

# 2.5: Standard Deviation (4 of 4)

# Learning Objectives

· Use mean and standard deviation to describe a distribution.

# Deciding Which Measurements to Use

We now have a choice between two measurements of center and spread. We can use the median with the interquartile range, or we can use the mean with the standard deviation. How do we decide which measurements to use?

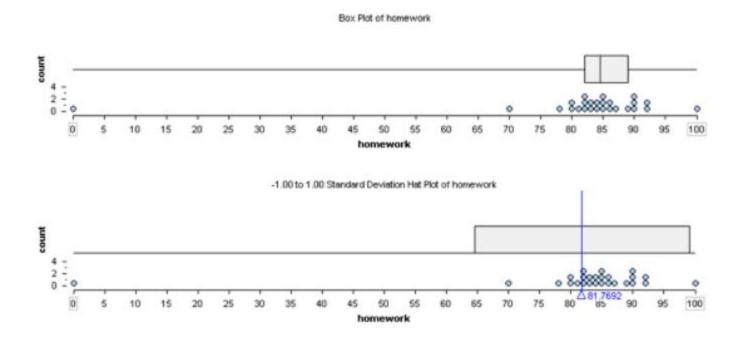
Our next examples show that the shape of the distribution and the presence of outliers helps us answer this question.

## Example

### Homework Scores with an Outlier

Here are two summaries of the same set of homework scores earned by a student: a boxplot and an SD hatplot. Notice that the distribution of scores has an outlier. This student has mostly high homework scores with one score of 0. Here are some observations about the homework data.

- Five-number summary: low: 0 Q1: 82 Q2: 84.5 Q3: 89 high: 100
- Median is 84.5 and IQR is 7
- Mean = 81.8, SD = 17.6



The typical range of scores based on the first and third quartiles is 82 to 89.

The typical range of scores based on Mean  $\pm$  SD is 64.2 to 99.4 (Here's how we calculated this: 81.8 - 17.6 = 64.2, 81.8 + 17.6 = 99.4.)

Which is the better summary of the student's performance on homework?

The typical range based on the mean and standard deviation is not a good summary of this student's homework scores. Here we see that the outlier decreases the mean so that the mean is too low to be representative of this student's typical performance. We also see that the outlier increases the standard deviation, which gives the impression of a wide variability in scores. This makes sense because the standard deviation measures the average deviation of the data from the mean. So a point that has a large deviation from the mean will increase the average of the deviations. In this example, a single score is responsible for giving the impression that the student's typical homework scores are lower than they really are.

The typical range based on the first and third quartiles gives a better summary of this student's performance on homework because the outlier does not affect the quartile marks.

### Example

# **Skewed Incomes**

In this example, we look at how skewness in a data set affects the standard deviation. The following histogram shows the personal income of a large sample of individuals drawn from U.S. census data in the year 2000. Notice that it is strongly skewed to the right. This type of skewness is often present in data sets of variables such as income.