Statistical Literacy: A Short Introduction

1. What is statistical literacy?

As defined by the W. M. Keck Statistical Literacy Project, statistical literacy

- is critical thinking about numbers, about statistics used as evidence in arguments.
- is the ability to read and interpret number in statements, surveys, tables and graphs.
- studies how statistical associations are used as evidence for causal connections.

2. Why is statistical literacy important?

"The world of the twenty-first century is a world awash in numbers" (Steen, 2005).

Numbers are not only important because they are pervasive; they are pervasive because they are important. It is because numbers have both the power to influence and the power to inform that we need to educate citizens to attend to numbers, to understand them, and to think thoughtfully and critically about them. (Lutsky, 2008)

Statistical literacy is an important asset for full citizenship Peter Campos

Widespread statistical illiteracy ... is cause for immediate concern." Charles Murray, 2010

3. Who needs statistical literacy?

Statistical literacy is needed by data consumers – anyone who tries to evaluate numerical information.

Statistical literacy is needed most by journalists, policy analysts, decision makers and by political, economic and social leaders, but most of all by the citizens of a modern democracy.

4. What should a statistically-literate person be able to do?

Statistical literates should be able to evaluate number-based claims in the media. Consider these newspaper headlines:

- *Soft Drinks Could Boost Pancreatic Cancer Risk.*
- *Absent Dads cause Earlier Puberty in Girls.*
- *Weddings boost mood*
- *Shooter video games can improve decision making*
- *Female Viagra Doesn't Improve Sexual Desire*

A statistically-literate reader can tell that the first two claims will have much weaker support because the outcome is not repeatable for a given person: you only get cancer once; you only go through puberty once. Comparing different people weakens the argument.

They can tell that the last three claims have stronger support since the outcomes can be measured before and after the event or condition in question. But only the last one can have fairly strong support. The last study is the one in which the outcome is repeatable AND the subject can be assigned – unknowingly to either get Viagra or to get a placebo.

Statistical literacy is like speed reading or speed dating. You get more information faster.

Statistical Literacy: A Short Introduction

Statistical literates should be able to read and critically evaluate statistically-based arguments involving public policy such as *The Bell Curve* (Herrnstein and Murray), *More Guns -- Less Crime* (Lott), *Population and Development* (Simon), *The Tyranny of Numbers* (Eberstadt), *Economics and Politics of Race* (Sowell) and *Criminal Justice* (Bidinotto).

Statistical literates can spot the difference between association and causation. Consider this set of newspaper titles – all involving the same underlying story:

- 45,000 deaths *attributable to* uninsurance (PNHP 9/17/09)
- 45,000 American deaths *associated with* lack of insurance (CNN 9/18/09)
- Study: Uninsured Americans Have 40 Percent *Higher* Death Risk (Ivanhoe, 9/18/09)
- Study: 45,000 Uninsured *Die* a Year (CBS News, 9/17/09)
- No health coverage *tied to* 45,000 deaths a year (Rueters MSNBC 9/17/09)
- Lack of insurance *linked to* 45,000 deaths (White Coat News, 9/17/09)
- Study *links* 45,000 U.S. deaths to lack of insurance (Reuters, 9/17/09)
- Study: 45,000 U.S. Deaths *From* Lack of Insurance (MoneyNews 9/17/09)
- One American dies every 12 minutes *due to* no health insurance (blog DR 9/17/09)
- 45,000 Americans die … *because of* lack of health insurance (blog MyDD 9/17/09)
- Lack of Health Insurance *Kills* 45,000 a Year (Health Insurance com, Inst.)
- Lack of Health Insurance *cause* 44789 deaths in United States every year (blog)
- Lack of insurance *to blame for* almost 45,000 deaths: Study (HealthDay 9/17/09)

A statistically-literate reader knows that words like "kills", "causes" and "blame" make "causal claims" that are much more disputable than "association claims" involving words like "attributed to", "associated with", "tied to", "linked to" or "due to".

