By Mike Aamodt, Associate Editor

Mean versus Median: It Can Make a Difference

I know what you are thinking. How lame has the Technical Affairs column become when the topic is the mean versus the median? As boring as the topic sounds, it really does matter and it is not a topic that we spend much time thinking about, but probably should. After all, on a daily basis, how often do we make references to such things as "the average salary" or "the average tenure" of our employees? Probably fairly often. So, let's start the discussion with the basics that everyone knows and then move on to the good stuff.

When we refer to the average/typical person, situation, or organization, we are basing our reference on a measure of central tendency. A measure of central tendency is a statistic that summarizes the most common values in a dataset. The three measures of central tendency are the mean, median, and mode. The mean is the arithmetic average of a set of scores, the median is the middle score in the dataset (the point at which an equal number of scores fall above and below), and the mode is the most frequently occurring score. Because there are three measures of central tendency, rather than using a term such as the average tenure of our employees, it is important to be specific and use such terms as the median tenure, the mean tenure, or the modal tenure.

The mode is the least useful and least commonly used measure of central tendency. The mode should be used when the data are categorical or the goal of the analysis is to determine the most likely event that will occur. In fact, for categorical/nominal/discrete data, the mode is the only appropriate measure of central tendency. Take, for example, the data shown in Table 1 in which you have coded race as 1=White, 2=Black, 3=Hispanic, 4=Asian American, and 5=Native American. The modal race code—most frequently occurring—is 1 (White). Taking the mean of race code would result in a value of 2.33, a statistic that is meaningless (e.g., is our typical employee 1/3 of the way between being Black and Hispanic?)!

Table 1: Example of Categorical Data

Race Code	Race	Frequency
1	White	20
2	Black	10
3	Hispanic	15
4	Asian American	7
5	Native American	3
	Total	55
2.33	Mean	

The other use for the mode is when we want to infer from our data the event that is most likely to occur in the future. For example, if you like to solve the Cryptoquote in the daily paper, an analysis of letters used in English writing indicates that the most commonly used letter (the mode) is an "e." Thus, a typical strategy is to substitute an "e" for the most frequently occurring code (my wife tells me that there are many other strategies, but that will have to be the topic of another column). If you are not a Cryptoquote fan, here is another example. If a district attorney and a defense attorney were trying to reach a plea agreement, the defense attorney would probably be interested in the sentence most commonly administered by a particular judge (mode) rather than the mean or median sentence.

Theoretically Speaking: Mean versus Median

In deciding whether or not to use the mean or the median, consideration must be given to both statistical and theoretical issues. Let's start the discussion with the theoretical issues that are often ignored in selecting a measure of central tendency and end with the statistical issues.

The theoretical issues involve what you are going to do with the mean or the median. Typically, people use a measure of central tendency to do one of three things: Describe the average/typical person/situation, predict future behavior or make inferences about future needs, and/or compare two or more groups.

If our purpose is to describe the norm or the typical person/organization, then the median is the most appropriate measure of central tendency. If our purpose is to predict future behavior, to make an inference about necessary resources, or to compare two groups, the mean is the theoretically most appropriate measure of central tendency.

Describing Data

If the measure of central tendency is to be used for *descriptive purposes only*, the mean and median often tell very different stories, and the median is the most appropriate measure. Examples of descriptive questions might be:

- What is the average/typical tenure of our employees?

(continued on next page)

- What is the average/typical amount of money spent per tourist?
- What is the average/typical salary for a billing clerk in Boise, Idaho?
- How long does it take the average/typical unemployed person to find a job in Roanoke, Virginia?

These questions are descriptive in nature, because the answer is not being compared to another measure of central tendency to see if the two measures are significantly different from one another. That is, we are asking, “What is the typical tenure of our employees?” rather than “Is the typical tenure of our employees significantly different from employees at a competing organization?”

