

# Statistical Thinking in Medicine, Part 5: Descriptive Statistics and Quantifying ‘Unusual’

Robert A. Calder, MD, MS; Jayshil J. Patel, MD

In part 4 of the “Statistical Thinking in Medicine” series, we discussed that statistics allows us to learn about a population from a random sample. Now, we need a succinct way to summarize both our sample and the corresponding population. The two most important characteristics of both a population and a sample are the “average” and the “spread” of the values. These are among the most important functions of statistics and normally they are among the first concepts taught in most statistics courses. We have chosen to delay this discussion, hoping to first inspire our readers with an understanding of the main uses of statistics in medicine before presenting detailed methods. Our principal focus here will be conceptual rather than mathematical since the mathematics sometimes obscures the overall ideas. Therefore, in this installment, our aims are to describe (1) how to best measure the “average,” (2) how to quantify the spread of the data, (3) how the spread of the data in a population differs from the

spread of the mean in a random sample, and (4) how these measures help us to determine what is “unusual.”

## What is an “Average” Value? The Mean

Imagine you measure the height of some sample of Wisconsin high school students. What one number would best represent the heights of the students in this sample? The word “average” may come to mind. There are multiple ways to calculate an average, but the three most frequent ways to pick the average are the mean, median, and mode.

The mean is the arithmetic average of some group of numbers. For example, for the numbers 1, 2, and 3, the mean is the sum of the numbers—6—divided by the total number of values 3. In this example, the mean is 2. The mean often is used as the average. However, when the data are not evenly spaced, ie, heavily skewed, the mean can be heavily weighted toward the higher values. For example, the average of 1, 4, 5, and 30 is 40 divided by 4, which is 10, a number that would not be most people’s first choice of the average of these numbers. A practical example of skewed data is when the mean is cited as the “average” salary in the United States. The mean is not very representative of all salaries since it is heavily weighted by the small number of very high salaries. A humorous example of a right-skewed dataset would be the mean salary today of students in Bill Gates’ fifth grade class!

Mathematically, the population mean is computed as:

$$\mu = \sum x_i / N$$

where the Greek letter mu ( $\mu$ ) represents the population mean, the Greek letter sigma ( $\Sigma$ ) means to add up each of the individual “ $x_i$ ” values, and the capital N represents the number of units in the whole population.

From our example above, suppose instead of merely taking a sample, we measure the height of every Wisconsin high school student. If we then added up all of those measurements and divided by the total number of students, we would derive the *population* mean. Suppose that population mean was 5 feet 6 inches. It is the “population” mean because we measured the height of every Wisconsin high school student.

A random sample is a subset that is representative of the population. A sample mean is calculated as:

$$\bar{x} = \sum x_i / n$$

where the Latin letter “x” with a bar over it represents the sample mean and small case “n” is the number in the sample. In general, Greek letters are used to represent the key features (“parameters”) of the population, such as the mean ( $\mu$ ) and standard deviation ( $\sigma$ ), and Latin letters are used to represent estimates of these parameters in a sample.

## The Median

Recall that values may not be evenly distributed (they are skewed). When the distribution

• • •

**Author Affiliations:** Medical College of Wisconsin, Milwaukee, Wisconsin (Calder); Division of Pulmonary and Critical Care Medicine, Medical College of Wisconsin, Milwaukee, Wisconsin (Patel).

**Corresponding Author:** Robert A. Calder MD, Adjunct Assistant Professor, Medical College of Wisconsin, Milwaukee, WI; email rcalder@mcw.edu.

of values is skewed, the mean is less representative of the distribution, and the median is often a better measure of average. The median is the “middle value” in some group of numbers. For example, the median of 1, 2, 3, 4, and 10 is 3, the middle value. (The mean of these numbers is 4, which is skewed right, or toward the highest value.) If there are an even number of values, the median is the mean of the middle two values, when listed from lowest to highest. For example, in the distribution 1, 2, 3, 4, 5, 10, the median is 3.5. The median is obviously less affected by extreme values. Therefore, when reporting the “average” US salary, which is skewed, it is often presented as the median: half are above that value and half are below.

### The Mode

Sometimes the mode, or the most common value, is the best measure of average. In our first-year medical school classes for example, the average number of years of education after high school would be best represented by the mode, since almost everyone would have the same value (eg, 4 years of undergraduate studies). The mode is used as the average when a substantial number of values are the same; however, there is no generally accepted rule for what constitutes “substantial number of values.” In the US, the modal salary would most likely be the minimum wage.

