What Educated Citizens Should Know About Statistics and Probability

Jessica UTTS

Much has changed since the widespread introduction of statistics courses into the university curriculum, but the way introductory statistics courses are taught has not kept up with these changes. This article discusses the changes, and the way the introductory syllabus should change to reflect them. In particular, seven ideas are discussed that every student who takes elementary statistics should learn and understand in order to be an educated citizen. Misunderstanding these topics leads to cynicism among the public at best, and misuse of study results by policy-makers, physicians, and others at worst.

KEY WORDS: Coincidences; Practical significance; Statistics education; Statistical literacy; Survey bias.

1. INTRODUCTION

Statistical studies are prominently featured in most major newspapers on a daily or weekly basis, yet most citizens, and even many reporters, do not have the knowledge required to read them critically. When statistics courses were first introduced, they were taken primarily by students who intended to pursue their own research, or were in disciplines that required them to analyze data as part of their training. The focus of those courses was on computation, and little emphasis was placed on how to integrate information from study design to final conclusions in a meaningful way. Much has changed since then, in three ways: the audience, the tools available to students, and the world around us.

At many universities, students in a large proportion of majors are required to take an introductory statistics course. Most of these students will never actually do statistical analyses of their own. Therefore, we should be preparing them to read and understand studies conducted and analyzed by others, published in journals, and reported by the media. Anecdotally, the audience has changed in two other ways as well. First, students in introductory statistics courses seem less adept at quantitative reasoning than in earlier days, perhaps because of the broader representation of majors. And there are more returning students, who may have different interests than traditional college-age students.

There is no question that the tools available for use by the students have changed in recent years. Most students come to college with a sophisticated calculator, at least capable of finding means and standard deviations, and often capable of doing most of the procedures taught in introductory statistics courses. Further, there is universal access to computers, and programs like Excel have standard statistical features. Even the use of statistical software has changed over the past few decades, with programs like Minitab and SPSS being menu-driven, making them easy for novices to learn and use.

The world around us has changed as well. Statistical studies are reported regularly in newspapers and magazines, so students are likely to encounter them on a routine basis. And, for classroom use, there is an abundance of examples available on the Internet through sites like those of the Gallup Organization, *USA Today*, the Bureau of Labor Statistics, and so on. Further, many journal articles are available on-line, so it is easy for instructors and students to find complete examples of the design, implementation, analysis, and conclusions of statistical studies.

The consequence of all of these changes is that there is less need to emphasize calculations, and more need to focus on understanding how statistical studies are conducted and interpreted. Relevant and interesting examples are readily available. Yet many instructors have not made any changes in how they teach introductory statistics.

2. SEVEN IMPORTANT TOPICS

There are of course many important topics that need to be discussed in an elementary statistics course. For this article, I have selected seven topics that I have found to be commonly misunderstood by citizens, including the journalists who present statistical studies to the public. In fact researchers themselves, who present their results in journals and at the scientific meetings from which the journalists cull their stories, misunderstand many of these topics. If all students of introductory statistics understood them, there would be much less confusion and misinterpretation related to statistics and probability and findings based on them. In fact the public is often cynical about statistical studies, because these misunderstandings lead to the appearance of a stream of studies with conflicting results. This is particularly true of medical studies, where the misunderstandings can have serious consequences when neither physicians nor patients can properly interpret the statistical results.

A summary of the seven topics covered in this article is presented first, followed by a more in-depth explanation with examples for each topic:

1. When it can be concluded that a relationship is one of cause and effect, and when it cannot, including the difference between randomized experiments and observational studies.

2. The difference between statistical significance and practical importance, especially when using large sample sizes.

Jessica Utts is Professor, Department of Statistics, University of California, Davis, CA 95616 (E-mail: jmutts@ucdavis.edu). This article is adapted from an invited paper for the Sixth International Conference on Teaching Statistics, held in July 2002, and appears here in revised form with permission from the International Association for Statistical Education. The author thanks the editor, Jim Albert, for inviting this special section on Statistical Literacy, and the editor and three referees for insightful comments on this article.

  DOI: 10.1198/0003130031630

3. The difference between finding "no effect" or "no difference" and finding no statistically significant effect or difference, especially when using small sample sizes.

4. Common sources of bias in surveys and experiments, such as poor wording of questions, volunteer response, and socially desirable answers.

5. The idea that coincidences and seemingly very improbable events are not uncommon because there are so many possibilities.

