TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
January 2006 — High School Edition

www.rpdp.net

The *Nevada State Mathematics Standards in Data Analysis* requires students to know and be able to "select and use measures of central tendency" and "analyze validity of statistical conclusions" including "inappropriate uses of measures of central tendency." The *Nevada High School Proficiency Exam* has questions dealing with these standards, as do other standardized tests. In this issue of *Take It to the MAT*, we'll look at the mean and median and see why one is sometimes better than another for describing the center of data.

There's an old story about 1984 University of North Carolina geography graduates earning average salaries of over $1,000,000 right out of college. That must be one awesome program! Alas, it turns out that *one* of those graduates was Michael Jordan, whose geography knowledge came in handy when having to travel to the cities of the NBA. Without Jordan, the mean salary was probably closer to $25,000. This shows the effect that an outlier—an extreme value not in the general pattern of the data—can have on measures of center.

Take for example a company which has 7 employees. Their salaries, in thousands of dollars, are 20, 25, 30, 30, 30, 35, and 40. The *mean* salary is $\frac{20+25+30+30+30+35+40}{7}=\frac{210}{7}=30$, or $30,000. The *median* salary is also $30,000. There are no values that seem to be in the extreme. Either value would be a fair representation of the typical salary in the company. There are few that earn less, a few that earn more, but $30,000 is a good single value to represent all seven salaries.

Now, suppose that we include the owner's salary of $110,000. Now the mean salary is $\frac{20+25+30+30+30+35+40+110}{8}=320=40$, or $40,000. The median salary is still $30,000. Which is a better representative of the typical salary of the employees? The median still looks good at $30,000. If one were told the "average" salary of the current employees were $30,000, this would be a fair statement. One would not be surprised to learn that a couple people make less and a couple people make more, but $30,000 is in the center. If one were told that the "average" were $40,000, however, it may come as a shock when we learn that only one person makes more than that amount and three-fourths of all employees make less. Clearly, the median is the better representative measure here.

It seems that in the presence of extreme values, the median is the better choice for a measure of center. This is generally the case. Since the median is a true measure of *location*, in this case the center, we get a representative number regardless of the data values. But since the mean is computed using those data values, it can be affected by outliers. Where the data goes, so goes the mean.

Does this imply that we should always use the median and never use the mean? Not at all. There are advantages and disadvantages to both. Much of statistics is build upon the mean because of its properties, so it is more desirable in many cases. But in the presence of extreme data values, it's safer—and more appropriate—to report the median. Simply put, the median is less susceptible to outliers.