

Statistical literacy focuses on the use of ordinary English to describe statistical ideas.

According to the McKenzie (2004) survey, statistical educators see the association-causation distinction as one of the two most important ideas in statistics. The problem is that ordinary English is often ambiguous when making this distinction.

The goal of this survey is to see how math/stat educators view various statements involving association and causation.

This survey involves 30 multiple-choice scenarios. Each scenario has the same four choices:

- a. Association: the statement indicates a pure association – a change in focus.
- b. Between: the statement is "between" pure association" and "pure causation".  
Technically it may state an association, but generally it demotes/connotes causation.
- c. Causation: Statement clearly states causation – an internal change within a subject.
- d. Don't know.

Instructions: Choose the category that you think is MOST LIKELY. Use "Don't know" only to indicate there is no modal category

Samples:

Classify each of the following statements under one of these four categories:

Choices: a. Association    b. Between or ambiguous    c. Causation    d. Not sure

- S1:    X is strongly associated with Y.  
Some may choose 'a' saying 'associated' always indicates an association.  
Some may choose 'b' saying 'strongly associated' implies causation.
- S2:    X may be a cause of Y.  
Some may chose 'c' saying 'cause' always indicates 'causation'.  
Some may chose 'b' saying 'may be' implies that 'X may be associated with Y'
- S3:    X is causally linked to Y  
Some may choose 'a' saying that 'linked' technically implies an association.  
Some may choose c saying that 'cause' trumps 'linked'  
Some may choose b saying that this mixture is ambiguous.

As you can see, this is a survey of what you think when viewed from two perspectives:

- 1) your view as a professional in a discipline with a well-understood technical vocabulary
- 2) your understanding of how non-professionals view these same words or phrases.

Question: What other statements should be included in this survey? What should be excluded?

McKenzie, John (2004). . Conveying the Core Concepts. 2004 ASA *Proceedings of the Section on Statistical Education*. P 2755-2757. Copy at [www.statlit.org/pdf/2004McKenzieASA.pdf](http://www.statlit.org/pdf/2004McKenzieASA.pdf)

Instructions: Classify each of the following statements under one of these four categories:

- a. Association    b. Between or ambiguous    c. Causation    d. Not sure

**Part 1: X must be discrete (usually binary); Y is continuous.**

Continuous variables include "likely to [be] Y" or "the \_\_\_\_\_ of Y", where the space is a named-ratio such as chance, risk, odds, likelihood, probability, rate, percentage, fraction or percent.

- Q1. X1 is or has more Y than X2. Y can be a variable or an attribute: durable, reliable, etc.  
Q2. X1 is <comparative> than X2. Comparatives include taller, stronger, safer, cheaper, etc.  
Q3. Y occurs after X.  
Q4. X stops or prevents Y.  
Q5. X protects against Y. (X is a Y protector)  
Q6. Those who do X have more Y [than those who don't do X]  
Q7. X has more Y than non-X. X prefer Y. X voted [more] for Y [more than non-X]  
Q8. Those who do X have more Y than they would have if they hadn't done X.

**Part 2: X and Y are both continuous variables.**

- Q9. As X increases, Y increases. If X increases, then Y increases. If X is X<sub>0</sub>, then Y is Y<sub>0</sub>.  
Q10. The more X [is], the more Y [is]. The bigger [that] X [is], the bigger [that] Y [is].  
Q11. People with more X have more Y [then people with less X].  
Q12. The more X you do/have, the more Y you will get/have or can expect.  
Q13. Controlling for X reduces or eliminates P% of the variation in Y  
Q14. P% of the variation in Y is attributable to X. N instances of Y are attributable to X.  
Q15. X explains P% of [the variation in] Y. N instances of Y are explained by X.  
Q16. P% of [the variation in] Y is due to X. N instances of Y are due to X.  
Q17. P% of [the variation in] Y occurred because of X. N instances of Y occurred because of X.

**Part 3: X and Y can be either binary or continuous variables.**

- Q18. X causes Y.  
Q19. X helps or may help cause Y.  
Q20. X is a causal factor in Y.  
Q21. X is causally connected to Y.  
Q22. Y is an effect or consequence of X.  
Q23. X results in Y.  
Q24. X cuts/reduces Y.  
Q25. X is a causal factor in Y.  
Q26. Y exists or occurs from, due to or because of X.  
Q27. X ups/increases Y.  
Q28. X leads to Y.  
Q29. X is linked or tied to Y.  
Q30. X is to blame for Y.