6. "Confusion of the inverse" in which a conditional probability in one direction is confused with the conditional probability in the other direction.

7. Understanding that variability is natural, and that "normal" is not the same as "average."

3. CAUSE AND EFFECT

Probably the most common misinterpretation of statistical studies in the news is to conclude that when a relationship is statistically significant, a change in an explanatory variable is the *cause* of a change in the response variable. This conclusion is appropriate only under very restricted conditions, such as for large randomized experiments. For single observational studies, it is rarely appropriate to conclude that one variable caused a change in another. Therefore, it is important for students of statistics to understand the distinction between randomized experiments and observational studies, and to understand how the potential for confounding variables limits the conclusions that can be made from observational studies.

As an example of this problem, an article appeared in *USA Today* titled "Prayer can lower blood pressure" (Davis 1998). The article reported on an observational study funded by the United States National Institutes of Health, which followed 2,391 people aged 65 or over for six years. One of the conclusions reported in the article read:

> Attending religious services lowers blood pressure more than tuning into religious TV or radio, a new study says. People who attended a religious service once a week and prayed or studied the Bible once a day were 40% less likely to have high blood pressure than those who don't go to church every week and prayed and studied the Bible less (Davis 1998).

The headline and the displayed quote both indicate that praying and attending religious services actually *causes* blood pressure to be lower. But there is no way to determine a causal relationship based on this study. It could be that people who are healthier are more able to attend religious services, so the causal relationship is the reverse of what is attributed. Or, it could be that people who are more socially inclined are less stressed and thus have lower blood pressure, and are more likely to attend church. There are many other possible confounding variables in this study that could account for the observed relationship. The problem is that readers may mistakenly think that if they alter their behavior with more prayer and church attendance, it will cause their blood pressure to lower.

Another example illustrates that even researchers can make this mistake. An article in *The Sacramento Bee* (Perkins 1999) reported on an observational study of a random sample of more than 6,000 individuals with an average age of 70 when the study began. The study followed them over time and found that a majority, over 70%, of the participants did not lose cognitive functioning over time. One result was quoted as "Those who have diabetes or high levels of arteriosclerosis in combination with a gene for Alzheimer's disease are eight times more likely to show a decline in cognitive function" (Perkins 1999). So far, so good, because the reporter is not implying that the increased risk is causal. However, one of the original researchers (if quoted accurately) was not so careful. The researcher was quoted as follows: "That has implications for prevention, which is good news. If we can prevent arteriosclerosis, we can prevent memory loss over time, and we know how to do that with behavior changes—low-fat diets, weight control, exercise, not smoking, and drug treatments" (Perkins 1999).

In other words, the researcher is assuming that high levels of arteriosclerosis are *causing* the decline in cognitive functioning. But there are many possible confounding variables that may cause both high levels of arteriosclerosis and decline in cognitive functioning, such as genetic disposition, certain viruses, lifestyle choices, and so on.

Resisting the temptation to make a causal conclusion is particularly difficult when a causal conclusion is logical, or when one can think of reasons for how the cause and effect mechanism may work. Therefore, when illustrating this concept for students, it is important to give many examples and to discuss how confounding variables may account for the relationship. Fortunately, examples are easy to find. Most major newspapers and Internet news sites report observational studies several times a week, and they often make a possibly erroneous causal conclusion.

4. STATISTICAL SIGNIFICANCE AND PRACTICAL IMPORTANCE

Students need to understand that a statistically significant finding may not have much *practical* importance. This is especially likely to be a problem when the sample size is large, so it is easy to reject the null hypothesis even if there is a very small effect.

As an example, the *New York Times* ran an article with the title "Sad, Lonely World Discovered in Cyberspace" (Harmon 1998). It said, in part:

> People who spend even a few hours a week online have higher levels of depression and loneliness than they would if they used the computer network less frequently. . . it raises troubling questions about the nature of 'virtual' communication and the disembodied relationships that are often formed in cyberspace (Harmon 1998).

It sounds like the research uncovered a major problem for people who use the Internet frequently. But on closer inspection, the magnitude of the difference was very small. On a scale from 1 (more lonely) to 5, self-reported loneliness decreased from an average of 1.99 to 1.89, and on a scale from 0 (more) to 3 (less), self-reported depression decreased from an average of .73 to .62.

