In the second example, it could be that higher paying jobs go to those workers with more education or it could be that people with more education come from better-off families.

In these examples, the explanatory variable is confounded with a lurking variable.

Two variables are **confounded** when their effects on a response variable cannot be distinguished from one another.

In the first example, it could be that being part of a religious community raises one's spirit and elongates life, but it is also possible that people who regularly attend religious services also tend to take better care of themselves.

In the second example, it could be that higher paying jobs go to those workers with more education or it could be that people with more education come from better-off families.

In these examples, the explanatory variable is confounded with a lurking variable.

Two variables are **confounded** when their effects on a response variable cannot be distinguished from one another.

Summary: Be careful to distinguish the following scenarios:

1. direct causation
2. common response
3. confounding

Summary: Be careful to distinguish the following scenarios:

1. direct causation
2. common response
3. confounding

Summary: Be careful to distinguish the following scenarios:

1. direct causation

2. common response

3. confounding

The best way to establish causation is in a controlled experiment where all lurking variables are controlled.

In an observational study, good evidence of causation requires:

- a strong association that appears consistently in several studies,
- a clear explanation for the claimed causal link, and
- a careful examination of possible lurking variables.

The best way to establish causation is in a controlled experiment where all lurking variables are controlled.

In an observational study, good evidence of causation requires:

- a strong association that appears consistently in several studies,
- a clear explanation for the claimed causal link, and
- a careful examination of possible lurking variables.

The best way to establish causation is in a controlled experiment where all lurking variables are controlled.

In an observational study, good evidence of causation requires:

- a strong association that appears consistently in several studies,
- a clear explanation for the claimed causal link, and
- a careful examination of possible lurking variables.

The best way to establish causation is in a controlled experiment where all lurking variables are controlled.

In an observational study, good evidence of causation requires:

- a strong association that appears consistently in several studies,
- a clear explanation for the claimed causal link, and
- a careful examination of possible lurking variables.

Assignment: Page 312, Problems 4.41,4.42, 4.45 and 4.48; and Page 316, Problems 4.49 and 4.50 (due Friday).