## What is the Spread of the Values?

### The Range

There are several ways to describe the spread of the values in a dataset. The simplest is the range: listing the lowest to the highest value in a data set. For example, when measuring heights for each student in our Wisconsin high school example, suppose the range of heights (shortest to tallest) is from 4 feet 4 inches to 6 feet 8 inches.

The interquartile range is the range from the 25th to the 75th percentile. Suppose interquartile range is heights between 5 feet 2 inches to 5 feet 10 inches. In that case, 25% of the students would be shorter than 5 feet 2 inches and 25% would be taller than 5 feet 10 inches. So, the interquartile range represents the “middle half” of the data with one quarter below that range and one quarter above it.

### The Variance

The variance is another way to express the spread of the values. It is the average squared difference from the mean and measures how much the data points are spread out from the mean. To compute the variance for the population, each individual value is subtracted from the mean, that quantity is then squared, and all of these squared values are added up and divided by the total number of values in the population, giving the average (mean) of all the squared values. Squaring these values is done to avoid having to work with absolute values, which are much more difficult to work with. Mathematically, the population variance ( $\sigma^2$ ) is computed as:

$$\sigma^2 = \sum (x_i - \mu)^2 / N$$

In our high school example, the population mean ( $\mu$ ) was 5 feet 6 inches. To calculate the population variance, the population mean height (5 feet 6 inches) would be subtracted from each individual value; that result would then be squared, and each squared result would be added up. That sum would be divided by the total number of students in the population to derive the population variance. Suppose the population variance is 36 (inches squared). That is the average squared difference from the mean.

As with calculating the population mean, the calculation for the sample variance ( $S^2$ ) is similar but uses a different denominator. Mathematically it is:

$$S^2 = \sum (x_i - \bar{x})^2 / (n - 1)$$

Once again, note that Latin letters are used in sample calculations. This equation means that the sample variance is equal to the sum of the squares of the difference between each sample value and the sample mean, divided by “ $n-1$ ”. The complete explanation for dividing by  $n-1$  is beyond the scope of this discussion. However, the key idea is that the mean minimizes the sum of squared differences in the numerator (easily proven with elementary calculus). Since the sample mean is almost never exactly the same as the population mean, the sum of squared differences in the numerator of the sample variance calculation is **smaller** than it would be if the true population mean ( $\mu$ ) were used in the calculation. Dividing by  $n-1$ ,

rather than just “ $n$ ,” helps to correct this issue, making the quotient larger and therefore making it an “unbiased” estimate of the population variance.

### The Standard Deviation

The most frequent measure of spread is the standard deviation ( $\sigma$ ), which is merely the **square root of the variance**. Mathematically, the population standard deviation is:

$$\sigma = \sqrt{\sigma^2}$$

and the sample standard deviation is:

$$S = \sqrt{S^2}$$

One of the virtues of standard deviation is that it is expressed in the same units as the individual measurements. Recall in our high school example, the population variance was 36 inches squared. The square root of that variance is 6 inches, and that is the standard deviation. In a normal distribution, about 68% of the values of the population are within 1 standard deviation of the mean. So, in our high school example, about 68% of the students would be between 5 feet and 6 feet tall (the mean – 6 inches and the mean + 6”). Moreover, about 95% of the population in a normal distribution will be within 2 standard deviations of the mean: 5 feet 6 inches +/- 12 inches.

### Standard Error of the Mean (SEM)

When a sample mean is calculated, how much variation would there be in that calculated mean if it were calculated many times with the same random sample size? The answer is: the “standard error of the mean.” This is really just the standard deviation of the means computed from many samples. But it gets this new name, “standard error,” because the sample mean is an “estimator” of the population mean. *The standard deviation of any “estimator” (eg, relative risk, odds ratio, hazard ratio, etc) is called the “standard error.”* This serves to differentiate the standard deviation of an *estimator* from the standard deviation of *individual values in the population*. The standard deviation of individual values quantifies the overall spread of the data. The standard error of the mean quantifies the standard deviation of the sample mean. Another important difference between standard deviation and standard error is that the standard error of any estima-

tor depends on the *size of the sample used to calculate it*. The larger the sample, the smaller the standard error. The difference between the standard error of the mean and population or sample standard deviation is confusing and one of the most frequently asked questions by students.