Here is another example of how a very large sample size resulted in a statistically significant difference that seems to be of little practical importance to the general public. The original report was in *Nature* (Weber, Prossinger, and Seidler 1998), and a Reuters article on the Yahoo Health News Web site ran a headline "Spring Birthday Confers Height Advantage" (Feb. 18, 1998). The article described an Austrian study of the heights

of 507,125 military recruits, in which a highly significant difference was found between recruits born in the spring and the fall. The difference in average heights was all of .6 centimeters, or about 1/4 inch. While that may be important to researchers who are studying growth issues, the difference is hardly what most of us would consider to be "a height advantage."

A related common problem is when multiple comparisons or analyses are done, but only those that achieve statistical significance are reported. In most studies a variety of relationships are examined, but only those achieving statistical significance are reported in the media. For instance, a randomized experiment studying the effect of taking aspirin or hormones may examine their relationship with multiple outcomes, such as heart disease, stroke, and various types of cancer. If the researchers have not adjusted for multiple comparisons, it is misleading to focus on the relationships that achieved statistical significance as if those were the only ones tested. Although the multiple analysis problem is not discussed in detail here, it is important to discuss it with students when explaining cautions about interpreting statistical significance.

5. LOW POWER VERSUS NO EFFECT

It is also important for students to understand that sample size plays a large role in whether or not a relationship or difference is statistically significant, and that a finding of "no difference" may simply mean that the study had insufficient power. For instance, suppose a study is done to determine whether more than a majority of a population has a certain opinion, so the test considers $H_0 : p = .5$ versus $H_a : p > .5$. If in fact as much as 60% of the population has that opinion, a sample size of 100 will only have power of .64. In other words, there is still a 36% chance that the null hypothesis will not be rejected. Yet, reporters often make a big deal of the fact that a study has "failed to replicate" an earlier finding, when in reality the magnitude of the effect mimics that of the original study, but the power of the study was too low to detect it as statistically significant.

As an example with important consequences, a February 1993 conference sponsored by the United States National Cancer Institute (NCI) conducted a meta-analysis of eight studies on the effectiveness of mammography as a screening device. The conclusion about women aged 40–49 years was: "For this age group it is clear that in the first 5–7 years after study entry, there is no reduction in mortality from breast cancer that can be attributed to screening" (Fletcher et al. 1993).

The problematic words are that there *is no reduction.* A debate ensued between the NCI and American Cancer Society. Here are two additional quotes that illustrate the problem:

> A spokeswoman for the American Cancer Society's national office said Tuesday that the . . . study would not change the group's recommendation because it was not big enough to draw definite conclusions. The study would have to screen 1 million women to get a certain answer because breast cancer is so uncommon in young women (*San Jose Mercury News*, Nov. 24, 1993).

> Even pooling the data from all eight randomized controlled trials produces insufficient statistical power to indicate presence or absence of benefit from screening. In the eight trials, there were only 167,000 women (30% of the participants) aged 40–49, a number too small to provide a statistically significant result (Sickles and Kopans 1993).

The confidence interval for the relative risk after seven years of follow-up was .85 to 1.39, with a point estimate of 1.08, indicating that there may be a small reduction in mortality for women in this age group, or there may be a slight increase (see Utts 1999, p. 433). The original statement that there was "no reduction in mortality" is dangerously misleading.

The lesson to convey in this context is that students should be wary when they read that a study found no effect or relationship when the researchers expected there to be one. Generally, this conclusion is newsworthy only when it contradicts earlier findings or common wisdom. It is important in such cases to find out the size of the sample, and if possible, to find a confidence interval for the results. If the confidence interval is wide or if it tends to be more to one side of chance than the other, there is reason to suspect that the study may not have had sufficient power to detect a real difference or relationship.

Power is no longer a topic to be avoided in an introductory course because it is easy to find software that can do the calculations, and the concept is no more difficult than the concept of Type 1 and Type 2 errors. Minitab will calculate power for most of the tests taught in an elementary statistics course, and there are Web sites available as well. A good Internet source with links to hundreds of sites for statistical calculations, including power, is http://members.aol.com/johnp71/javastat.html, maintained by John Pezzullo.

6. BIASES IN SURVEYS

There are many different sources through which bias can be introduced into surveys. Some of the more egregious are difficult to detect unless all of the details are understood. For example, a Gallup Poll released on July 9, 1999, based on a random sample of 1,016 U.S. adults, asked two different questions in random order, each of which could be used to report the percentage of people who think creationism should be taught in public schools in the United States. The two questions and the proportion that answered "Favor" were:

Question 1: Do you favor or oppose teaching creationism *ALONG WITH* evolution in public schools? (68% favor).

