

# Measures of Spread

## Reading

- Section 2.7

## Practice Problems

**2.7 (Page 134)** 69, 70, 72

**2.7 (Page 143)** 94, 95, 97, 98, 99, 100, 101

**Overall 2 (p. 145)** 109, 110, 113, 114, 115

## Notes

- The goal of measures of spread is to assess the “variation” in the data in some way.
- Different measures achieve this in different ways.
- The two main measures of spread are:

**Std. Dev.** Standard Deviation. Measures distance of data from the mean.

Often denoted by  $s$ .

Not resistant.

**IQR** Interquartile range. Measures range occupied by the middle 50% of the data.

Resistant.

- Let us review how the standard deviation is computed. Its formula is a bit complicated, and it looks something like this:

$$s = \sqrt{\frac{1}{n-1} \sum (x - \bar{x})^2}$$

Basically:

- Compute the mean of all the values.
- Subtract the mean from each value (the result of this step is the “**deviations**”).
- Square all the deviations.
  - \* Ensures values are positive before computing average.
- Average these squared deviations: Add them all up, then divide by  $n - 1$ . The result at this stage is called the **variance**.
  - \* Why  $n - 1$ : Technical reason, and won't really matter for large  $n$ .

- \* One way to think about it: The deviations always add up to 0, so once you know  $n - 1$  of them the last one is determined.
  - \* In this context,  $n - 1$  is called the “degrees of freedom”.
- Take a square root at the end.
  - \* Fixes the units of measurement.
- Outliers have a considerable effect on this formula.
  - They have a very large deviation, and because they pull the mean towards themselves they cause larger deviations in the other values as well.
  - Because we square them before adding, those large deviations will dominate the equation even more.
- What to use depends on the distribution:
 

**Symmetric** Mean for center, Standard Deviation for spread  
**Skewed/Outliers** Median for center, IQR for spread
- Chebyshev’s Rule:
  - At least 75% of the data is within 2 standard deviations from the mean
  - At least 89% of the data is within 3 standard deviations from the mean
- For Bell Shaped data (Empirical Rule):
  - Approximately 68% of the data is within one standard deviation of the mean.
  - Approximately 95% of the data is within two standard deviations of the mean.
  - More than 99% of the data is within three standard deviations of the mean.