Mathematically, the standard error of the mean, if we know the population standard deviation is:

$$SEM = \sigma / \sqrt{n}$$

If the population standard deviation is not known—which is usually the case in practice—mathematically, the estimated standard error of the mean is:

$$\text{Estimated SEM} = S / \sqrt{n}$$

Returning to our high school example, recall that the standard deviation of the population was 6 inches. Again, that means that in the *overall population*, about 68% of the heights of the students were within 6 inches of the population mean. If we were to take a random sample of 16 students, for example, the standard deviation of that sample mean would no longer be 6 inches; it would be much less, because in sampling 16 students and then computing a mean, if that were done many times, *the range for the computed means* would be much less than the range for random picks of individual students. Every group of 16 probably would consist of a few very short people, a few very tall people, and the rest of average height. In fact, the variance of the mean in a sample of 16 students would be cut 16-fold from the population variance ( $\sigma^2/16$ ). So, in this example, the variance of the mean for a random sample of 16 students would be the population variance divided by 16, which is 36/16. The square root of this (6/4 or 1.5) would be the standard error of the mean for a sample of 16. So, about 68% of the means calculated from groups of 16 would be within 1.5 inches of the sample mean—that is, *there would be much less variation for means calculated from groups of size 16 than for individual values*. If the sample size were larger, the standard error would be even less.

## How Do We Determine What Is “Unusual”?

The point of calculating and understanding the

average and spread of values is to determine whether some individual value is “unusual,” whether the mean of some random sample is “unusual,” and whether the difference between two means is “unusual” (analogous to the difference between two treatments).

### Is This Individual Value “Unusual”?

Suppose in our high school example, we randomly chose a student who was 6 feet 6 inches tall. Would that be unusual (in the sense of being statistically significantly tall)? Given that the population mean is 5 feet 6 inches and the standard deviation is 6 inches, someone 6 feet 6 inches tall would be 2 standard deviations above the mean. In a normal distribution, which we assume our high school population has, about 95% of students will be within 2 standard deviations of the mean, 2.5% would be taller than 6 feet 6 inches, and 2.5% would be shorter than 4 feet 6 inches since a normal distribution is symmetric and “bell shaped.” Since only about 2.5% of students would be taller than 6 feet 6 inches, therefore, the “1-tail” *P* value of that student’s height would be 0.025, meeting the traditional definition of statistical significance ( $P < 0.05$ ). Typically, a 2-tailed test is done where deviations that are equally extreme in the other direction (ie, shorter than the mean by the same amount) also are counted. However, a 1-tail test is justifiable here because we were only interested in how unusual that height was (ie, that height or taller). An example below will use a 2-tailed test.

### Is This Sample Mean “Unusual”?

Suppose we randomly choose 16 people from one of the best basketball teams in Wisconsin and we start by assuming (null hypothesis) that they have the same height as all other students in Wisconsin. If we then calculate the mean height of these 16 basketball players to be 5 feet 10 inches, would that imply that the players on this basketball team are significantly different in height than the overall school population in the state? To determine this, we need to know how many standard deviations this sample mean is from the population mean. Above, we calculated the standard error of the mean height for a group of 16 to be 1.5

inches. Since 5 feet 10 inches is 4 inches above the population mean of 5 feet 6 inches and that represents 2.67 “standard errors” for the sample mean ( $4/1.5$ ), the question becomes: how unusual would it be for a value in a normal distribution to be more than 2.67 standard deviations away from the mean (in either direction, above or below)? A table for the normal distribution shows that the probability of being *less than* 2.67 standard deviations is 0.9962. That means that  $1 - 0.9962$  or 0.0038 is the probability of being *more than* 2.67 standard deviations above the mean, and 0.0038 is also the probability of being 2.67 or more standard deviations below the mean. Therefore,  $0.0038 \times 2$  or 0.0076 of the normal distribution is more than 2.67 standard deviations away from the mean. Therefore, that is the *P* value for a “2-tailed” test of whether the sample height of 5 feet 10 inches is different from the mean for the population. In other words, if we assume (the null hypothesis) that the basketball players have the same height as everyone else in the school population, there is less than a 1% chance ( $P = 0.0076$ ) that we would get these or more extreme results. Therefore, we can reject the null hypothesis and accept the alternative hypothesis that the heights of the basketball players on this team are different from the school population overall.

