Online Statistics Education: An Interactive Multimedia Course of Study

Developed by Rice University (Lead Developer), University of Houston Clear Lake, and Tufts University

http://onlinestatbook.com/2/summarizing_distributions/variability.html

Measures of Variability

Authored by David M. Lane

What is Variability?

Variability refers to how spread out or dispersed the data are.

To see what we mean by spread out, consider the graphs in Figure 1. Each graph represents the scores on one of two quizzes, Quiz 1 and Quiz 2.

The mean score for each quiz is 7.000. However, even though the two quizzes have the same mean score, you can see that the distributions are quite different. The students' scores on Quiz 1 are more clustered together, whereas the students' scores on Quiz 2 are more spread out. The students' scores on Quiz 1 are less variable than the students' scores Quiz 2, even though the mean score is the same on the two quizzes.

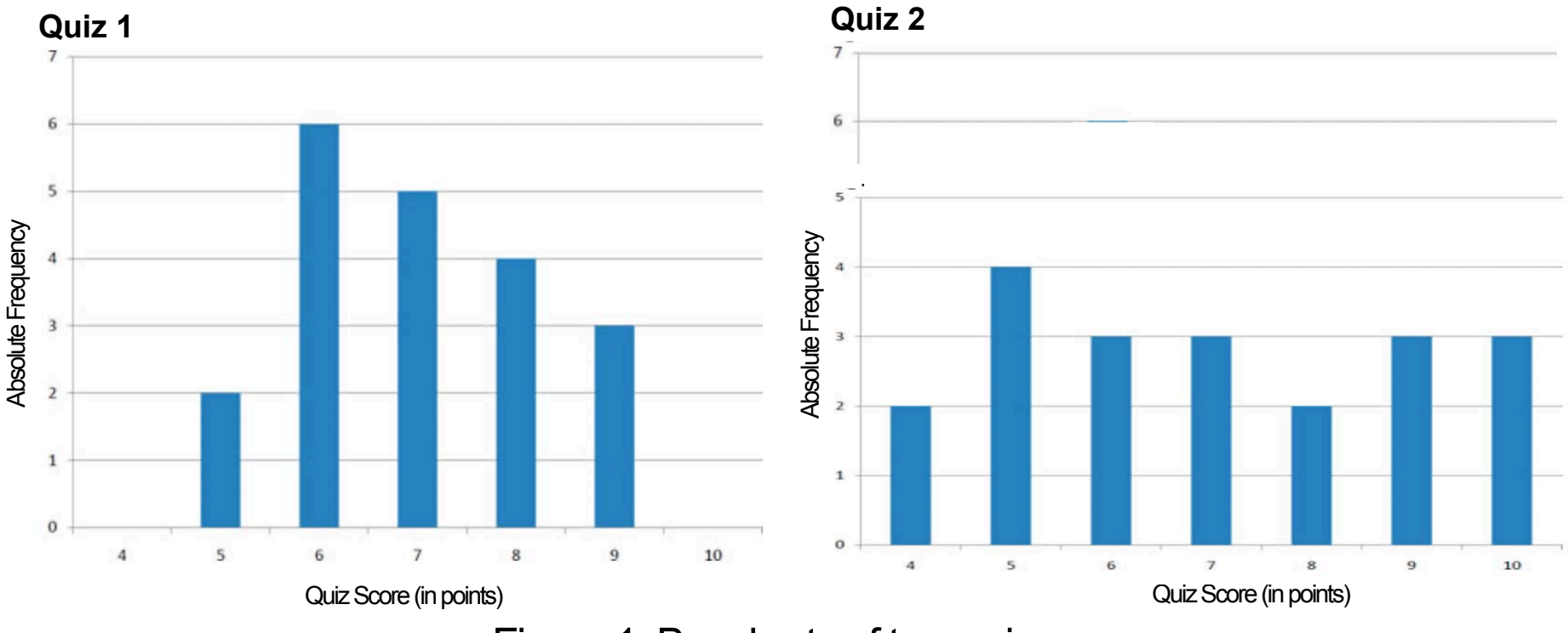

Figure 1. Bar charts of two quizzes.

