

Statistics...

The importance of numbers and statistics to the modern world cannot be overstated. In his book *How to Think About Statistics*, author John L. Phillips put it succinctly: "The culture of any industrialized society is suffused with quantitative information. Some quantitative messages are simple and direct; others involve a relatively complicated process of inference. Knowing how to think statistically makes possible the comprehension of both."

Understanding statistics is even more imperative given that numerical results are often used (and misused) to manipulate or distort information. In *Statistical Deception at Work*, John Maura writes, "If you cannot understand simple statistics, you can be fooled by news stories, advertisements and daily encounters with other people. You are likely to be taken in by modern-day medicine men who are out there seeking ways to dupe unsuspecting [individuals] into becoming their agents."

And as Cynthia Crossen writes in *Tainted Truth*, "People know enough to be suspicious of some numbers in some contexts, but we are at the mercy of others. We have little personal experience or knowledge of the topics of much modern research, and the methodologies are incomprehensibly arcane. Nevertheless, we respect numbers, and we cannot help believing them."

The Basics...

Statistics and statistical methods are of two basic types:

- **Descriptive statistics** summarize some facet of a complete population. They are used when an entire population is small or easy enough to measure. For example, the average height or weight of everyone in your family is a descriptive statistic. Because all members of the population are included in the calculation, the result is a totally accurate, and thus completely reliable, measurement.
- **Inferential statistics** are used to predict or infer something about a very large population by measuring samples, or subsets, of that population. This is done when it is virtually impossible, or prohibitively expensive, to obtain data about all members of a particular population.

Many of the statistics we normally come in contact with while reading the paper, watching TV, or talking to colleagues are of the inferential variety. Examples include the number of people projected to carry the HIV virus in 1998, the average growth rate of maple trees, and the odds of incurring a side effect when taking a new drug.

These types of statistics are thus used to make far-reaching policy decisions regarding everything from the number of streetlights needed per city block, to the level of funding allocated to school lunch programs, to the amount of money spent to protect the grizzly bear population of the Western United States. Thus, it is critical that we develop a good understanding of how best to use, and not abuse, inferential statistics.

Key Concepts...

Proportion

The concept of proportion allows us to compare relative differences in size, quantity, etc. between or within samples. The emphasis is on relative, because we don't know the absolute difference or the magnitude of that difference.

Proportion can be measured as a...

...ratio. If there are 10 girls and 5 boys in the choir, the ratio of girls to boys is 2 to 1.

...percentage. The choir can also be described as 67% female and 33% male.

Percentages also play a major role when trying to determine the "odds" that an event might happen. If it is discovered that 7 out of 10 mice are brown, we can infer that a.) 70% of mice are brown, or that b.) there is a 70% chance that the next mouse we see will be brown. This idea leads us nicely into the concept of....

Probability

When it comes to statistics, the term probability is used to describe...

...the likelihood that an event will or will not happen. (There is a 90% probability that it will rain tomorrow.)

...the degree of certainty regarding the relationship of two or more variables. (90% of a tree's growth rate can be explained by the amount of rainfall received during the spring and summer months.)

...the level of confidence that what you think is real actually is real. (We are 90% certain that half of the children that will be born next year will be female.)

Sample Size

Generally speaking, the larger the sample size, the higher the level of probability that the statistics actually mean what they say they mean. For example, using a sample of 10 people to draw inferences about the behaviors of a million people is not a smart thing to do, since there is only a small probability that the sample is representative of the whole population. However, a sample of 300 may be more than adequate to allow for statistically sound inferences.

Watch out when samples that are of reasonable size become subdivided. This situation occurs fairly frequently when analyzing survey results, as there is a tendency to learn as much as possible about very specific (but usually very small) sub-samples. This type of analysis is usually referred to as a "cross-tabulation."

Example: Say that out of 300 people asked about their eating habits, 15 indicate that they are vegetarians. From this we can reasonably infer that about 5% of the population eats no meat. (So far, so good.)

What if we ask those 15 vegetarians to name their favorite sport and 60% say "soccer". Does this mean that most vegetarians prefer soccer? No! We only asked 15 people, a sample size that is far too small to give a reliable result. To make this claim, we would have to start again by asking the question of enough vegetarians (at least 100) that we could be reasonably sure that the answer we received was projectable to all vegetarians.

