Statistics, Day 5

Objective: To compare the shapes, center and spreads of two distributions; To determine if a distribution is symmetric or skewed

What does the standard deviation tell you about a data set?

If two data sets have the same mean, the data with the larger standard deviation is more spread out.

Example: A company that makes raisin bread advertises that there are at least 10 raisins in every slice of bread. Sample loaves are tested 20 times per day, and a slice of bread from each loaf is examined. If there are fewer than 8 raisins in the slice, the loaf cannot be sold. Examine the following data to determine which baker is doing a better job.

Baker 1 Baker 2

6 12 8 13 15 10 12 8 13 10

7 11 9 15 10 6 12 10 9 7

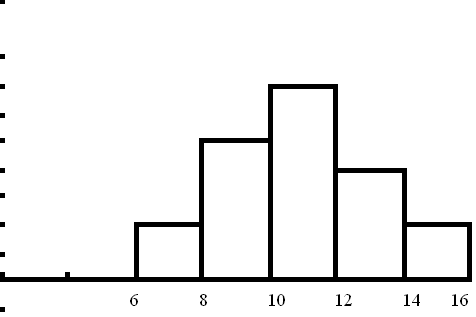
13 5 7 11 13 14 11 10 9 11

13 10 12 3 12 10 12 8 9 14

Distributions are frequently described by their shape.

Look at the histogram for Baker 2. The graph is relatively symmetric. For a symmetric distribution, the mean and the median are relatively the same.

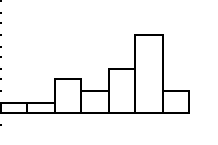


Baker 2

= 10.25

Median = 10

s = 2.17

Now look at the histogram for Baker 1. This graph is called skewed left (or negatively skewed) because there is a long tail of data on the left. For this data, the mean is less than the median because it is being pulled down (to the left) by the lower numbers.

Baker 1

= 10.25

Median = 11

s = 3.34

The shapes of the distributions can also be compared using box plots. Your calculator can make up to three box plots stacked on top of each other.

Example: Describe the shape of the following set of data

75, 104, 72, 71, 95, 69, 75, 73, 77, 76, 90

What are the values of the mean and the median?

Which would be a better measure of center for the distribution?

Are any of the data points outliers?