Question 2: Do you favor or oppose teaching creationism *INSTEAD OF* evolution in public schools? (40% favor).

Notice that depending on one's own opinion, these results could be misused to advantage. Someone in favor of creationism could report that 68% think it should be taught, while someone opposed to creationism could report that only 40% think it should be taught.

It's not just the wording of questions that can cause bias to be introduced. There are many other details of how a survey is done that may seem minor but that can have major consequences. For instance, sometimes the order in which questions are asked can change the results. Clark and Schober (1992, p. 41) reported on a survey that asked the following two questions:

1. How happy are you with life in general?
2. How often do you normally go out on a date? About _____ times a month.

There was almost no relationship between the respondents' answers to the two questions. But when the survey was done

again with Question 2 being asked first, the answers were highly related. Clark and Schober speculated that in that case, respondents interpreted Question 1 to mean "Now, considering what you just told me about dating, how happy are you with life in general?" Respondents naturally think that questions on a survey are supposed to be related, and any issues brought to mind by one question may influence subsequent answers.

There are many other ways in which question wording, question order, method of sample selection and other issues can bias survey results. See Tanur (1992), Utts (1999), or Utts and Heckard (2003) for more discussion and examples.

7. PROBABLE COINCIDENCES

Most people have experienced one or more events in their lives that seem to be improbable coincidences. Some such events are so surprising that they attract media attention, often with estimates of how improbable they are. For instance, Plous (1993) reported a story in which a Mr. and Mrs. Richard Baker left a shopping mall, found what they thought was their car in the parking lot, and drove away. A few minutes later they realized that they had the wrong car. They returned to the parking lot to find the police waiting for them. It turned out that the car they were driving belonged to *another* Mr. Baker, who had the same car, with an identical key! Plous reported that the police estimated the odds at a million to one.

The problem with such stories and computations is that they are based on asking the wrong question. The computation most likely applies to that exact event happening. A more logical question is: What is the probability of that or a similar event happening sometime, somewhere, to someone? In most cases, that probability would be very large.

For instance, I was once on a television talk show about luck with a man who had won the million-dollar New York State lottery twice, and the host of the show thought this demonstrated extraordinary luck. Although it may have been wonderful for that individual, Diaconis and Mosteller (1989) reported that there is about an even chance of the same person winning a state lottery in the United States in a seven-year period. That was precisely the interval between the two wins for this person.

It is not easy to calculate precise probabilities for coincidences, but it is possible to show students calculations that approximate the order of magnitude. For instance, there are many stories about twins raised separately who meet as adults and discover that they have striking characteristics in common. Perhaps their wives or children have the same names, and they drive the same kind of car, and they work in the same profession. As a crude approximation, suppose the probability of a "match" on any given item for two people of the same age and sex is 1/50 and that whether there is a match on one item is independent of whether there is a match on other items. Further, suppose in the course of getting to know each other they discuss 200 items, certainly not an unrealistic number. Then the number of "matches" is a binomial random variable with $n = 200$ and $p = .02$. The expected number of matches is four, and even the probability of 6 or more matches is relatively high, at about .21. But the focus in this kind of encounter is on the striking matches, and not on the many dozens of topics that were discussed but did not match.

Even if an event with extremely low probability of occurrence is reported, remember that there are over six billion people in the world, with many circumstances occurring to each one daily. Therefore, there are surely going to be some that seem incredible. In fact if something has only a one in a million probability of happening to any particular person in a given day, it will happen, on average, to over 6000 people in the world, each day. When the media reports an incredible coincidence it should be viewed from this perspective.

8. CONFUSION OF THE INVERSE

Most teachers of statistics know that probability can be very confusing to students, and that intuition about probability is not very good. Psychologists have identified a version of this problem that leads to important misunderstandings, called "confusion of the inverse." The basic problem is that people confuse the conditional probability $P(A|B)$ with the conditional probability $P(B|A)$.

As an example, Eddy (1982) presented this scenario to 100 physicians:

> One of your patients has a lump in her breast. You are almost certain that it is benign, in fact you would say there is only a 1% chance that it is malignant. But just to be sure, you have the patient undergo a mammogram, a breast X-ray designed to detect cancer.
>
> You know from the medical literature that mammograms are 80% accurate for malignant lumps and 90% accurate for benign lumps. In other words, if the lump is truly malignant, the test results will say that it is malignant 80% of the time and will falsely say it is benign 20% of the time. If the lump is truly benign, the test results will say so 90% of the time and will falsely declare that it is malignant only 10% of the time.
>
> Sadly, the mammogram for your patient is returned with the news that the lump is malignant. What are the chances that it is truly malignant?