The terms variability, spread, and dispersion are synonyms and refer to how variable a set of data is. Three frequently used measures of variability are the **range**, the **variance**, and the **standard deviation**.

Range

The Range is the simplest measure of variability, and one you have probably encountered many times in your life. The Range is simply the highest value in the data set minus the lowest value in the data set.

For example, what is the Range of the following set of data? 10, 2, 5, 6, 7, 3, 4

The highest value is 10, and the lowest value is 2. Because 10 minus 2 = 8, the Range for this data set is 8.

On Quiz 1, shown in Figure 1, the lowest value is 5, and the highest value is 9. Therefore, the Range on Quiz 1 is 4. On Quiz 2, the lowest value is 4, and the highest value is 10. Therefore, the range on Quiz 2 is 6.

The larger the range, the more spread out the data are; the smaller the range, the more clustered the data are.

Variance

Variability can also be defined in terms of how close the values are to the central tendency. We can use the mean as the measure of central tendency.

The variance is defined as the average squared difference of each value from the mean.

The data from Quiz 1 are shown in Table 1. The mean score is 7.000. Therefore, the column "Deviation from Mean" contains each score minus 7.

The column "Squared Deviation" is simply the Deviation from the Mean squared.

Table 1. Calculation of Variance for Quiz 1 scores.

Scores	Deviation from Mean	Squared Deviation
9	2	4
9	2	4
9	2	4
8	1	1
8	1	1
8	1	1
8	1	1
7	0	0
7	0	0
7	0	0
7	0	0
7	0	0
6	−1	1
6	−1	1
6	−1	1
6	−1	1
6	−1	1
6	−1	1
5	−2	4
5	−2	4
Means		
7	0	1.5

In this example, the mean of the Squared Deviations is 1.500. Therefore, the Variance of Quiz 1 is 1.500.

The more tightly the data values cluster around the mean, the lower the Variance will be. The lower the Variance, the less spread out the data will be.

The Variance of Quiz 2 is 6.700. Because the Variance of Quiz 1 is 1.500, and the Variance of Quiz 2 is 6.700, we know that, on average, the scores on Quiz 1 are more tightly clustered around the mean of Quiz 1, whereas, on average, the scores on Quiz 2 are more dispersed.

Standard Deviation

The Standard Deviation is simply the square root of the Variance. For Quiz 1, because the Variance is 1.500, the Standard Deviation is 1.225 (which is the square root of 1.500).

The Standard Deviation is an especially useful measure of variability when the distribution is normal ("bell shaped") or approximately normal because the Standard Deviation affects the shape of the "bell curve."

Figure 2 shows two normal distributions. The red distribution has a mean of 40.000 and a standard deviation of 5.000; the blue distribution has a mean of 60.000 and a standard deviation of 10.000. As you can see, the red distribution is more clustered together -- its bell shape is more peaked, whereas the blue distribution is more spread out -- its bell shape is flatter.

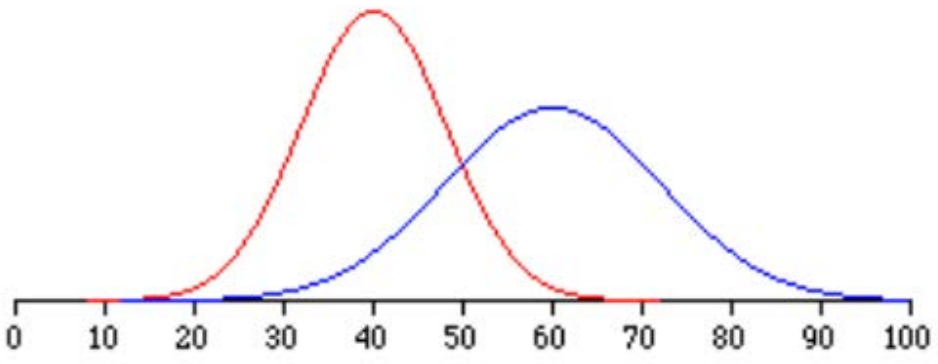

Figure 2. Normal distributions with standard deviations of 5 and 10.

Figure 2. Normal distributions with standard deviations of 5 and 10.

NOTE: This article has been modified.