A statistically-literate reader can spot obvious errors in news stories. Consider these:

- *Racial Imbalance Persists at Elite Public Schools* New York Times 11/08/2008. "at Stuyvesant...2% of blacks, 3% of Hispanics, 24% of whites and 72% of Asians were accepted." The 100% total is suspicious. These are not parts of a pie so there is no reason for that total. The report should have said "among those accepted, 2% are blacks, 3% are Hispanics, 24% are whites and 72% are Asians." Since these are parts of the same "pie", they should total 100%.
- *Study says too much candy could lead to prison.* AP 9/30/2009. "Of the children who ate candies or chocolates daily at age 10, 69 percent were later arrested for a violent offense by the age of 34." This is an incredible statistic. Do you believe eating candy daily can predict criminal behaviour 20 years in advance? No! The truth: "69% of respondents who were violent criminals by the age of 34 years reported that they ate confectionary nearly every day during childhood." The AP reversed the order: "69% of daily candy-eating kids became violent criminals by 34" is very different from "69% of violent criminals by age 34 had been daily candy-eaters as kids".

Statistically-literate readers look for weasel words: count words that imply much but assert very little – words like *many*, *some* or *few*, *lots* or *little*, *high* or *low*, *often* or *seldom*.

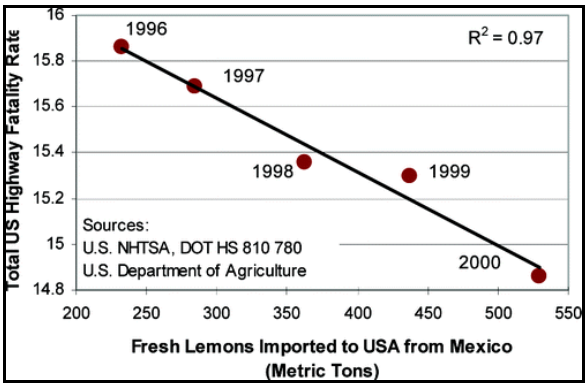
- *Many teens share prescription drugs. Some elderly get futile care.*
- *Adult video gamers often overweight, depressed. Older drivers in fewer crashes.*

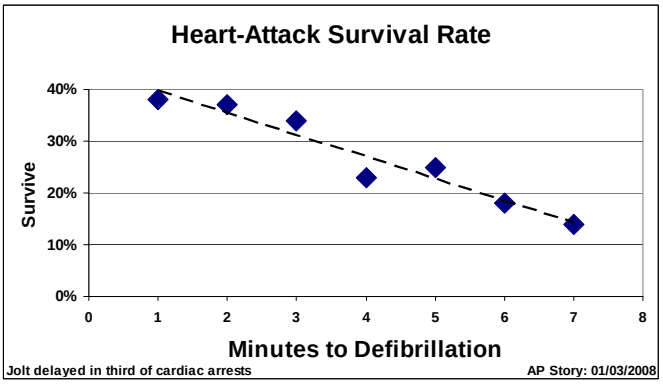
Statistically-literate readers look out for ideas that are vague. Consider these headlines:

- *High exposure to BPA linked to low sperm count.* How low is *low*? By choosing a higher sperm count cutoff, the number of men with low sperm count is increased.
- *Too much TV psychologically harms children*. Exactly how much TV is *too much*? What did they consider *harm*? What did they consider *psychological harm*?

Statistical Literacy: A Short Introduction

A statistically-literate reader has an idea of when an association is not likely to be causal.
Consider the traffic fatality-rate graph on the left: As the amount of lemons imported from Mexico increased, the US traffic fatality rate decreased. The lack of a plausible mechanism and the small effect size (a 6% drop from 15.8 to 14.8) all but invite alternate explanations.





A statistically-literate reader has an idea of when association is likely to be causal.
Consider the heart-attack survival rate graph on the right: As the time to defibrillation after a heart attack increased, the survival rate decreased. The presence of a plausible mechanism (lack of oxygenated blood causes death) and the large effect size (a 62% reduction from 40% to 15%) ward off alternate explanations.

A statistically-literate reader knows the difference between *frequently* and *likely*.

