

THE EDUCATED CITIZEN'S GUIDE TO STATISTICS: Twelve Important Concepts for Statistical Literacy

Concept 1:

Do the results of that study apply to me?

The Basic Idea

The results of a study can be extended to a larger group if the units in the study are representative of the larger group *for the questions asked in the study*. One way to make sure this is true is to use a random sample from the larger group as the units in the study, and this is done commonly for surveys and polls. However, using a random sample is rarely feasible for randomized experiments, and generally volunteers are recruited and used instead. For observational studies, convenience samples often are used, such as students or hospital patients.

Example

A randomized experiment was done in which some of the volunteers were taught to meditate and others were not, and they were studied over a four-month period. The results showed that those who practiced meditation had strengthened immune systems. The participants were right-handed, healthy employees of a biotech company in Madison, Wisconsin. There were male and females in the study. (The handedness factor was important in this study so that the same region of the brain could be measured for all participants. If left-handers had participated, the side of the brain measured would probably have to be reversed.) Would the results apply to you, if you were to learn meditation?

News story: "Mending Mindfully" *Yoga Journal*, Nov 2003, p. 35

Original source: Richard J. Davidson, PhD, Jon Kabat-Zinn, PhD, Jessica Schumacher, MS, Melissa Rosenkranz, BA, Daniel Muller, MD, PhD, Saki F. Santorelli, EdD, Ferris Urbanowski, MA, Anne Harrington, PhD, Katherine Bonus, MA and John F. Sheridan, PhD "Alterations in Brain and Immune Function Produced by Mindfulness Meditation" *Psychosomatic Medicine* 65:564-570 (2003)

Concept 2:

If I change my behavior will it matter?

The Basic Idea

If an observational study finds a relationship between an explanatory variable and an outcome, there is no way to know if the relationship is one of cause and effect. It is quite common for the media to present such results as if there is a causal relationship. The advice will then be that if you change your behavior, or diet, or whatever, you will benefit by the change in outcome. Remember that *a cause and effect conclusion can only be made based on a randomized experiment.*

Example 1

"Attending religious services lowers blood pressure more than tuning into religious TV or radio, a new study says..." *USA Today*

"Weekly attendance at religious services increases the longevity of worshippers, a new study has found, possibly because of a greater inner peace." *Sacramento Bee*

News story: "Churchgoers live longer, study finds" *Sacramento Bee*, December 24, 2001, p. A3, Nancy Weaver Teichert

Original source 10: Oman D, Kurata JH, Strawbridge WJ, Cohen RD, "Religious attendance and cause of death over 31 years," *Intl J Psychiatry Med* 32:69–89, 2002

Example 2

"[The observed relationship between arteriosclerosis and Alzheimers'] 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..."

Note: A **confounding variable** is a variable that *affects the response variable* (e.g. longevity) and also *is related to the explanatory variable* (e.g. church attendance). The effect of a confounding variable on the response variable cannot be separated from the effect of the explanatory variable on the response variable. Thus, with confounding variables present we can't conclude that a change in the explanatory variable will *cause* a change in the response variable.

Concept 3:

Can I believe the results of that poll?

The Basic Idea

There are many ways in which bias can enter surveys and polls, in addition to the method of selecting the sample, covered in Concept 1. Some of the many factors that can lead to bias include the wording of questions, who asks, the order in which questions are asked and whether they are open or closed questions. Unfortunately, news stories based on surveys and polls rarely give you the details necessary to ascertain problems. When you read the results of surveys and polls, try to find a website that has a more complete description. Many polling organizations provide press releases on their websites for a few days after the results are released, so if you act fast you may be able to get these details.

