UNINTENTIONAL LIES IN THE MEDIA: DON'T BLAME JOURNALISTS FOR WHAT WE DON'T TEACH

Jessica Utts
Department of Statistics, University of California, United States of America
jutts@uci.edu

It's easy to find misleading and even harmful reporting of statistical results. For example, a 2008 study titled "You Are What Your Mother Eats," asserted that children born to mothers who eat breakfast cereal are more likely to be boys than are children born to mothers who do not eat breakfast cereal. A 2009 analysis by statistician Stan Young and colleagues showed that the result was almost surely a false positive, but by then the study had gained widespread media attention. Many students who take introductory statistics come away from the course able to compute a standard deviation, yet unable to spot an egregious example of poor statistical reporting such as the one illustrated by this example. We are doing an inadequate job of educating the next generation of medical researchers, journal referees, policy-makers, journalists, and so on. I will discuss some ways we can do a better job.

INTRODUCTION

The majority of students in an introductory statistics course in college will never take another statistics course. Yet they will need to use statistical information to make decisions throughout their lives. Some of them will enter professions such as journalism or medicine that require them to inform others based on statistical information. Therefore, those of us who teach introductory statistics courses have an opportunity to change peoples' lives, and the lives of those with whom they will come in contact, for years to come. But most of us are not taking full advantage of this opportunity. Although there have been positive changes in the way introductory statistics courses are taught, there is so much more we could do. We can't do it unless we are willing to give something up, but the trade-off is worth making.

It is easy to find examples of misleading reporting of statistical studies and to criticize the media for them. But I speculate that most of the journalists who wrote those stories took an introductory statistics class as part of their education. What were they taught? Most likely, they came out of the course knowing how to construct a histogram and compute a standard deviation, but never being exposed to issues like the problems with multiple testing, how to know when a causal connection can be made, why it is important to consider baseline risk when assessing relative risk, and a variety of other topics that are critically important to decision-making in daily life. We are doing the world a disservice by clinging to the teaching of topics that only a small subset of our students will ever need to know, while ignoring topics that would benefit the vast majority of them. It is not enough to separate "statistical literacy" into a course of its own. Every student who takes one statistics course should first and foremost be learning the statistical ideas they need to make informed decisions in daily life.

SOME TOPICS WORTH LEARNING

There are certain statistical ideas that all educated citizens should understand. In Utts (2003) I discussed seven such topics, and in this paper I repeat one of them (because it is so important and so often misunderstood), and I introduce three additional ones. For some of these topics students will need to understand very basic probability, for others they will need to understand the fundamentals of hypothesis testing. But some of these topics require no background beyond what is covered in primary and secondary school.

What Educated Citizens Should Know about Statistics and Probability

The topics covered in Utts (2003) include:

- Unwarranted causal connections based on observational studies,
- Statistical significance versus practical importance, especially for large studies,

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

- Lack of statistical significance does not mean there is no population effect, especially for small studies,
- Potential forms of bias in surveys and other studies,
- Unusual coincidences are highly likely to happen someday, somewhere, to someone, by chance,
- Conditional probabilities in one direction should not be confused with conditional probabilities in the other direction, such as confusing the (low) probability of having a disease given a positive test with the (high) probability of a positive test given that one has a disease,
- "Normal" is not the same thing as "average," for instance when claiming that rainfall in a given year is way above or below "normal." Variability is natural and ubiquitous.

The topics covered in this discussion include:

- Unwarranted causal connections based on observational studies (again),
- Multiple testing and the consequences for statistically significant findings,
- Absolute risk, relative risk, personal risk and risk trade-offs,
- Some psychological influences on probability assessments.

When teaching these topics in the classroom, case studies can be an excellent mechanism for illustrating more than one of them at a time. Students are more likely to remember the cautions when they are presented with concrete examples than when presented with abstractions. In the hope that the same is true of teachers, this paper will use examples to illustrate the above topics.

CEREAL AND SEX

A headline in *New Scientist* on April 23, 2008 claimed "Breakfast cereals boost chances of conceiving boys." The article claimed that "A survey of 740 pregnant women found that boys were slightly more likely if a women had high energy intake prior to conception, and that the individual food with most impact was breakfast cereal." Additional details in the article noted that "When the researchers divided the women into groups with high, medium and low intake of energy, they found that 56% of women in the high-energy group had boys, compared with 45% in lowest group." It then went on to report that cereal was the only specific food that made a difference, "Of women eating cereals daily, 59% had boys, compared with only 43% who bore boys in the group eating less than a bowlful per week (Coghlan, 2008)."

