Statistics

By Robert Niles

Numbers can't "talk," but they can tell you as much as your human sources can. But just like with human sources, you have to ask!

So what should you ask a number? Well, mathematicians have developed an entire field - statistics - dedicated to getting answers out of numbers. Now, you don't have to have a degree in statistics in order to conduct an effective "interview" with your data. But you do need to know a few basics.

Here, described in plain English, are some basic concepts in statistics that every writer should know...

So, You're a Beginner?

Mean

This is one of the more common statistics you will see. And it's easy to compute. All you have to do is ***add*** *up all the values in a set of data and then* ***divide*** *that sum by the number of values in the dataset. Here's an example:*

Let's say you are writing about the World Wide Widget Co. and the salaries of its nine employees.

The CEO makes $100,000 per year,
Two managers make $50,000 per year,
Four factory workers make $15,000 each, and
Two trainees make $9,000 per year.

So you ***add*** *$100,000 + $50,000 + $50,000 + $15,000 + $15,000 + $15,000 + $15,000 + $9,000 + $9,000 (all the values in the set of data), which gives you $278,000. Then* ***divide*** *that total by 9 (the number of values in the set of data).*

That gives you the ***mean****, which is $30,889.*

Not a bad average salary. But be careful when using this number. After all, only three of the nine workers at WWW Co. make that much money. And the other six workers don't even make half the average salary.

So what statistic should you use when you want to give some idea of what the average ***worker*** *at WWW Co. is earning? It's time to learn about the* ***median****.*

Median

Whenever you find yourself writing the words, "the average worker" this, or "the average household" that, you don't want to use the mean to describe those situations. You want a statistic that tells you something about the worker or the household in the middle. That's the ***median****.*

Again, this statistic is easy to determine because the median literally ***is*** *the value in the middle. Just line up the values in your set of data, from largest to smallest. The one in the dead-center is your median.*

For the World Wide Widget Co., here are the worker's salaries:

$100,000
$50,000
$50,000
$15,000
$15,000
$15,000
$15,000
$9,000
$9,000

That's 9 employees. So the one halfway down the list, the fifth value, is $15,000. That's the median. (If halfway lies between two numbers,split 'em.)

Comparing the mean to the median for a set of data can give you an idea how widely the values in your dataset are spread apart. In this case, there's a somewhat substantial gap between the CEO at WWW Co. and the rank and file. (Of course, in the real world, a set of just nine numbers won't be enough to tell you very much about anything. But we're using a small dataset here to help keep these concepts clear.)

Here's another illustration of this: Ten people are riding on a bus in Redmond, Washington. The mean income of those riders is $50,000 a year. The median income of those riders is also $50,000 a year.

Joe Blow gets off the bus. Bill Gates gets on.

The median income of those riders remains $50,000 a year. But the mean income is now somewhere in the neighborhood of $50 million or so. A source now could say that the average income of those bus riders is 50 million bucks. But those other nine riders didn't become millionaires just because Bill Gates got on their bus. A reporter who writes that the "average rider" on that bus earns $50,000 a year, using the median, provides a far more accurate picture of those bus riders' place in the economy.

(Statisticians have a value, called a standard deviation, that tells them how widely the values in a set are spread apart. A large SD tells you that the data are fairly diverse, while a small SD tells you the data are pretty tightly bunched together. If you'll be doing a lot of work with numbers or scientific research, it will be worth your time to learn a bit about the standard deviation.)

On to Percent change

Percent

Percent changes are useful to help people understand changes in a value over time. Again, figuring this one requires nothing more than third-grade math.

*Simply **subtract** the old value from the new value, then **divide** by the old value. **Multiply** the result by 100 and slap a % sign on it. That's your percent change.*

Let's say Springfield had 50 murders last year, as did Capital City. So there's no difference in crime between these cities, right? Maybe, maybe not. Let's go back and look at the number of murders in those towns in previous years, so we can determine a percent change.

Five years ago, Capital City had 42 murders while Springfield had just 29.

Subtract the old value from the new one for each city and then divide by the old values. That will show you that, over a five year period, Capital City had a 19 percent increase in murders, while Springfield's increase was more than 72 percent.

That's your lead.

Or is it? There's something else to consider when computing percent change. Take a look at ***per capita*** *to find out.*

Per capita

Percent change in value tells you only part of the story when you are comparing values for several communities or groups. Another important statistic is each group's ***per capita*** *value. This figure helps you compare values among groups of different size.*

Let's look at Springfield and Capital City again. This year, 800,000 people live in Springfield while 600,000 live in Capital City. Five years ago, however, just 450,000 people lived in Springfield while 550,000 lived in Capital City.

Why is this important? The fact that Springfield grew so much more than Capital City over the past five years could help explain why the number of murders in Springfield increased by so much over the same period. After all, if there are more people in a city, one might expect there to be more murders.

To find out if one city really is more dangerous than another, you need to determine a per capita murder ***rate****. That is, the number of murders for each person in town.*

To find that rate, simply ***divide*** *the number of murders by the total population of the city. To keep from using a tiny little decimal, statisticians usually multiply the result by 100,000 and give the result as the number of murders per 100,000 people.*

In Springfield's case, 50 murders divided by 800,000 people equals a murder rate of 6.25 per 100,000 people. Capital City's 50 murders divided by 600,000 people equals a murder rate of 8.33 per 100,000 people.

Five years ago, Springfield's 29 murders divided by 450,000 people equaled a murder rate of 6.44 per 100,000 people. And Capital City's 42 murders divided by 550,000 equaled a murder rate of 7.64 per 100,000 people.

In Percent, we found that the number of murders in Springfield increased 72 percent over five years, while the number of murders in Capital City grew by just 19 percent. But when we now compare **per capita** murders, Springfield's murder rate decreased by almost 3 percent, while Capital City's per capita murder rate increased by more than 9 percent.

There's the real story....

This material comes from Niles' detailed explanation of how numbers work