Example (from *Seeing Through Statistics*)

In February 1995, the *Washington Post* added this fictitious question to its weekly poll of 1000 randomly selected respondents: "Some people say the 1975 Public Affairs Act should be repealed. Do you agree or disagree that it should be repealed?" Almost half (43%) of the sample expressed an opinion, with 24% agreeing that it should be repealed and 19% disagreeing. The *Post* then tried another trick that produced even more disturbing results. This time, they polled two separate groups of 500 randomly selected adults. The first group was asked: "President Clinton [a Democrat] said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?" The second group was asked: "The Republicans in Congress said that the 1975 Public Affairs Act should be repealed. Do you agree or disagree?" Respondents were also asked about their party affiliation. Overall, 53% of the respondents expressed an opinion about repealing this fictional act. The results by party affiliation were striking: For the "Clinton" version, 36% of the Democrats but only 16% of the Republicans agreed that the Act should be repealed. For the "Republicans in Congress" version, 36% of the Republicans but only 19% of the Democrats agreed that the act should be repealed.

Concept 4:

Can I be unusual and still be normal?

The Basic Idea

In everyday language "normal" and "average" are often confused. For instance, weather reporters talk about the normal rainfall for a given month of the year, or normal high temperature for a given date. What they really mean is the *average* rainfall or high temperature. (Actually, they use a more complicated method than a simple average, but the result is similar.) In the other direction, we talk about whether someone is of "average intelligence" or gets "average grades." We don't mean that they are *exactly* at the average; we mean that they are in a range that includes the majority of people. This wouldn't be such a problem if people understood natural variability, and that for almost all things we measure, there is a range that would be considered normal, but only one single number that is the average.

Example (from *Mind On Statistics*)

A company (that will remain unnamed) located near Davis, California, was having an odor problem in its wastewater facility. Blame it on the weather: [According to a company official] "Last year's severe odor problems were due in part to the 'extreme weather conditions' created in the Woodland area by El Niño." 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." (Amy Goldwitz, *The Davis Enterprise*, March 4, 1998, p. A1)

Rainfall that year was 29.7 inches. Five number summary for 47 years is 6.1, 12.1, 16.7, 25.4, 37.4

Is it really true that Woodland saw 170 to 180 percent of its normal rainfall?

Concept 5: Relationships are not always what they appear to be!

The Basic Idea

Chapter 11 is devoted to explaining various ways in which relationships can come about, in addition to a direct cause and effect. The most common of these ways is that there is a third variable affecting both the explanatory and response variables. A special case is that the third variable is time, and that both the explanatory and response variables are changing over time. The section on Concept 2 already asked you to explore the role of confounding variables in observational studies. This Concept includes many additional ways in which false relationships may turn up.

Example (from *Seeing Through Statistics*)

The correlation between divorce rate in the United States and the percentage of criminals admitted to state prisons for drug offenses is quite strong, at .67. Based on this correlation, advocates of traditional family values may argue that increased divorce rates have resulted in more drug offenses. But they both simply reflect a trend across time. The correlation relating year and divorce rate is much higher, at .92. Similarly, the correlation between year and percent admitted for drug offenses is .78. Any two variables that have either both increased or both decreased across time will display this kind of correlation.

*Some reasons for observed relationships (from *Seeing Through Statistics*):*

1. The explanatory variable is the direct cause of the response variable.
2. The response variable is causing a change in the explanatory variable.
3. The explanatory variable is a contributing but not sole cause of the response variable.
4. Confounding variables may exist.
5. Both variables may result from a common cause.
6. Both variables are changing over time.
7. The association may be nothing more than coincidence

Concept 6: Combining Groups: Beware of Simpson's Paradox!

The Basic Idea

Combining groups or situations when studying a relationship between two variables can completely change the apparent nature of the relationship. A special case (Simpson's Paradox) is when the change is so extreme that the direction of the relationship actually reverses when the groups are combined. Example 7 in Section 12.4 of the book illustrates this extreme case. When you read about a relationship of importance to you, think about whether the people or objects studied might have included different groups, and if so, find out if the same relationship held for all of them.