- Car most *frequently* stolen: Honda Civic
- Car most *likely to be* stolen: Cadillac Escalade

Frequently is a count. Honda Civics are common so the number stolen is higher. Here *likely* is a rate per car. Cadillac Escalades are less common so the theft-rate is higher.

A statistical-literate knows the difference between real statistics and speculative statistics.
Which counts are real: deaths due to poisoning or drownings versus deaths due to obesity, radon or second-hand smoke? Answer: the former (corner-certified); the latter are all speculative.

A statistical-literate knows how to read statements involving rates and percentages.
Do these statements say the same thing?

- Percentage of women who smoke vs. percentage of smokers who are women.
- Death rate of men vs. male rate of death.

In both cases, the answer is "No". How about these statements? Here the answer is "Yes."

- Percentage of women who smoke vs. percentage of smokers among women.

A statistical-literate knows how to interpret different definitions of a group.
Given the following two definitions, which group is bigger? Answer: the latter.

- Bullying: the use or threat of force vs. the use or threat of force, or rumors or gossip

The latter. The additional groups added with an "or" makes the latter definition less restrictive.

Given the following two definitions, which rate is bigger?

- Birth rate among women vs. Birth rate among women age 15-44.

The latter. While *women* is bigger than *women age 15-44*, they are the denominators. Both definitions have the same numerator (births) so the smaller denominator gives the higher rate.

Statistical Literacy: References

5. Where can I learn more about the W. M. Keck Statistical Literacy Project?

The foregoing is the tip of the iceberg. Statistical literacy is a separate discipline with its own math and statistics, its own goals and standards. Here are some articles by the project Director.

- Statistical Literacy and Liberal Education at Augsburg College, 2004 AACU Peer Review
- Statistical Literacy: Thinking Critically About Statistics, 1999 APDU *Of Significance*
- Statistical Prevarication: Telling Half-Truths Using Statistics. 2005 IASE Sydney.
- Association-Causation Problems in News Stories. 2010 ICOTS Ljubljana.
- The Social Construction of Rankings. 2010 ASA Proceedings Vancouver
- Percentage Graphs in USA Today Snapshots Online, 2006 ASA Proceedings
- Confound those Speculative Statistics, 2009 ASA Proceedings
- Presenting Confounding and Standardization Graphically 2006 ASA *STATS Magazine*.
- Statistical Literacy and Mathematical Thinking. 2000 ICME Tokyo, Japan.
- Quantitative Literacy and School Mathematics: Percentages and Fractions in *Calculation vs. Context: Quantitative Literacy And Its Implications for Teacher Education.* MAA 2008.
- Information Literacy, Statistical Literacy and Data Literacy, 2004 IASSIST IQ Journal.

For the most detailed presentation, see Assessing Statistical Literacy: Take CARE in *Assessment Methods in Statistical Education: An International Perspective*, Wiley 2010. Excerpts

RELATED ARTICLES:

Grawe, Nathan (2010). Integration with Writing Programs: A Strategy for Quantitative Reasoning Program Development. *Numeracy* 2,2.

Isaacson, Marc (2005). Statistical Literacy: An Online Course at Capella University. ASA Proceedings of Section on Statistical Education.

Lutsky, Neil (2008). Arguing with Numbers: Teaching Quantitative Reasoning through Argument and Writing in *Calculation vs. Context*.

Schmit, John (2010). Teaching Statistical Literacy as a Quantitative Rhetoric Course. ASA Proceedings of Section on Statistical Education.

TEACHER RESOURCES:

Statistical literacy 2007 brochure, 2008 brochure, course outcomes, textbook, Moodle exercises online syllabus and website.

Milo Schield *is the Director of the W. M. Keck Statistical Literacy Project at Augsburg College, Vice President of the National Numeracy Network, US Representative to the International Statistical Literacy Project and webmaster of www.StatLit.org (Google-rated as the #1 site for statistical literacy since 2004). He directed a 2009 MAA survey finding that 19% of US four-year colleges responding offer a course they call statistical literacy.*

Milo has taught critical thinking, statistics and statistical literacy for more than a decade. He has given talks on statistical literacy in 14 countries, has written more than 50 papers on statistical literacy and is currently working on a statistical literacy textbook.