### Are These Sample Means “Unusual”?

Another question we can now answer with this general method is whether the difference between two means is unusual. This is a very common question in medicine. For example, is the mean systolic blood pressure reduction for drug A different than for drug B? To answer this question, subjects could be randomly chosen to receive either drug A or drug B. The mean systolic blood pressure reductions for each group could be calculated, then we could test whether the difference between those mean reductions is statistically significant (at whatever level we decide before doing the study). Our null hypothesis would be that there is no difference in the mean systolic reductions between the drugs. After calculating the difference in the mean blood pressure reductions between the drugs, we would then divide this difference (if any) by the *standard error of the difference*

**between the two means.** Assuming the groups are independent (ie, the results in one group do not affect the other group) the standard error of the difference between the means is the square root of the sum of the variances for each mean. The mathematical formulas for this can be found in any standard statistics text. The key concept here is that some difference, such as the mean systolic pressure reductions with drugs A and B, is calculated and that difference **is then divided by the appropriate standard error** (in this case for the difference between two means), **then the probability of that number of standard error units is determined** and this determines whether the difference is statistically significant. This answers the question: if there is no difference between these drugs, how likely would we see these – or more extreme results – by chance alone?

## Conclusion

In summary, the best measure of “average” depends on the skewness of the data. Variance is the average squared difference from the mean. The calculated sum of squares in the sample variance tends to be smaller than that for the population variance; therefore, the sample

sum of squares is divided by  $n-1$  to create an “unbiased” estimate of the population variance. Standard deviation is the square root of the variance, and it measures variation using the same units as the original data. Standard error of the mean (SEM) is the standard deviation of sample means – the larger the sample, the smaller the standard error of these samples. “Unusual” is quantified by how far some quantity is from what would be expected under the null hypothesis. If our data follow a normal distribution, which they often do because of the central limit theorem, we can then determine where our data fall in a normal distribution and read off in a table the probability of seeing such differences or greater if the null hypothesis is true.

In our next (and final) article in this series, we will discuss what constitutes evidence and how to know when the evidence is sufficient to conclude that something causes something else. Finally, you may wish to test your comprehension of the concepts presented in this article (answers will be provided in part 6).

**Funding/Support:** None declared.

**Financial Disclosures:** None declared.

## Practice Questions/Problems

1. What is the mean of the following numbers: 1, 2, 3, 4, 7, 8?
2. What is the median of the following numbers: 2, 5, 9, 16, 22, 25?
3. If the mean systolic blood pressure of all patients in your practice is 130 mmHg and the standard deviation is 6 mmHg, what percent of your practice would you expect to have systolic pressures above 142 mmHg assuming the systolic pressures follow a normal distribution?
4. Suppose you record the blood pressures of the next 9 patients in your office and calculate the mean systolic pressure of that sample to be 134 mmHg. Would that mean surprise you? What is the standard error of the mean for this random sample of 9?
5. What would the standard error of the mean be for a random sample of 36 of your patients?

## Part 4: Probability Practice Questions and Answers

1. When rolling a die, what is the probability of rolling a 1 or a 2?

*Since rolls of the die are mutually exclusive and the probability of each is  $1/6$ , the probability of rolling a 1 or a 2 is  $1/6 + 1/6 = 2/6$  or  $1/3$  by the addition rule.*

2. If the probability that a laboratory test is positive is 40%, assuming test results on different days are independent, what is the probability of at least one positive test when testing is done on two separate days?

*$0.4 + 0.4 - 0.16 = 0.64$  or  $1 - \text{the probability that both are negative: } 1 - 0.6^2 = 0.64$*

3. As with the conditions in question 2, what is the probability that at least one

test is positive if tests are done 5 days in a row?

*$1 - \text{the probability that all 5 tests are negative} = \text{probability that at least 1 test is positive: } 1 - 0.6^5 = 0.92224$*

4. To make a diagnosis, suppose you order 20 independent laboratory tests, each of which is “normal” in 95% of people. What is the probability that at least one test is abnormal?

*$1 - \text{the probability that all of the tests are normal} = \text{the probability that at least 1 is abnormal: } 1 - 0.95^{20} = \sim 0.64$*

5. How many ways are there to shuffle a standard deck of 52 cards?

*$52! = 8.06 \times 10^{67}$  – a number greater than all of the atoms in the Milky Way Galaxy!*

# WMJ

*Let us hear  
from you!*

If an article strikes a chord or you have something on your mind related to medicine, share it with your colleagues. Email your letter to the editor to

**wmj@med.wisc.edu**

advancing the art & science of medicine in the midwest

**WMJ**

*WMJ* (ISSN 2379-3961) is published through a collaboration between The Medical College of Wisconsin and The University of Wisconsin School of Medicine and Public Health. The mission of *WMJ* is to provide an opportunity to publish original research, case reports, review articles, and essays about current medical and public health issues.

© 2025 Board of Regents of the University of Wisconsin System and The Medical College of Wisconsin, Inc.

**Visit [www.wmjonline.org](http://www.wmjonline.org) to learn more.**