There were numerous media reports about this study in print, internet and on television news conveying the same message, that eating cereal increases the chances of having a boy. The original study was published in the prestigious *Proceedings of the Royal Society B* (Mathews et al., 2008). In the body of the paper the authors of the original study never said that eating cereal actually *caused* women to be more likely to have a boy. Instead, they used phrases such as "The consumption of breakfast cereals was also strongly associated with having a male infant (p. 1665)." But their title certainly suggested a causal connection (see references), and the media coverage and interpretation of the results accentuated the causal connection between diet and birth outcome.

Mistake 1: Observational Studies and the Implication of Causation

Probably the most common mistake made in the media regarding statistical studies is to conclude that an association is causal when that conclusion is not warranted because the results are based on an observational study. The study of cereal and birth outcome is a classic example. The women in the study were not randomly assigned to eat certain foods. When they were recruited to the study at approximately 14 weeks of gestation they were asked to give a retrospective account of their typical diet for the year preceding conception, and at about 28 weeks of gestation they were asked to give an account of their typical diet during the pregnancy.

The authors of the original study contributed to the problem because they gave what they considered to be a plausible explanation for the connection between a high-energy diet and an increased likelihood of having a male child: "Our results support hypotheses predicting investment in costly male offspring when resources are plentiful (Mathews et al., 2008, p. 1661)."

If there are factors that contribute to increased likelihood of conceiving a baby of one sex or the other, they have yet to be found. So it is difficult to know what the possible confounding variables might be in this study. But we know that diet is linked to many other lifestyle variables, and possibly to genetic variables, so if there really are factors that change the likely sex ratio it is quite reasonable that they would have some relationship with diet, without necessarily having a causal connection.

As may be obvious by now to the astute reader, in this example there is a bigger statistical problem. It was identified by Young et al (2009) in an article published in the same journal, nine months after the original study was published. We turn to that issue now.

Mistake 2: Selective Reporting of Multiple Hypothesis Tests

The women is this study used a food-frequency questionnaire to report their consumption of 133 different foods for each of the two time periods (preconception and during pregnancy). Young et al. (2009) requested the dataset and received information on 132 foods for two time periods, resulting in the possibility of 264 hypothesis tests. Each test would look for a difference in the proportion of male and female births for mothers who had high and low consumption of the food. Young et al. computed the 264 actual *p*-values and plotted them against the expected order statistics for a sample of 264 observations from a uniform (0,1) distribution. The *p*-values lined up almost perfectly with the expected order statistics. As an additional analysis Young et al. used simulations to find multiplicity-adjusted *p*-values. While the original *p*-value for the relationship between cereal consumption and birth outcome was 0.0034, the multiplicity-adjusted *p*-value was 0.2813. They concluded that "the claimed effects are readily explainable by chance (Young et al., 2009, p. 1211)."

Ioannidis (2005) looked at the replication rate for 45 high-impact medical studies for which the treatment being studied was found to be effective. The studies were chosen because they were published in specific high-impact medical journals between 1990 and 2003 and had been cited at least 1000 times. Six of them were observational studies and the remaining 39 were randomized controlled trials. There were attempted replications of similar or larger sample size and similar or better controlled designs for 34 of these 45 studies. Replications for five of the six observational studies found smaller or reversed effects, compared with nine of the 39 randomized controlled trials. Ioannidis speculated on various possible explanations for the lack of replication (in both types of studies), and multiple testing was one of the possible explanations.

AVOIDING RISK MAY PUT YOU IN DANGER

Gigerenzer et al. (2008) discuss illustrative examples of what they refer to as "collective statistical illiteracy," and make recommendations for what statistics educators can do to help reduce its prevalence in society. One striking example of the harm statistical illiteracy can cause was the issuance of a warning from the UK Committee on Safety of Medicines in October 1995, which included letters to over 190,000 medical practitioners, pharmacists and public health officials. The warning, which was also presented as an emergency announcement to the media, was that "third-generation oral contraceptive pills increased the risk of potentially life-threatening blood clots in the legs or lungs twofold—that is, by 100%." The announcement caused massive anxiety and sales of contraceptive pills plummeted.