Most of the physicians responded with an answer close to 75%. But in fact, given the probabilities presented, the correct answer is only about 7.5%! Eddy reported: "When asked about this, the erring physicians usually report that they assumed that the probability of cancer given that the patient has a positive X-ray was approximately equal to the probability of a positive X-ray in a patient with cancer (1982, p. 254)." In other words, the physicians confused the probability of a positive test given that the woman has cancer with the probability that the woman has cancer *given* that the test was positive.

Most medical tests have low false positive and false negative rates, yet the probability of having the disease, given that a test result is positive, can still be quite low if the initial probability of having the disease is low. In that case, most positive test results will be false positives.

I find that the easiest way to illustrate this concept for students is through what I call a "hypothetical hundred thousand" (Utts and Heckard 2003, p. 228), which is a table showing the theoretical breakdown of results for 100,000 people. Table 1 illustrates the breakdown using the numbers for the example Eddy presented to the physicians. Notice that of the 10,700 patients whose test is malignant, only 800, or about 7.5% actually had a malignancy. Because there were so many more women with benign lumps than malignant lumps, the 10% of them with a false positive test made up the large majority of positive test results.

There are numerous other situations where confusion of the inverse may apply. For example, a study released by the American

Table 1. Breakdown of Actual Status Versus Test Status for a Rare Disease

	Test is malignant	*Test is benign*	*Total*
Actually malignant	800	200	1,000
Actually benign	9,900	89,100	99,000
Total	10,700	89,300	100,000

Automobile Association Foundation for Traffic Safety (Stutts et al. 2001) was widely publicized because it found that only 1.5% of drivers in accidents reported that they were using a cell phone, whereas, for example, 10.9% reported that they were distracted by another occupant in the car. Many media reports concluded that this meant that talking on a cell phone was much less likely to cause an accident than other distractions, like talking with someone in the car or attending to the radio.

But notice that this is confusing two conditional probabilities. The reported proportion of accidents of .015 (1.5%) for which the driver was using a cell phone is an estimate of the probability that a driver was using a cell phone, *given* that he or she had an accident. The probability of interest is the inverse—the probability that a driver will have an accident, given that he or she is using a cell phone. That probability cannot be found from the reported data because it depends on the prevalence of cell phone use. But, it is almost certainly true that many more drivers are talking with other occupants of the car than talking on a cell phone at any given time. This study was criticized for other aspects as well; for an interesting critique see the article by the hosts of the "Car Talk" radio show (Magliozzi and Magliozzi 2001); one of whom (Tom) has a Ph.D. in Management from Boston University and a good understanding of statistics.

9. AVERAGE VERSUS NORMAL

The seventh concept students need to understand is that of natural variability and its role in interpreting what is "normal." Here is a humorous example, described by Utts and Heckard (2003). A company near Davis, California was having an odor problem in its wastewater facility, which they tried to blame on "abnormal" rainfall:

> Last year's severe odor problems were due in part to the extreme weather conditions created in the Woodland area by El Niño [according to a company official]. She said Woodland saw 170 to 180 percent of its normal rainfall. "Excessive rain means the water in the holding ponds takes longer to exit for irrigation, giving it more time to develop an odor (Goldwitz 1998).

The problem with this reasoning is that yearly rainfall is extremely variable. In the Davis, California area, a five-number summary for rainfall in inches, from 1951 to 1997, is 6.1, 12.1, 16.7, 25.4, 37.4. (A five-number summary includes the low, first quartile, median, third quartile, and high values.) The rainfall for the year in question was 29.7 inches, well within the "normal" range. The company official, and the reporter, confused "average" with "normal." This mistake is very common in reports of temperature and rainfall data, as well as in other contexts. The concept of natural variability is so crucial to the understanding of statistical results that it should be reinforced throughout the introductory course.

10. CONCLUSION

The issues discussed in this article constitute one list of common and important misunderstandings in statistics and probability. There are obviously others, but I have found these to be so prevalent that it is likely that millions of people are being misled by them. It is the responsibility of those of us teaching introductory statistics to make sure that our students are not among them.