Randomness

All samples are not created equal. Generally, the fairest way to generate a sample is to do so randomly, letting the laws of probability spin their magic. Random samples of a large enough size will do a surprisingly nice job of modeling a large population.

Sometimes, non-random samples make sense, especially when trying to draw inferences about specific population sub-groups (e.g., Americans of Irish heritage or fruit flies that were born with an extra pair of wings). **But be careful of their use, especially when reviewing advertising claims.** The use of non-random samples to "stack the deck" is a favorite trick of unscrupulous advertisers.

Example: A car manufacturer will claim in big letters that its new model is preferred over a competing model. It's only in the small print that we find out that the study was conducted among a specific sample: First time buyers over the age of 65. Thus, the results are projectable only among the portion of the population that has never bought a car but has probably filed for Social Security.

Reliability

When working with inferential statistics, it is important to remember that the numbers are approximations of what the total population is like, not numbers describing the total population itself (unless, of course, the sample IS the population). They are thus subject to some amount of error, leading to a need to express their degree of reliability.

One way to do so is through the concept of sampling error. For example, if a survey is said to include a sampling error of +/- 3 points, it means that any quoted figure could actually be 3 points above or 3 points below that figure. So if a study infers that 10% of the population has 12 toes, a more accurate description is that "the percent of people with 12 toes falls between 7 and 13%." While this answer is less precise, it is certainly more accurate!

Another way that reliability is presented is as a percentage. A Confidence Level of 95% indicates that there is a 95% probability that the number quoted is in fact the actual number that would be found by studying the entire population. Traditionally, differences that occur at the 95% Confidence Level are considered to be significant, and those at the 99% level are considered to be very significant.

It is very important to know the reliability factor when comparing two different pieces of information.

Let's say that you are conducting a long-term study designed to measure the male/female proportion among silver foxes. If the percentage of male foxes changes from 50 to 52%, should you note an increase? The answer is "yes", but only if differences of 2 or more points are considered meaningful (statistically significant). Otherwise, you can only claim a directional increase, which technically is no increase at all.

This concept of statistical significance is rarely discussed in enough detail when surveys are presented by the media, industry and special interest groups. Since it is only natural for a group to present data in the best possible light, it is extremely important to be able to assess the value of the raw numbers and the findings drawn from them. Statistical significance is an excellent way to do so.

Independence

This is one of the most misunderstood concepts in all of statistics and the reason why otherwise smart people will consistently bet (and lose) on the lottery. Let's start with a coin. If you flip it, the odds of having it land heads-up are 50%. So if you flip the coin and it comes up tails, what are the odds that it will come up heads the next time? It's fairly obvious that the odds remain at 50% because each coin toss is an independent event, unaffected by the previous toss.

The independent nature of coin tosses is easy to understand. But what about the chances of winning while playing the lottery? If the numbers 3,5,7 came up today, does that mean you should do what most people do and not bet on them tomorrow? Not at all! Each day's drawing is an independent event, just like a coin toss. Thus, the odds of drawing 3,5,7 tomorrow are the same as they were today. Remember: Someone is going to win. But the odds that it will be you are exceedingly small.

Absolute vs. Relative Change

Should you be frightened to learn that the number of airplane accidents doubled between 1995 and 1996? Maybe, but you'd probably feel differently if the number went from only 1 to 2 than if it increased from 500 to 1000.

Or, should you be alarmed if the concentration of a certain toxic chemical in the water supply tripled? Probably "no" if it went from 1 part per trillion to 3 parts per trillion, but probably "yes" if it went from 1000 parts per million to 3000 parts per million. Thus, it is always important to view relative change in an absolute context, and vice versa.

Not looking at both relative and absolute differences leaves us ripe for being manipulated. For example, if the incidence of skin cancer in a sample increased from 2% to 3%, optimists might say it only went up by 1% (absolute), while pessimists would argue that it jumped by 50% (relative). Both are right! That's why you have to see the actual before and after percentages, not just the changes.

By the way, relative effects are far more exaggerated at the small end of the spectrum than in the middle: A two percentage point increase from 2 to 4% is a 100% relative gain, but a two percentage

point increase from 50 to 52% is only a 4% relative gain! That's another good reason to always look at the numbers that are changing and not just the changes themselves.

Measurement Tools...

