

2.4: Quantifying Variability Relative to the Mean

In the previous lesson, we examined variability with respect to the median. In this lesson, we will develop a measure of variability that depends on the mean. Recall that the mean is the average of all values in a given set. We use all data values in its computation. The sample mean is denoted by \bar{x} and is found by adding all the values in the sample and dividing by the sample size. Its formula is $\bar{x} = \frac{\sum x_i}{n}$. When referring to a population mean, we use the symbol μ ("mu"). It is computed in the same way; we add all values in the population, and divide by the population size.

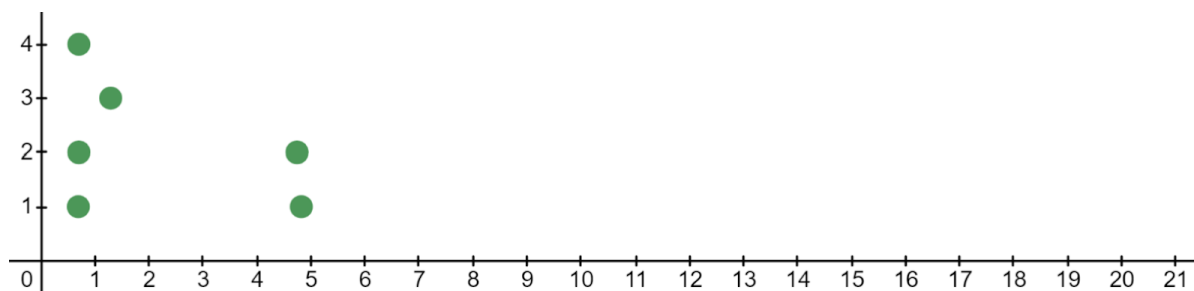
We have seen a few measures of variability so far. The range is the difference between the maximum and minimum values in a set of data. The range is not resistant to outliers because its calculation involves only the two most extreme values in the set of data. The interquartile range or IQR uses only two values as well, but it *is resistant* because it is the difference between the third and first quartiles. Quartiles are found using the resistant measure of center, the median. Therefore, the IQR depends on the median. We will now develop a measure of variability that depends on the mean and uses all values in a given set of data.

The table and dotplots below show the 2022 salaries (in millions of dollars) of a sample of players from the Kansas City Royals and Los Angeles Dodgers.

2022 Salaries (in millions of dollars)

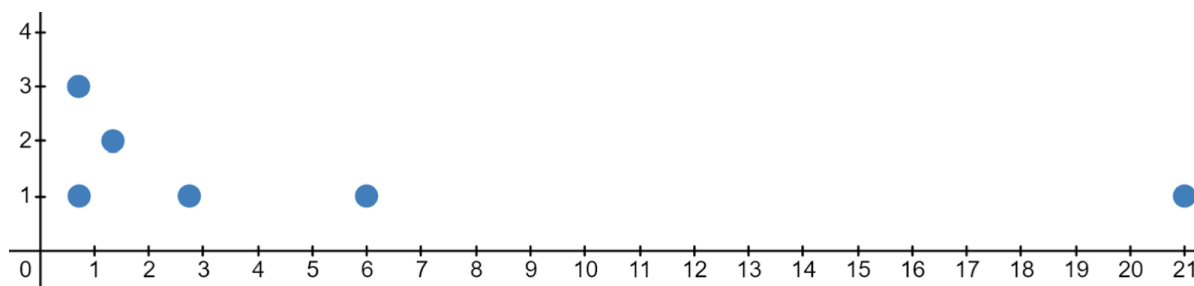
Kansas City Royals	4.83	0.7	4.75	0.71	1.3	0.71
Los Angeles Dodgers	6	0.73	2.75	1.35	0.72	21

Kansas City Royals



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Los Angeles Dodgers



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

1. Let's start by computing the mean salary for each sample. How do they compare?

A measure of variability should tell us how spread out data is. Since we are using the mean as our center, it would be useful to know the distances between each value and the mean. These distances are called **deviations**. Data values that are above the mean have positive deviations. Data values that are below the mean have negative deviations. Formulaically, we say

$$\text{deviation} = (\text{data value} - \text{mean}) = x_i - \bar{x}$$

2. Calculate the deviations for each of the samples and enter them in the tables below.

Kansas City Royals

Value	Deviation
4.83	
0.7	
4.75	
0.71	
1.3	
0.71	

Los Angeles Dodgers

Value	Deviation
6	
0.73	
2.75	
1.35	
0.72	
21	

3. Based on the deviations, which sample is more spread out?

Data values with large deviations (that are farther away from the mean) contribute more to the variability in the data set. Values with small deviations do not contribute as much to the overall variability of a data set. To measure the total amount of variability, we need to combine the deviations into a single number.