Example (from *Seeing Through Statistics*)

A well-known example of Simpson's Paradox, published by Bickel, Hammel, and O'Connell (1975), examined admission rates for men and women who had applied to graduate programs at the University of California at Berkeley. The actual breakdown of data for specific programs is confidential, but the point can be made with similar, hypothetical numbers. For simplicity, we will assume there are only two graduate programs. The figures for acceptance to each program are shown in the following Table.

	Program A			Program B		
	Admit	Deny	Total	Admit	Deny	Total
Men	400	250	650	50	300	350
Women	50	25	75	125	300	425
Total	450	275	725	175	600	775

Program A admitted:

62% of men

67% of women

Program B admitted:

14% of men

29% of women

Overall admitted:

$450/1000 = 45\%$ of men

$175/500 = 35\%$ of women

WHO FARED BETTER? MEN OR WOMEN?

In each program, women. Overall, men! That's the paradox.

Concept 7:

Wow! Was that coincidence really amazing?

The Basic Idea

When a coincidence occurs we tend to focus on the fact that the specific sequence of events had very low probability of occurring. But it is not at all surprising that extraordinary events happen to each of us once in awhile. And it's not at all surprising that events that seem almost impossible, for instance, with probabilities as low as $1/1,000,000$, happen to someone, somewhere once in awhile.

Example

"My next-door neighbor, Mr. George D. Bryson, was making a business trip some years ago from St. Louis to New York. Since this involved weekend travel and he was in no hurry, . . . and since his train went through Louisville, he asked the conductor, after he had boarded the train, whether he might have a stopover in Louisville. "This was possible, and on arrival at Louisville he inquired at the station for the leading hotel. He accordingly went to the Brown Hotel and registered. And then, just as a lark, he stepped up to the mail desk and asked if there was any mail for him. The girl calmly handed him a letter addressed to Mr. George D. Bryson, Room 307, that being the number of the room to which he had just been assigned. It turned out that the preceding resident of Room 307 was another George D. Bryson" (Weaver, 1963, pp. 282-283).

Note that the correct question to ask is *not* how unlikely is it that this would have happened to George D. Bryson, but how unlikely is it that it would happen to someone, somewhere, some day. In fact it would be incredible if no hotel room was ever occupied consecutively by two people with the same name!

Concept 8: Don't get your probabilities backwards!

The Basic Idea

The name given in the book for this idea is "confusion of the inverse;" see Section 18.4, pages 340-342. The idea is that we are interested in the probability of a particular event *given* knowledge of another event. We confuse this with the probability of the second event, *given* knowledge of the first event. The example in the book is that many physicians think that the probability that you have a disease *given* that your test is positive is the same as the probability that you will have a positive test, *given* that you have the disease. These can be quite different.

Example (from *Seeing Through Statistics*)

Suppose a medical test is 80% accurate for malignant lumps and 90% accurate for benign lumps. About 1% of lumps tested are actually malignant. The test comes back showing a malignant lump. What is the probability that it is actually malignant? In a survey of 100 doctors, most thought it was about .75, or 75% (Eddy, 1982). Let's look at a "hypothetical 100,000" table:

	Test malignant	Test benign	Total
Actual Malignant	800	200	1,000
Actual Benign	9,900	89,100	99,000
Total	10,700	89,300	100,000

Since there are about 1% malignancies, out of 100,000 there are 1000. Of the 1000 malignancies, 80% or 800 test malignant. Of the 99,000 benign lumps, 90% or 89,100 test benign. Therefore, there are 10,700 tests showing malignancy. Of the 10,700 tests showing malignancy, $800/10,700 = .075$ (7.5%) are malignant.

The physicians seemed to confuse the probability of a positive test *given* malignancy with the probability of malignancy *given* a positive test.

Concept 9: Knowing when not to trust your intuition.

The Basic Idea

Psychologists have shown that we do not have very good intuition about assessing probabilities in everyday circumstances. It is very easy to be misled unless you understand how you can be manipulated. Think carefully when you have to make a decision based on incomplete information. Make sure you are familiar with the psychological influences on probability discussed in Chapter 17.