Frequency Distributions

As any carpenter will tell you, it's critical to match the tool to the job. The same goes for statistics. Before you can choose the right way to measure data, you have to examine the way that the data distributes itself. There are three frequency patterns that account for a very large percentage of sample distributions:

- **Normal Distribution**

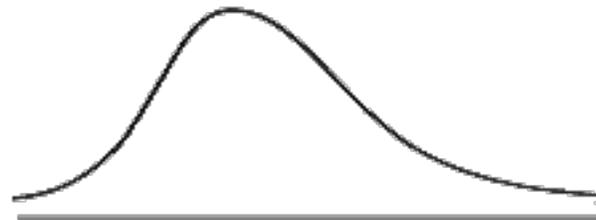
A common frequency distribution that characterizes many human and biological phenomena is called the normal distribution. This type of distribution describes a population in which the majority of score values occur in a central "average" range. The frequency of other scores is the same above and below the average, with fewer and fewer scores occurring as one looks farther from the center. This type of distribution is the familiar bell curve.



Example: Using the test scores from a biology test, the horizontal axis represents the actual scores the students achieved, and the vertical axis represents the frequency with which each particular score occurred. (By the way, the "curve wrecker" who made all of our lives miserable in high school had a score at the far right of the horizontal axis!)

- **Skewed Distribution**

Like a normal distribution, a skewed distribution describes a population in which the majority of score values occur within an average range. However, unlike the normal curve, the minority scores in a skewed curve are unequally divided among the possible higher and lower values outside the average range. The curve that's created looks like a mountain that falls off sharply to one side and more gradually to the other.



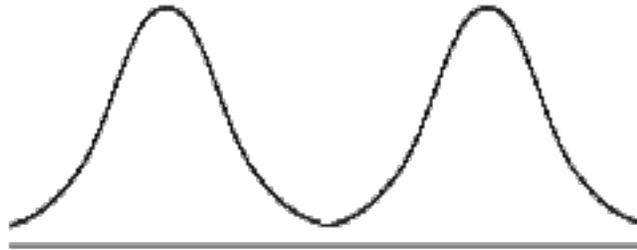
Example: Let's say you decide to analyze sales of ball point pens. You would probably find that most of them were fairly inexpensive, since they were made to be disposable. But there were still quite a few that were fairly costly (silver plated barrels) and a few that were downright expensive (gold barrels). By graphing pens by price, you'd find a frequency distribution that skewed toward the large volume inexpensive items and then tailed off toward the expensive types.

- **Bimodal Distribution**

Some score patterns have two (or more) central clusters, rather than one. When graphed, they look like two bell curves next to each other. In many cases, they are!

Example: Measuring the heights of a sample of people will produce a bimodal distribution -- one cluster for men and one for women. If instead you measure height separately by sex,

the results will be two normal distributions. Thus, a bimodal frequency distribution may be a sign that what is being measured as one group should really be reconsidered as two separate populations.



Averages

Next, we have to examine the way in which scores cluster within these distributions. A mathematical device frequently used to do so is the measure of central tendency, commonly known as the average.

Much of the time we use what's known as the "arithmetic average," which is found by adding up all of the scores in a sample and dividing by the number of scores. The statistical term for this commonly used and widely understood measurement is the mean. However, the mean is only one of five types of average that may need to be calculated and analyzed. Most of the time, we only need to concern ourselves with three:

The mean is the arithmetic average, found by adding up all the scores and dividing by the number of scores. In many cases, the mean is a perfectly suitable measure. But when there are extreme values in the distribution, the mean starts to lose its luster.

The median is the value in a sample that cuts the sample into two equal parts: an equal number of scores are above it, and an equal number are below it. Unlike the mean, the median is more concerned with the total number of scores, rather than with their values.

The mode is the score that occurs most often in a frequency distribution. It is not used much in scientific work, but pops up on TV when you see "person on the street" interviews: ("We asked five people if they like spaghetti. Three said yes, two said no. Thus, the average person likes spaghetti.")

When looking at a normal distribution, all three averages provide the same result: The mean, median and mode all represent the arithmetic average, the score in the middle of the distribution and the score that occurs the most frequently. It's when distributions are not bell shaped that we have to be careful.

Example: Household income distributions tend to be skewed toward the lower end of the scale. However, there are a relatively few households with ENORMOUS incomes. When looking at the

mean, these very few households will produce an average that is rather high and doesn't really present a true picture of the typical household.

