

MEASURES OF DISTRIBUTION

You work for Math Behind the Medicine (MBTM), a consulting company that helps pharmaceutical companies analyze data. You and your team have been assigned to help SBQ analyze the data gathered from samples of two production methods. The product being produced is a 500 mg acetaminophen tablet (over-the-counter pain killer). Your job is to let SBQ know if there is a mathematical difference between the two production processes. One hundred (100) tablets from each process were randomly collected. The data is given in the charts below. Each chart shows the number of tablets that contained a particular amount of acetaminophen.

Production Method One

mg of Acetaminophen	# of tablets
490	6
491	6
492	2
493	5
494	3
495	5
496	8
497	3
498	8
499	4
500	2
501	3
502	5
503	4
504	6
505	3
506	8
507	6
508	1
509	7
510	5

Production Method Two

mg of Acetaminophen	# of tablets
490	1
491	0
492	0
493	0
494	2
495	13
496	3
497	12
498	7
499	5
500	12
501	5
502	8
503	9
504	7
505	9
506	6
507	0
508	0
509	0
510	1

1. If you were just looking at the data, not doing any math, would you say the methods are the same or different? What is it about the data that suggests this?
2. What mathematical methods/calculations do you know of that are used to describe the characteristics of a set of data? Calculate these values for the given data.
3. Can you use the calculated values from question 2 as evidence for your answer to question 1?

Standard Deviation

As you considered the data and the calculated mean, one of your teammates remembers something they learned in their Algebra 2 class years ago. They remind you that the mean is a measure that helps describe the center of a set of data. They suggest you use a measure for distribution (or spread) instead of a measure for center.

A measure of distribution, your teammate explains, is a measure of how spread out data is, or how the data is distributed from its smallest values to its largest values. Suppose, for instance, that Joe has test scores of 60, 68, 69, 78, 90, 95, and 100. Sammy scores 78, 78, 79, 79, 82, 82, and 82. Note that both Joe and Sammy have a mean test score of 80. Joe's scores are more spread out than Sammy's. Measuring the mean will not tell you much about the characteristics of the test takers. A measure of distribution, or spread, will help you see that Sammy consistently scores near 80, while Joe's scores are spread out, or distributed, over a much larger range.

One common measure of distribution is called standard deviation. Calculating the standard deviation can be broken down into several steps. An example is given in the following table. The steps for calculating the standard deviation are in the left column. An example is given in the middle column. You should calculate Sammy's standard deviation in the third column.

How to calculate the Standard Deviation	Joe	Sammy
1. Calculate the mean, symbolically \bar{x} , of the data. Standard deviation is a measure of spread <i>about the mean</i> of the data. To find the mean find the sum of the data and divide by the number of pieces of data.	<p><i>Mean:</i></p> $\bar{x} = \frac{60 + 68 + 69 + 78 + 90 + 95 + 100}{7}$ $\bar{x} = 80$	
2. Find the deviation, or distance from the mean, for each piece of data. This is done by subtracting the mean from the piece of data. If \bar{x} is the mean of the data set $\{x_1, x_2, \dots, x_n\}$, then the deviations of the data from the mean would be found by computing $x_1 - \bar{x}$, $x_2 - \bar{x}$, ..., and $x_n - \bar{x}$.	<p><i>Deviation:</i></p> $60 - 80 = -20$ $68 - 80 = -12$ $69 - 80 = -11$ $78 - 80 = -2$ $90 - 80 = 10$ $95 - 80 = 15$ $100 - 80 = 20$	

<p>3. Once we have found the deviation, or how far from the mean each data point is, we find the <i>mean squared deviation</i>, also called the <i>variance</i>, symbolically s^2. Just as the name suggests we square each deviation, then find the mean of these squares. With the variance, however, we divide by one less than the number of pieces of data in our sum.</p>	<p><i>Variance for Joe's scores:</i> $(-20)^2 = 400$, $(-12)^2 = 144$, $(-11)^2 = 121$, $(-2)^2 = 4$, $(10)^2 = 100$, $(15)^2 = 225$, $(20)^2 = 400$</p> $s^2 = \frac{400 + 144 + 121 + 4 + 100 + 225 + 400}{7 - 1}$ $s^2 = \frac{1394}{6} = 232\frac{1}{3}$	
<p>4. The <i>standard deviation</i>, or s, is the square root of the variance. Symbolically this is $s = \sqrt{s^2}$.</p>	<p><i>Standard Deviation:</i></p> $s = \sqrt{s^2} = \sqrt{232\frac{1}{3}} \approx 15.2425$	

Questions about the standard deviations you just calculated:

1. Why do you think the standard deviation of Joe's test scores is higher than the standard deviation of Sammy's test scores? (It may help to consider the intro paragraphs on the last page.)
2. What does the standard deviation tell you about a data set (what does the standard deviation measure)?

Back to the pain killer production methods:

Now that we know about standard deviation and how to calculate it, let's get back to the data on acetaminophen levels in the samples of tablets taken from our two production methods. It is the responsibility of the SBQ to ensure that the medication produced is safe and effective. If the amount of acetaminophen in a tablet is too high it could be dangerous to the consumer. If the amount of acetaminophen is too low, the tablet will be ineffective.

1. What would happen if SBQ produced an ineffective tablet?
2. What would happen if SBQ produced a tablet that had overly high levels of acetaminophen?
3. Use a calculator or computer spreadsheet application (or if you have it, some statistics software) to calculate the standard deviation of each sample. Record your results below.
 - Standard deviation for Production Method 1:
 - Standard deviation for Production Method 2:
4. Write a half page persuasive report detailing which method SBQ should use, if any at all. As support for your recommendation be sure to include answers to the following questions:
 - How effective is the medication produced using each method?
 - Use statistical data such as the mean and the standard deviation to support your analysis of each method's effectiveness.
 - Discuss possible liabilities that would exist as the result of using each method.