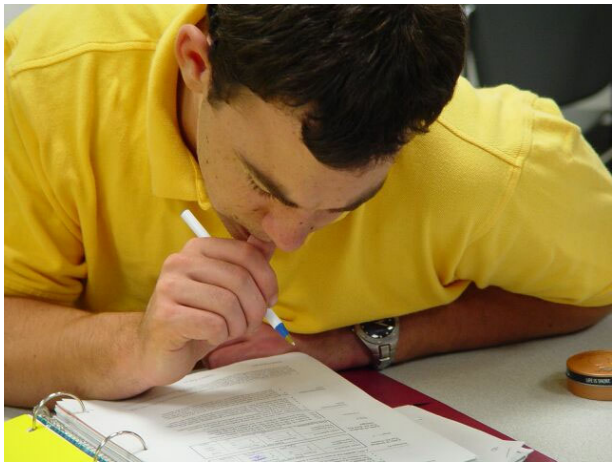
Statistical Literacy©

Critical thinking about arguments that use statistics as evidence



ASSOCIATION

CAUSATION

StatLit-Project-Brochure 2B 6/12 2002 ©

Association-Causation

Causation is much more than a mere association. This distinction is fundamental to our thinking.

In some cases, causation is quite clear.

- Intent: The batter hit the ball.
- Reason/Goal: He worked to make money.
- Nature: Acorns develop into oak trees.
- Enabler: Accidents due to the icy roads.
- Trigger: Earthquake triggers avalanche.
- Suppressor: Dampness suppresses fire.

Association can be a sign of causation.

- Jack and Jill are spending time together. Are they romantically involved?
- My boss seems to be avoiding me. Did I do something wrong? Am I being laid off?

But all too often there are alternate explanations for an observed association.

- Jack and Jill are spending a lot of time together because they are in the play.
- My boss is avoiding everybody because she is having problems with her boss.

Moral:

Always look for alternate explanations of associations in non-experimental situations! Always look for plausible confounders.

Grammar

A few words clearly distinguish association (correlate) from causation (cause, effect).

Many words can indicate either association or causation depending on the context. E.g., *due to, because of, the reason for, factor, condition, component, influences, linked, follows after, then, so, contribute to, lead to, follow after, result of,* ***changes****, determine, affect,* and *is responsible for.*

Mathematicians use these contextual terms to describe associations since causality is not a mathematical concept.

Users interpret these contextual terms as being causal when causality is the issue. They expect that mere associations would be referenced explicitly in such situations.

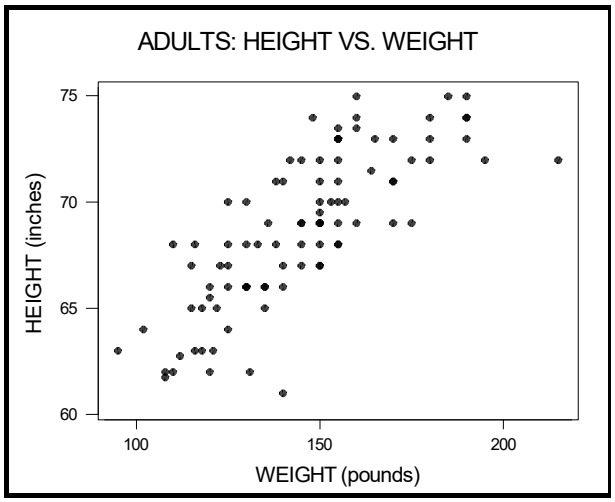
Statisticians have not yet formulated any guidelines on the ethical use of words to distinguish association from causation in non-experimental studies.

Moral:

Users must treat all these contextual words as causally ambiguous even when they are used by statisticians or social scientists!

Comparisons

In describing comparisons in scatter-plots, one can use different forms of grammar:



A. Pure association: People who are above average in weight tend to be above average in height. People who are heavier tend to be taller.

B. Technical association: As weight increases, height [tends to] increases. Weight is correlated with height.

C. Pure causation: As a person gets heavier, they [tend to] get taller.

X Ambiguous: As people get heavier, they [tend to] get taller. Increasing weight, increases height.

Moral:

The truth of A (association) or of B does not guarantee the truth of C (causation).

Change

A "change" can be a physical activity in reality or a mental activity: a change in focus. But mental constructs are often described using physical equivalents, so can be difficult to tell if a change is physical or mental.

Consider the statement, "As weight increases, height increases." This can be interpreted two ways:

- Mental: *As weight increases* [by shifting our focus from people who are below average in weight to people who are above average], *height increases*.
- Physical: *As weight increases* [for each person who eats more], *height increases*.

Mathematicians typically interpret change as being mental: a change in focus on the value of a variable (e.g., just association).

Non-quantitative people often interpret change as being physical: a change in an entity (e.g., a statement of causation).

Moral:

In reading statements of comparison or change written by mathematically-trained statisticians or social scientists, readers must be alert to inadvertently assuming that a change is physical when the author was viewing it as purely mental.

Attributable To

Epidemiology uses the phrase "attributable to" to identify an association. Non-professionals may view this as meaning 'due to' or 'probably caused by". The following is an example of the transition from pure association to pure causation.

Suppose that lung cancer deaths are five times as likely among smokers as among non-smokers, and that 100,000 smokers die of lung cancer each year.

- Pure association: The **risk** of lung cancer death is 5 times as great among smokers as among non-smokers. The **relative risk** of lung cancer death is 5 for smokers.
- Technical association: 80% [(RR-1)/RR] of the lung-cancer deaths among smokers are **attributed to** smoking.
- Ambiguous: 80% of the lung-cancer deaths among smokers are **due to** smoking. 80,000 smokers died **from** (**because of)** cigarette smoke. Smoking is **directly linked** with 80,000 lung-cancer deaths.
- Pure causation: Smoking **caused** 80,000 lung-cancer deaths.

Moral:

Associations are seen; causes are inferred. Words or phrases used by some to indicate association may be mistakenly taken by others as indicating causation.