The median, on the other hand, will produce a more accurate and somewhat lower income picture, as it examines the distribution of scores without regard to their value. (This is why you generally hear the term median income, rather than mean income.)

Thus, beware of the term "average!" Always try to find out something about the distribution and whether or not average refers to the mean, median or mode. If the distribution is normal (bell curve), the mean is generally the preferred measure, although any of the three will do. If the distribution is skewed, the median is probably the better choice. In the case of skewed distributions, substitution of the mean for the median should make you suspicious.

Standard Deviation

Although it sounds complex, the concept of standard deviation is actually fairly simple: How far from the mean do the scores actually fall? Obviously, the tighter that a group of scores clusters around the mean, the easier it is to make accurate predictions about the value of additional scores. Thus, samples with lower standard deviations provide more reliable and predictable data than samples with higher standard deviations. For this reason, the standard deviation is one of the most important concepts in all of statistics.

Example: Start with two numbers: 1 and 99. Their mean score is 50 (1 + 99 divided by 2). The deviation of each of these numbers from the mean is 49 (1 is 49 less than 50; 99 is 49 greater than 50). This indicates that the range in which you can safely predict where another score might fall is huge: 98, or 49 units below 50 and 49 units above 50.

On the other hand, start with the numbers 49 and 51. Their mean score is also 50, but each of their deviations from it is only 1! Thus, the predictive range is 2 -- 1 unit below 50 to one unit above 50.

In more concrete terms, what if you were William Tell and had to pick one of two archers to shoot the apple on top of your son's head? By holding a contest and reviewing the scores, you note that both archers put 4 out of 6 arrows in the bull's eye. Archer A put the other two at the far end of the target and at opposite sides from each other. Archer B put the other two just outside the bull's eye, right next to each other.

If you're fairly sane, you'd choose Archer B, because her arrows all fell in a nice tight cluster. In statistical terms, her scores would have a lower standard deviation, meaning that her chances of putting an arrow where she wanted to put it were better than what you could expect from Archer A.

Coefficient of Correlation

This measure helps us determine the strength and direction of a relationship between two variables. The Coefficient of Correlation ranges between -1.00 (perfect negative correlation) to 0.00 (no correlation) to +1.00 (perfect positive correlation). An example of a negative correlation is the

relationship of helmets to deaths caused by motorcycle accidents: the more that helmets are worn, the fewer the resulting accidental deaths. A positive correlation would be the amount of rainfall in a given year and average tree growth.

Correlations can be powerful tools, but need to be considered carefully because...

...they don't necessarily indicate causality, but rather the presence of a relationship. While we know intuitively that rainfall causes tree growth, we can't say for sure whether depression causes alcoholism or alcoholism causes depression. But we do know that the two are correlated.

...strong mathematical correlations may not actually mean a relationship exists. For example, the cost of buying a new home has risen dramatically over the last few years, as has the cost of breakfast cereal. Would you be willing to believe that a strong relationship exists between the two?

Remember, correlations are mathematical constructs with mathematical meanings. Thus, there is always the chance that they do not accurately describe real-world relationships. When in doubt, fall back on one of the best tools at your disposal: common sense!

A Few Examples...

Election Campaigns

As you watch the news on TV, the latest election poll results are announced. The reporter remarks on the fact that "Candidate Brilton is pulling away from his adversary." Behind her, this chart flashes up on the screen:

Approval Ratings (%)			
Candidate	Last Week	This Week	Change
Rob Mole	35	34	-1
Clint Brilton	53	55	+2
Russ Moreau	8	8	0
Undecided	4	3	-1
Sampling error is +/- 3 points.			

At first glance, the reporter seems right: Brilton's score looks as if it increased by 2 points, while Mole's declined by 1 point. But the small print at the bottom of the chart tells a different story. Because the standard error is +/- 3 points, any change equal to or less than 2 points is really no change at all. Thus, the candidate's ratings didn't truly change at all.