Example

A classic example, provided by Kahneman and Tversky (1982, p. 496), was a study in which they presented subjects with the following statement: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. Respondents were then asked which of two statements was more probable:

1. Linda is a bank teller.
2. Linda is a bank teller who is active in the feminist movement.

Kahneman and Tversky report that "in a large sample of statistically naive undergraduates, 86% judged the second statement to be more probable" (1982, p. 496).

What's wrong with that logic?

Concept 10: What's important may not be significant!

The Basic Idea

If a study fails to find a statistically significant difference or relationship, don't interpret that to mean that there isn't an important difference or relationship in the population. Unless the number of individuals in a study is quite large, even a moderate population relationship may not produce data that achieves statistical significance. If possible, examine a confidence interval to find out if it falls mainly to one side or the other of chance. Be particularly wary of studies based on a small number of individuals.

Example (from *Seeing Through Statistics*)

A new story on mammograms reported:

"For [women 40-49] 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..." [underline added]

In fact the relative risk after 7 years of follow-up was 1.08, with 95% confidence interval of 0.85 to 1.39. So, in fact it is unknown if there is an effect, and the *conclusion* of "no reduction in mortality" is dangerously misleading.

Note: Make sure students understand that it is not appropriate to *accept* a null hypothesis, and that sample size is a big factor in determining the *p-value*.

Concept 11: **What's significant may not be important!**

The Basic Idea

Statistical significance is not the same thing as practical significance or importance. Almost any null hypothesis can be rejected if the sample size is large enough. Be careful when you read the results of a large study. Try to find a correlation, relative risk, difference or other summary that will give you some idea of the magnitude of the effect. It's even better if you can find a confidence interval to accompany the summary.

Example:

"Spring birthday conveys height advantage" (*Reuters* article, reporting a study in *Nature*)

The heights of 507,125 military recruits were examined, and the researchers found a highly statistically significant difference in heights for recruits born in spring and fall.

However, the difference in means was only 0.6 centimeters, or about $\frac{1}{4}$ inch! The huge sample size resulted in a very small p -value, even though the difference in means was slight.

Concept 12:

Why does one out of 20 foods seem to cause cancer?

The Basic Idea

Using a level of significance of .05 means that for all tests in which the null hypothesis is true, we will decide to reject the null 5% (.05) of the time by chance. (This is not quite true if the tests use data from the same individuals and are thus not independent, but it will be close.) That means that in a large study for which many relationships are investigated, about 5% (1 in 20) of all of the ones that don't exist will appear as if they do exist. For instance, if many foods are tested for their relationship to a disease, and none are related, about one in 20 will appear to be related. There are sophisticated methods for dealing with this problem, called "multiple comparisons." But when you read reports of studies in the news you would not generally be told if they were used. Therefore, use caution when interpreting results that were clearly part of a larger study. News articles may mention that other things were tested, but will generally focus on the results that were statistically significant and ignore the rest.

Example (from *Mind On Statistics*)

"*Tea Doubles Chance Of Conception*: Tea for two might make three, according to a new study. The investigators asked 187 women, each of whom said they were trying to conceive, to record information concerning their daily dietary intake over a one-year period. According to the researchers, an analysis of the data suggests drinking one-half cup or more of tea daily approximately doubled the odds of conception per cycle, compared with non-tea drinkers. The mechanism behind tea's apparent influence on fertility remains unclear. But the investigators point out that, on average, tea drinking is associated with a 'preventive or healthier' lifestyle, which might encourage conception." (Reuters News, Feb. 27, 1998; study published in the *American Journal of Public Health* (1998), Vol. 88, No. 2: 270 -274.)

Note: The researchers measured intake of six different beverages, and caffeinated tea was the only one that showed a statistically significant relationship with odds of conception. That was not mentioned in the news story.