4. One way we might do this is by finding the average deviation from the mean. Let's try it! Add all the deviations for the Kansas City Royals. Interpret the result.

Standard Deviation

The standard deviation is a measure of variability that describes the typical deviation from the mean for all values in a set of data. To compute the **standard deviation of a sample**, we complete the following steps:

Step 1: Find the mean and find all deviations from the mean.

Step 2: Square each deviation.

Step 3: Add all squared deviations and divide by one less than the sample size.

Step 4: Take the square root of the result from step 3.

-
5. Complete the table with the square deviations for each value in the sample. Find the total square deviations.

Kansas City Royals

Value	Deviation	(Deviation) ²
4.83		
0.7		
4.75		
0.71		
1.3		
0.71		

Total square deviations:

6. The sum of squared deviations is one way to represent the variability in a distribution. But it is not a commonly used measure. A more commonly used measure is the **sample variance** which is found by dividing the sum of squared deviations by one less than the sample size. In other words, the sample variance is an average of the squared deviations. Formulaically, we say

$$s^2 = \frac{\text{sum of squared deviations}}{\text{sample size minus 1}} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Compute the sample variance for the Kansas City Royals.

7. The **sample standard deviation**, denoted by s , is the square root of the sample variance. Calculate the sample standard deviation for the Kansas City Royals. Include units in your answer.
8. Can the standard deviation be negative? Explain.
9. Can the standard deviation be zero? Explain.

The process of calculating the standard deviation by hand has many steps and is time consuming, even for small sets of data. Therefore, we will usually use technology to compute the mean and standard deviation for us.

10. Use the following instructions to use the desmos graphing calculator to compute the sample standard deviation for the Los Angeles Dodgers.
- Go to <https://www.desmos.com/calculator>.
 - Copy the values from the data set into the first line. Type $D = [6, 0.73, 2.75, 1.35, 0.72, 21]$ Hit enter on your keyboard to go to the next line.
 - Type $\text{mean}(D)$ to compute the sample mean. Hit enter on your keyboard to go to the next line. $\bar{x} =$ _____
 - Type $\text{stdev}(D)$ to compute the sample standard deviation. $s =$ _____
11. Compare the sample standard deviations found in 7 and 8. Which standard deviation is higher? Which distribution is more spread out? Do these values adequately represent the spread/variability in each sample of salaries?

12. In the sample of salaries from the Los Angeles Dodgers, which value impacts the standard deviation the most?

13. Let's say that 21 was incorrectly recorded as 12. Use desmos to find the standard deviation of the set with the error in it:

$D = [6, 0.73, 2.75, 1.35, 0.72, 12]$ How does this error affect the sample standard deviation?

Old standard deviation:

New standard deviation (from set with error):

Outliers and skewing have a large effect on the standard deviation. Therefore, we say that the standard deviation is not a resistant measure of variability.

Summary

We have examined in detail two measures of center (mean and median) and three measures of variability or spread (range, interquartile range, and standard deviation). We calculated the IQR using the median. We calculated the standard deviation using the mean. In deciding which measures of center and spread to use, we need to remember two things:

- Mean and standard deviation go together. Median and IQR go together.
- Both the mean and standard deviation are influenced by outliers and skew.

When the data are skewed or contain outliers, we usually use the median and IQR to summarize the data. When the data are reasonably symmetric we use the mean and standard deviation. In addition, these summary values are never enough. We should always look at a graph as well. This can be a dotplot, histogram, or boxplot.

2.4: Quantifying Variability Relative to the Mean is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by LibreTexts.

- [Current page](#) is licensed [CC BY-NC-SA 4.0](#).
- [1.2: The Statistical Analysis Process](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).