Canvassing for Donations

As treasurer of Citizens for A Better Society (CABS), you are in charge of neighborhood fund-raising efforts. Past efforts have shown that the odds of receiving donations go down to practically zero if home prices fall below \$100,000. Further, it doesn't seem to matter how much more than \$100,000 the house is worth, only that it costs more than this apparently magic number. You send out three trusty lieutenants to find the best neighborhood. They come back with the following results:

- Neighborhood A: Average (mean) home price: \$125,000
- Neighborhood B: Average (mean) home price: \$130,000
- Neighborhood C: Average (mean) home price: \$140,000

Based on the above, your aides immediately opt to begin canvassing Neighborhood C. You're not sure, and ask them to show you their data. After compiling it and doing a few simple calculations, you come up with this chart:

Home Values for Neighborhood: (\$000)		
A	B	C
90	70	60
100	70	60
110	80	70
110	80	80
120	80	80
130	90	170
140	90	180
140	200	210
150	200	240
160	340	250
Mean=125	Mean=130	Mean=140
Median=125	Median=85	Bimodal! Mean 1= 70 Mean 2= 210

Now, the situation looks very different. Neighborhood C is really two very different neighborhoods, with an equal group of high and low value homes. The total number of homes worth over \$100,000 is five.

Next, you look at Neighborhood B, which has the second highest mean home value. But the results have been skewed by the fact that there are a few very expensive homes and many more modest ones. Thus, the median tells a truer story for this group, since at \$85,000 it is more representative of the entire neighborhood. In fact, only three homes are worth over \$100,000.

Finally, you examine Neighborhood A. Even though it has the lowest mean of the three, it has the most homes worth over \$100,000 (9 out of 10). Since the mean and the median are the same, the distribution should be more centrally clustered and less skewed, which in fact is the case.

Your conclusion? Go with Neighborhood A, which has the lowest mean but the largest number of homes that fit your criteria.

Buying A Car

After thoroughly testing a number of cars, you have narrowed the decision down to either a new Flotsam or a new Jetsam. You call the dealers and ask about consumer satisfaction ratings for their specific car.

The Flotsam dealer tells you that his car has seen the biggest jump in customer satisfaction ever recorded by the Powerless Company, which specializes in these surveys. In fact, he says, there's no need to even call the Jetsam dealer. Since the Flotsam satisfaction number jumped by 20 points and the Jetsam number didn't change in the last year, the choice is obvious.

Not quite convinced, you hang up and call the Jetsam dealer. Yes, she says, the Flotsam number did go up by 20 points -- from an approval rating of 23% to one of 43%. And yes, the Jetsam number didn't change -- it stayed the same at its old level of 96%.

Which car would you choose? Once you can look at both the absolute and relative numbers, your decision may change.

Summary...

A basic understanding of statistics is a critical component of informed decision making. Statistical concepts are not hard to master, and mastery will help ensure accurate use, and minimum misuse, of the large quantity of numerical information that confronts us every day.

Here is a simple table that you can refer to when thinking about statistics:

Statistics Crib Sheet

Here is a list of 10 questions to always ask when confronted with statistics:

- Was the study performed by a reputable, third party research organization?
- When was the study conducted and for whom?
- Is the sample size large enough to be projectable?
- Was the sample selected randomly? If not, is there a good reason?
- If "the average" is discussed, is it clear which one (mean, median, mode) is being used and why?
- Are confidence limits and sampling errors clearly indicated?
- Can you get a feel for both absolute and relative changes? Over what time period?
- Do the cited correlations make sense?
- Do the conclusions drawn by the study hold up in light of the questions listed above?
- Are other similar studies cited? Are the results compared and comparable to the findings from the current study?

References...

1. [Full House: The Spread of Excellence from Plato to Darwin](#), Stephen Jay Gould (Random House, 1997).
2. [How to Lie with Charts](#), Gerald E. Jones (SYBEX, 1995).
3. [How to Lie with Statistics](#), Darrell Huff (W. W. Norton, 1993).
4. [How to Think About Statistics](#), John L. Phillips (W. H. Freeman and Company, 1988).
5. [Introduction to Probability](#), John E. Freund (Dover Publications, 1973).
6. [Introduction to Statistical Analysis and Inference for Psychology and Education](#), Sidney J. Armore (Krieger Publishing Company, 1983).
7. [Statistics without Tears](#), Derek Rowntree (MacMillan, 1982).
8. [Statistics Concepts and Controversies](#), David S. Moore (W H Freeman and Company, 1996)
9. [Statistical Deception at Work](#), John Maura (Lawrence Erlbaum Associates, 1992).
10. [Tainted Truth](#), Cynthia Crossen (Simon & Schuster, 1994).