The median is the more appropriate measure of central tendency when the goal is to understand the “typical” person or situation. For example, in a situation in which an employee has missed six days of work in the past year and his supervisor wants to make a judgment about whether these six days are “above average,” the question is, “Compared to the typical employee, is six days atypical?” If we use the data from Table 2, using the median would result in the correct judgment that the six days is less than the seven days that the “typical” employee misses.

Table 2: Absenteeism Data

Employee		Number Days Missed
Carrie		11
Bo		9
Vonzell		9
Anthony		8
Scott		8
Constantine		7
Anwar		2
Nadia		1
Nikko		0
Jessica		0
Mikalah		0
	Total	55
	Mean	5.0
	Median	7.0

(continued on next page)

As another example, imagine that an organization is negotiating a settlement with the EEOC regarding back wages. A key statistic in such computations is the number of weeks that the average/typical/normal person is unemployed. In this case, the median is the proper measure of central tendency, because the goal is to describe the "typical unemployed person" rather than to project the total number of weeks that all unemployed people will be out of work. That is, there will be some people who will be employed for long periods of time, because they are not trying to find jobs, are not qualified for any jobs, or are in unusual occupations for which there are few openings. Likewise, there will be people who find a job in one day. Such people do not represent the "typical" applicant and greatly skew the data.

Why is the distinction important? Imagine a situation in which regional data indicate that the median number of weeks to find a job is 8 and the mean is 30 (the mode is 4). If back pay was based on the median and there were 10 employees entitled to back pay and the average weekly salary was $300, the estimated back pay would be $24,000 ($300 x 10 people x 8 weeks). The back pay estimate based on the mean would be $90,000—a substantial difference!

It is important to understand in the above example that neither the mean nor the median is more plaintiff-friendly or defendant-friendly than the other. When there are no outliers and the data are normally distributed, the mean and the median are identical (or very similar). As shown in Figures 1 and 2, when they are not similar, the median will be higher than the mean when the data are negatively skewed (more high scores); and the mean will be higher than the median when the data are positively skewed (more low scores). Thus, with normally distributed data, the choice of the mean or median should be based on how the measure of central tendency will be used.

Predicting and Making Inferences

As an example of making an inference about necessary resources, let's imagine that we are having a cookout and need to determine how many shrimp to purchase for our 20 guests. We just happen to have data from a previous cookout, and these data are conveniently located in Table 3. At our last cookout, the 11 guests ate a total of 66 shrimp—the mean number of shrimp eaten per guest was 6.0 and the median 5.0. Had we used the median to estimate the number of shrimp to buy for the "typical guest," we would have purchased 55 shrimp and would have been 11 short of what our guests actually ate.

Table 3: Cookout Data

Guest		Number of Shrimp Eaten
Taylor		12
Katharine		11
Elliott		10
Chris		8
Paris		6
Kellie		5
Ace		4
Bucky		4
Mandisa		2
Lisa		2
Kevin		2
	Total	66
	Mean	6.0
	Median	5.0

As another example, imagine that we want to anticipate staffing needs by looking at the number of days employees were absent last year. Assume that we have 50 employees and we want to base our estimate on the 11 employees we had last year. The number of days missed for each of the 11 employees is shown in Table 2 on Page 4. The mean number of days missed was 5.0 per employee and the median was 7.0. If we hired extra staff with the idea that our 50 employees would each miss the median number of days absent (50 x 7 = 350 days), we would overestimate our staffing needs compared to the estimate provided by the mean number of days absent (50 x 5 = 250 days).

Statistically Speaking: Mean versus Median

As mentioned previously, if the data are normally distributed, the mean and median will be identical or very similar. If the data are not normally distributed or the sample size is small, the median is the better choice because it is less susceptible to outliers than is the mean.

How do you know if your data are normally distributed or if there are outliers? The easiest approach is to simply look at the difference between the mean and the median. If they are far apart, the data are not normally distributed, and the median should be used. For example, as shown in Table 4, the mean and median weeks to find a job are very different, indicating that the data are not normally distributed and that the median is statistically the more appropriate measure of central tendency.

