





- The spread of a distribution refers to how much **variability** the data has.
- Common measures are the **sample variance** and **sample standard deviation**.
- A very basic measure is to look at just the **minimum** and **maximum** values.
 - The sample *range* is the difference between these two.



Sample Standard Deviation

• The standard deviation s is the square root of the variance.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$



Important Facts About s² and s

- Both variance and standard deviation will **ALWAYS** be non-negative values.
 - s=0 means that all observations have the same value.
- The units of **variance** are the square of the units of the data.
- The units of **standard deviation** are the units of the data.
- Both are **HEAVILY** affected by outliers.

Proposition

- Let *x*₁,..., *x*_n be a sample and *c* be a non-zero constant.
- If $y_1 = x_1 + c$, ..., $y_n = x_n + c$, then $s_y = s_x$
- If $y_1 = cx_1, ..., y_n = cx_n$, then $s_y = |c| s_x$

Three Standard Deviations

- Almost all observations for **any** data set lie within three standard deviations of the mean.
- Chebyshev's rule
 - At least 89% of observations fall within three standard deviations of the mean

The Empirical Rule

- If the distribution is roughly bell shaped, approximately:
 - 68% of observations fall within one standard deviation of the mean.
 - 95% of observations fall within two standard deviations of the mean.
 - 99.7% of observations fall within three standard deviations of the mean.