Mistake 3: Ignoring base rates, personal risk and risk trade-off

Learning that something you are doing voluntarily is doubling your chance of a life-threatening complication may scare you into stopping that exceedingly dangerous activity. But what about learning that something you are doing is increasing your risk of an adverse reaction from 1 in 7000 to 2 in 7000? You would probably be much less likely to panic if the statistical information were to be presented in that form, and that is the magnitude of the risk in this situation.

You might think that even one additional threatened life in 7000 is not worth taking a risk, but Gigerenzer et al explain that the risk posed by not taking contraceptive pills could be much greater. They quoted the following probable consequences of the substantial drop in pill sales:

- Estimated increase of 13,000 abortions the following year in England and Wales,

- Estimated similar increase in births, especially large for teenagers,
- Additional $70 million cost to the National Health Service for the abortions alone,
- Additional deaths and complications from abortions and births most likely far exceeded the pill risk.

It is crucial that we teach students to distinguish between relative risk and absolute risk. Framing risk in terms of number of additional deaths (or other consequences) per 1000 or 100,000 individuals will make much more sense to people than framing risk in terms of proportions. We must also teach people to think about the trade-off in risks if they change their behavior and may be substituting one risk for another. For example, caffeine has been shown to improve alertness for a few hours after ingestion, so giving up one's morning coffee may increase the likelihood of an accident on the way to work. Finally, we must teach people to consider their own individual risks. For example, if an intervention is found to increase the risk of breast cancer but decrease the risk of heart disease, then individuals must assess not only the baseline risks, but their own genetic and lifestyle risks for the two diseases before deciding whether or not to consider the therapy.

WHEN INTUITION AND PROBABILITY DISAGREE

Psychologists have shown that people have poor intuition about probability, and even statistically educated people can be misled if information is not presented in terms of probability. This reliance on intuition instead of analysis can be used to fool people on juries, in financial decisions and in many other aspects of life.

Mistake 4: Basing judgments on intuition instead of calculated probabilities

William James (1890) is credited with first suggesting that humans have both an intuitive mind and an analytic mind, and that the two minds process information in different ways. Recent research by Alter et al. (2007) has shown that it may be possible to create conditions in which decisions that initially would be made by the intuitive mind can be shifted to be made by the analytic mind instead. For example, psychologists know that people place higher probability on events they can readily bring to mind (the "representativeness heuristic" described in more detail below), and also that people tend to overrate their own abilities in tasks like driving (Plous, 1993). Thus, someone who intuitively feels that it is safer to drive than to fly to a destination most likely is placing higher than warranted probability on the likelihood of an air disaster (based on the extensive media coverage these receive when they happen) and lower than warranted likelihood on their own chances of a disastrous car accident, feeling that their own exquisite driving skills reduce their likelihood of an accident to near zero. Helping such a person perform an analytical assessment based on actual data, rather than relying on an intuitive assessment, may lead them to the correct conclusion that in most cases flying is actually a safer mode of transportation.

Let's examine a classic example that illustrates how people on a jury can be misled by trusting their intuitive minds instead of engaging their analytical minds. Tversky and Kahneman (1982) presented people with the following scenario:

Imagine you are a member of a jury judging a hit-and-run driving case. A taxi hit a pedestrian one night and fled the scene. The entire case against the taxi company rests on the evidence of one witness, an elderly man who saw the accident from his window some distance away. He says that he saw the pedestrian struck by a blue taxi. In trying to establish her case, the lawyer for the injured pedestrian establishes the following facts:

- *There are only two taxi companies in town, "Blue Cabs" and "Green Cabs." On the night in question, 85 percent of all taxis on the road were green and 15 percent were blue.*
- *The witness has undergone an extensive vision test under conditions similar to those on the night in question, and has demonstrated that he can successfully distinguish a blue taxi from a green taxi 80 percent of the time.*

Most people presented with this scenario placed a high probability on the witness being correct that the guilty taxi was blue. But a probability assessment reveals otherwise. In a quick intuitive assessment of the situation, respondents are likely to confuse the conditional probability that the witness said the taxi was blue given that it was blue (stated as 0.8) with the conditional probability that the taxi was blue given that the witness said it was blue (which must be calculated). A tree diagram can be used to find the correct conditional probability, or (my preference) what I call a "hypothetical hundred thousand table" can be used. The table illustrates how many out of 100,000 cases would fall into each set of possibilities. I prefer to use 100,000 rather than a smaller number like 1000 so that if events have very small probabilities the table still will include only whole numbers. In this scenario, a table of only 100 could be used, but to illustrate the most general procedure I use 100,000. Table 1 shows the hypothetical 100,000 table for this scenario.