(continued on next page)

Table 4: Duration of Unemployment in Weeks (2002)

State	Mean Duration	Median Duration
Alabama	17.8	9.2
Alaska	13.3	6.7
Arizona	13.7	7.1
Arkansas	15.8	8.3
California	17.3	9.2
Colorado	15.5	7.7
Connecticut	17.7	9.3
Delaware	16.1	7.8

A second, but rough, approach is to look at a plot of the data to see if there are extreme scores or if the plot looks very different from that of a normal curve. Examples of such plots are shown in Figures 1 and 2.

Figure 1: Example of a negatively skewed distribution

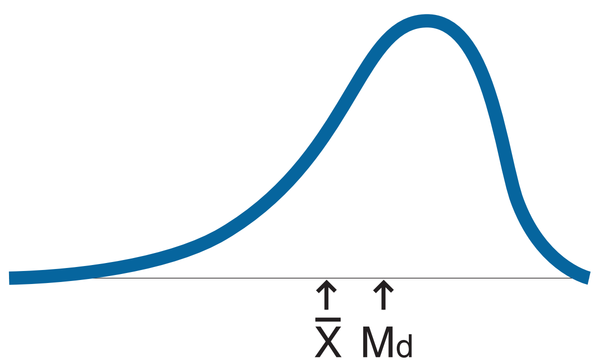
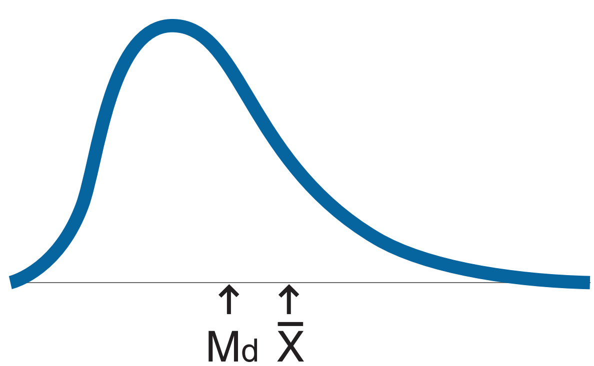


Figure 2: Example of a positively skewed distribution



A third approach, and the most accurate, is to look at measures of kurtosis and skewness (Excel, SAS, and SPSS provide these two measures as part of their descriptive statistics options). Kurtosis refers to the shape of the peak of a distribution (pointed or flat) and skewness refers to how symmetrical the left side of the curve is to the right side. If data are normally distributed, the skewness and kurtosis values will be close to zero. As shown in Figure 1, a negative skewness value indicates that the distribution has a negative skew (mostly high values and a few extreme low values). As shown in Figure 2, a positive skewness value indicates that the distribution has a positive skew (mostly low scores and a few extreme high scores). Kurtosis values that exceed zero indicate a sharper peak than normal, whereas values less than zero indicate a flatter peak than normal.

The million-dollar question becomes, how far from zero do skewness and kurtosis values need to be to suggest that the distribution is not normal? Though experts disagree about the answer to this question, a simple, yet accurate, approach is to compute the standard errors for skewness and for kurtosis. SPSS and SAS provide the standard error in their descriptive statistics output, but Excel does not. The formula for approximating the standard error for kurtosis is the square-root of (24÷N) and the formula for approximating the standard error of skewness is the square-root of (6÷N). For example, if you had 50 employees, the standard error for kurtosis would be the square root of (24÷50), which equals 0.48.

