Measures of Spread

Summary

1. Measures of spread, such as standard deviation, tell us how spread out the data is.
2. Standard deviation uses the same units as the values in the data set.

Range

Range

The **range** of a dataset is found by subtracting the minimum value from the maximum value.

Example 1. During a heat wave one summer, I decided to cool off by drinking milkshakes everyday for a week. The number of milkshakes I had each day is shown:

9, 2, 7, 10, 3, 4, 12

Find the range for the number of milkshakes I drank that week.

A disadvantage of relying solely on the range as a measure of variation is that it is heavily affected by outliers (extreme values).

Variance

Deviation from the Mean

The **deviation from the mean** refers to how far a data value, x, is from the mean; found by

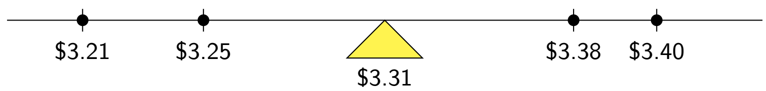
$$x - \text{mean}$$

Deviation can be positive, negative, or zero.

Example 2. Find the mean of the following gas prices:

\$3.25, \$3.40, \$3.21, \$3.38

We can think of the mean as a "balancing point" for our data set:



Each data point has a deviation (or *distance*) from the mean.

Example 3. Calculate each data point's deviation from the mean.

$3.25, $3.40, $3.21, $3.38 (Mean: $3.31)

Example 4. Find the mean of the deviations from the previous example.

- Positive deviations will always "cancel out" negative deviations.
- Two possible remedies:
  - Take absolute value of deviations
  - Square the deviations
- Absolute values will lead us to *mean absolute deviation*.
- For future calculations, squaring is the better choice.

Example 5. Square each of the deviation gas prices and then find the mean of the squared deviations.

The result of the previous example is known as the **population variance** and is denoted by

$$\sigma^2$$

Like the mean, variance also has a **sample variance** and is denoted

$$s^2$$

Population vs. Sample Variance

- Sample variance $\longrightarrow$ divide by *one less* than then number of observations (**degrees of freedom**).
- Sum of the deviations **must** be 0.
  - In a data set with 4 entries, the first 3 entries can be any number we want.
  - The final value must make it so that the sum of the deviations from the mean equals 0.

Population Variance	**Sample Variance**
$\sigma^2 = \frac{\sum(x_i-\mu)^2}{N}$	$s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$

Note: Variance always gives us squared units of measurement.

Standard Deviation

The **standard deviation** is the square root of variance.

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$

$$\sigma = \sqrt{\frac{\sum(x_i-\mu)^2}{N}} \quad \text{and} \quad s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$

Properties of Standard Deviation:

- It's a measure of how much (on average) the data values deviate from the mean.
- Can be positive or 0 only.
- Like the mean, it can be greatly affected by extreme values (outliers).
- The units are the same as those in the data set.
- Less and less differences between population and sample s.d. with larger sample sizes.

Example 6. Find population AND sample standard deviations of the gas prices.

Usually, data values will be within 2 standard deviations of the mean.

Example 7. The mean price of gas one day was \$3.58 with a standard deviation of \$0.33. What interval would represent a "usual" price of gas?