Table 1. Hypothetical Hundred Thousand Table for the Taxi Example

	Witness said green	Witness said blue	Total
Taxi was green	68,000	17,000	85,000
Taxi was blue	3,000	12,000	15,000
Total	71,000	29,000	100,000

It is now easy to see that P(Taxi was blue | Witness said blue) = 12000/29000 = 0.41. The intuitive answer, that the probability in question is about 0.80, is clearly wrong. We can help students understand how their intuition can lead them astray by presenting several examples of this sort.

The representativeness heuristic is the name psychologists use for the idea that people give higher probability to descriptions that are representative of how they think the world works, even when an analytic assessment would make it clear that a less representative scenario has higher probability. Following Plous (1993), who originally presented this scenario, in winter quarter 2010 I asked my introductory statistics class which of the following two scenarios would have been more likely during the Cold War:

- An all-out nuclear war between the United States and Russia
- An all-out nuclear war between the United States and Russia in which neither country intends to use nuclear weapons, but both sides are drawn into the conflict by the actions of a country such as Iraq, Libya, Israel, or Pakistan.

Of the 96 respondents, only 32, or 1/3, chose the first option. But the second option is a subset of the first one, so it cannot have a higher probability. When this question was posed to the students they had just studied the laws of probability, and thus they knew that P(A and B) must be less than or equal to P(A). Yet, the majority of them chose the scenario that was more representative of how they could envision a nuclear war happening. This tendency to assign higher probability to a subset of possibilities that seem more representative of how things could happen is called the "conjunction fallacy," because the conjunction of events A and B is assessed to have higher probability than one of the individual events.

Psychologists have identified many more heuristics that illustrate how decisions made by the intuitive mind can conflict with those made using probability rules and the analytical mind. For additional examples, see Plous (1993) or Utts (2005).

As Plous (1993) has noted, one of the most effective ways to help people make better judgments is to have them consider reasons why their initial judgment might be wrong. Helping students understand that the intuitive mind is ill-equipped to deal with judgments that require more analytical analysis is the first step in helping them correct those intuitive judgments.

CONCLUSION

The information explosion of recent years has led to an enormous number of media stories presenting the results of statistical studies in formats accessible to the general public. As statistics educators, we need to do a better job of educating our students to write these stories (as future journalists), to interpret them for decision-making (as future doctors, lawyers and other professionals) and to read them with a critical eye (as future consumers of information). When we have one opportunity to educate students on the important statistical ideas they will encounter in daily life, we should not squander it by insisting on teaching them technical details that they will neither use nor remember. Showing students how to use their analytical minds when the first instinct is to use their intuitive minds is an important lesson that should be conveyed in all introductory statistics classes. This paper has provided several examples of how we can do that, and I encourage readers to find more comprehensive sources for additional ideas and examples.

REFERENCES

Alter, A., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General, 136*(4), 569-576.

Coghlan, A. (2008). Breakfast cereals boost chances of conceiving boys. *New Scientist online* (23 April, 2008). Online: www.newscientist.com/article/dn13754-breakfast-cereals-boost-chances-of-conceiving-boys.html.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2008). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53-96.

Ioannidis, J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*, *294*(2), 218-228.

James, W. (1890) *The principles of psychology*. New York: Henry Holt and Company. Online: http://psychclassics.asu.edu/James/Principles.

Mathew, F., Johnson, P.J., & Neil, A. (2008). You are what your mother eats: evidence for maternal preconception diet influencing foetal sex in humans. *Proceedings of the Royal Society B, 275*, 1661-1668.

Plous, S. (1993). *The psychology of judgment and decision making.* New York: McGraw-Hill

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). New York: Cambridge University Press.

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician, 57*(2), 74-79.

Utts, J. (2005). *Seeing through statistics* (3rd edition). Belont, CA: Thomson Brooks/Cole.

Young, S. S., Bang, H. & Oktay, K. (2009). Cereal-induced gender selection? Most likely a multiple testing false positive. *Proceedings of the Royal Society B, 276*, 1211-1212.