Many universities now have statistical or numerical literacy courses in addition to the traditional introductory statistics course, and it may be tempting to think that these topics belong in those courses rather than in the traditional courses. But that misses the point. What good is it to know how to carry out a t test if a student can not read a newspaper article and determine that hypothesis testing has been misused?

It is not difficult to incorporate the topics covered in this article into the traditional curriculum, and in fact students enjoy hearing about them if they are presented with good examples. The discussion of topics 2 and 3, on the relationship between statistical significance and sample size, should be part of the discussion of Type 1 and Type 2 errors. Topics 5 and 6 can be incorporated into the syllabus with probability, and in fact make interesting examples of finding probabilities. Topic 7 on natural variability as part of what's normal, can be taught in the early part of the course when discussing averages and measures of variability.

Topic 1, on avoiding implications of cause and effect based on observational studies, and Topic 4 about biases in surveys are the only ones that may require additions to the syllabus. But I think it's important to give at least a brief overview of types of statistical studies and how they are done, so that the data collection process is not a complete mystery for students. One lecture explaining the difference between an observational study and a randomized experiment, and the role of confounding variables in the interpretation of observational studies would do more to prepare students for reading the news than a dozen lectures on statistical inference procedures.

The focus of this article has been on helping students interpret statistical studies. What about students who will eventually carry out their own research and data analysis? I think these ideas are even more important for those students to learn. I serve on many PhD exam committees for students who are doing research across a wide range of disciplines. There are two questions I ask every student. One is to explain the meaning of a p value. The other has to do with replicating a study with an important finding, but using a smaller sample size—the researcher is surprised to find that the replication study was not statistically significant. I ask students to give possible explanations. I am sorry to report that many students have difficulty answering these questions, even when they are told in advance that I'm going to ask them. I will know we have fulfilled our mission of educating the citizenry when any student who has taken a statistics class can answer these questions and similar ones on the topics in this article and related conceptual topics.

REFERENCES

Clark, H. H., and Schober, M. F. (1992), "Asking Questions and Influencing Answers," in *Questions About Questions*, ed. J. M. Tanur, New York: Russell

Sage Foundation, pp. 15–48.

Davis, R. (1998), "Prayer Can Lower Blood Pressure," *USA Today*, August 11, 1D.

Diaconis, P., and Mosteller, F. (1989), "Methods for Studying Coincidences," *Journal of the American Statistical Association*, 84, 853–861.

Eddy, D. M. (1982), "Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities," in *Judgment under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic, and A. Tversky, Cambridge, UK: Cambridge University Press, chap. 18.

Fletcher, S. W., Black, B., Harris, R., Rimer, B. K., and Shapiro, S. (1993), "Report on the International Workshop on Screening for Breast Cancer," *Journal of the National Cancer Institute*, 85, 1644–1656.

Goldwitz, A. (1998), "The Davis Enterprise," March 4, A1.

Harmon, A. (1998), "Sad, Lonely World Discovered in Cyberspace," *New York Times*, August 30, A3.

Magliozzi, T., and Magliozzi, R. (2001), "How AAA's Foundation for Traffic Safety Misused Otherwise Good Data," online at http://www.cartalk.cars.com/About/Drive-Now/aaa.html.

Perkins, K. D. (1999), "Study: Age Doesn't Sap Memory," *Sacramento Bee*, July 7, A1, A10.

Plous, S. (1993), *The Psychology of Judgment and Decision Making*, New York: McGraw Hill.

Stickles, E. A., and Kopans, D. B. (1993), "Deficiencies in the Analysis of Breast Cancer Screening Data," *Journal of the National Cancer Institute*, 85, 1621–1624.

Stutts, J. C., Reinfurt, D. W., Staplin, L., Rodgman, E. A. (2001), "The Role of Driver Distraction in Traffic Crashes," Technical Report; available online at www.aaafoundation.org, May 2001.

Tanur, J. (ed.) (1992), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, New York: Russell Sage Foundation.

Utts, J. M. (1999), *Seeing Through Statistics* (2nd ed.), Belmont, CA: Duxbury Press.

Utts, J. M., and Heckard, R. F. (2003), *Mind On Statistics* (2nd ed.), Belmont, CA: Duxbury Press.

Weber, G. W., Prossinger, H., and Seidler, H. (1998), "Height Depends on Month of Birth," *Nature*, 391, Feb. 19, 754–755.