If either the observed skewness or kurtosis value exceeds two standard errors, one might conclude that the distribution is not normal. For example, suppose that our data indicated a kurtosis of 0.11 and a skewness of 1.21. Assume that the standard error for kurtosis is .25 and for skewness is .12. Given that two times the kurtosis standard error is .50, and our kurtosis value is .11, we would not see any significant problems involving kurtosis. Skewness, however, is a different story. The value of two standard errors for skewness is .24. Our observed value of 1.21 greatly exceeds two standard errors indicating that our data are significantly positively skewed (we have a lot more low scores than high scores). Thus, the arithmetic mean would not be the appropriate measure of central tendency in this case.

SPSS and SAS provide two additional statistics that help determine if a distribution is normal: Shapiro-Wilk and Kolmogorov-Smirnov. The Shapiro-Wilk test is used for smaller samples (under 2,000) and the Kolmogorov-Smirnov for larger samples. Both tests consider skewness and kurtosis simultaneously. Other tests include the Stephens' test for normality, D'Agostino-Pearson test, D'Agostino's test for skewness, and the Anscombe-Glynn test for kurtosis.

(continued on next page)

If the values of these tests are not statistically significant, you can assume that your distribution is fairly normal. If, however, the value is significant, you may still need to look individually at the skewness and kurtosis values. The reason for this further investigation is because problems with kurtosis do not greatly affect t-test or ANOVA results, but problems with skewness do. So, if the problem involves kurtosis, you can still conduct a t-test or ANOVA. It is important to note that the above approaches tell you whether the amount of skewness or kurtosis is statistically significant; they don't, however, tell you if the amount of skew or kurtosis is practically significant. That is, is the departure from normality enough to actually affect the results of the t-tests, ANOVAs, or regression analyses?

An important caution is needed regarding the use of the tests discussed in the previous two paragraphs. Tests of normality are inaccurate with very small samples (<10) and can indicate statistically significant, yet practically insignificant, normality problems with large sample sizes. Due to these problems, some statisticians advise that these tests not be used. Furthermore, with large sample sizes, t-tests and ANOVAs are fairly immune to problems caused by non-normality.

Even with a Normal Distribution, It Matters

Statistically, if the choice between the mean and the median doesn't matter when the data are normally distributed, but the median is better when the data are not normally distributed, a logical question is, "Why not always use the median?" The answer depends on what you are going to do with the measure of central tendency—in this case, the mean or median.

If the purpose is to conduct further statistical analyses to answer such questions as, "Is the average/typical salary for women lower than the average/typical salary for men?" or "Do employees who complete a customer service training program perform better than employees without such training?" it is better to use the mean, because the common statistics used to test such questions (e.g., t-tests, ANOVAs) require the use of means rather than medians. There are many nonparametric statistics that can test differences between medians (e.g., Fishers exact test), but these statistics tend to be more complicated, less powerful, and more difficult to explain to a non-statistical audience.

Even if the data are not normally distributed, the mean can be used potentially for such analyses as t-tests and analyses of variance by forcing the data into a more normal

(continued on next page)

distribution using transformations or by using a mean other than the arithmetic mean. Common transformations include the square-root transformation, log-linear transformation, and inverse transformation. Square-root transformations are used for mildly skewed data, logarithmic transformations for moderately skewed data, and inverse transformations for more heavily skewed data. Though transformations usually make a distribution more normal, there is no guarantee that they will always result in a normal distribution.

The problem with using transformations is that, because they change the nature of the data, results are more difficult to explain. For example, if the mean salary for male accountants is $45,000 and the mean for female accountants is $43,000, it is easy to explain that there is a $2,000 mean difference in salary. However, if we used a logarithmic transformation, our analysis would indicate that the average salary for men is 4.65 and for women 4.63. Such numbers would not make practical sense to many managers or employees.

Another approach to make a skewed distribution more normal is to use a mean other than the arithmetic mean. As mentioned at the beginning of the column, the arithmetic mean is simply the sum of the scores divided by the number of scores. A trimmed mean is one in which the scores are ranked and a certain percentage (usually 5%) of the scores at the bottom and the top are removed. A trimmed mean is a good approach when the distribution is fairly normal but there are a few extreme scores. It is not a good technique, however, when the sample size is small.

An interesting, but complicated, approach to normalizing a distribution is to use M-estimators, means that are computed by providing less weight to values that are further from the center of the distribution. SPSS provides four such M-estimators: Huber's M-Estimator, Tukey's biweight, Hampel's M-estimator, and Andrews' wave. The *geometric mean* and the *harmonic mean* also provide less weight to values that are further from the center of the distribution. Rather than arithmetic means, geometric means are commonly used to compute average rates of growth (e.g., interest earned on a retirement fund) and harmonic means are used to compute the average distance per time (e.g., miles walked per hour).

Table 5 provides an example of the various means for a distribution of salaries with significant kurtosis and positive skewness problems. As you can see from the table, the arithmetic mean is higher than the other means, and the median is much closer to the other means and estimated means.

Table 5: Comparison of Measures of Central Tendency in a Positively Skewed Distribution

Measure of Central Tendency	Salary
Arithmetic mean	$105,973
Trimmed mean	$104,929
Geometric mean	$104,727
Harmonic mean	$103,559
Huber's M-estimator mean	$103,375
Hampel's M-estimator mean	$103,357
Median	$103,195
Tukey's biweight estimated mean	$102,268
Andrews' wave estimated mean	$102,245

Conclusion

So, what does all this mean from a practical or statistical standpoint? The choice of the mean or median goes beyond the simple statistical question of whether data are normally distributed. Choice of the proper measure of central tendency results in better decision making. From a practical perspective, to summarize the discussion:

- If we are making projections about future needs, the mean is the best choice.
- If we are describing the typical person or situation, the median is the better choice.
- If we want to determine if two measures of central tendency are significantly different from one another, the mean is the better choice as long as our data are normally distributed. If they are not, we should transform the data to force a more normal distribution. If the transformations do not result in a normal distribution, we must use the median.

Final Comments

While researching this column, I was surprised to see the extent to which experts disagree on the issues discussed in the column. If any *ACN* reader has some thoughts on this topic, please e-mail them to me and I will include them in the October column. I would like to thank Bobbie Raynes of New River Community College, David Cohen of DCI Consulting, Dan Biddle of Biddle Consulting Group, Jeff Aspelmeier of Radford University, and Michael Surrette of Springfield College for providing very useful comments and suggestions on earlier drafts of the column. I don't know what possessed them to agree to read the drafts, but I am grateful that they did.—ACN

HR Humor

Differences Between Employee and Management Behavior

When you take a long time, you're slow.
When your boss takes a long time, he's thorough.

When you don't get something done, you're lazy.
When your boss doesn't get something done, she's too busy.

When you make a mistake, you're an idiot.
When your boss makes a mistake, he's only human.

When you do it your own way, you don't do what you are told.
When your boss does it, she's showing creativity.

When you do it on your own, you're overstepping your bounds.
When your boss does it, he is demonstrating initiative.

When you take a stand, you're being bull-headed.
When your boss takes a stand, she's being firm.

When you're out of the office, you're wandering around.
When your boss is out of the office, he's on business.

When you call in sick, you're goofing off.
When your boss calls in sick, she must be very ill.

When you apply for leave, you must be going on an interview.
When your boss applies for leave, it's because he is overworked.

When you're seen shopping during work hours, you're a slacker.
When your boss is doing the same, she's picking up office supplies.

When you get a raise, you're lucky.
When he gets one, he really earned it.

When you do a good job, you get a pat on the back.
When she does a good job, she gets a bonus.

When you violate a rule, you're self-centered.
When your boss skips a few steps, he's being original.

When you please your boss, you're brown-nosing.
When your boss pleases his boss, she's being cooperative.

When you help a peer, you're not busy enough.
When your boss does it, he's a team player.

When someone else does your work, you're passing the buck.
When someone else does her work, she's assigning responsibility.—ACN