10.3: Measures of Dispersion

The mean, median, and mode can describe the "middle" of a data set, but none of them can describe how "spread out" the data is.

Range:

The *range* for ungrouped data is the difference between the largest and smallest values.
The *range* for grouped data is the difference between the upper boundary of the highest class and the lower boundary of the lowest class.

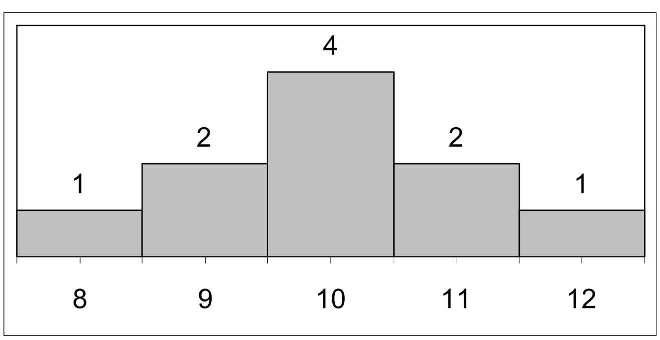
Example 1: Find the range.

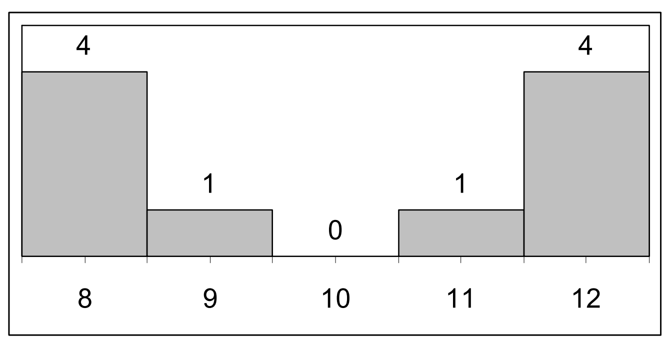
Commute Times							
0.3	0.7	0.2	0.5	0.7	1.2	1.1	0.6
0.6	0.2	1.1	1.1	0.9	0.2	0.4	1.0
1.2	0.9	0.8	0.4	0.6	1.1	0.7	1.2
0.5	1.3	0.7	0.6	1.1	0.8	0.4	0.8

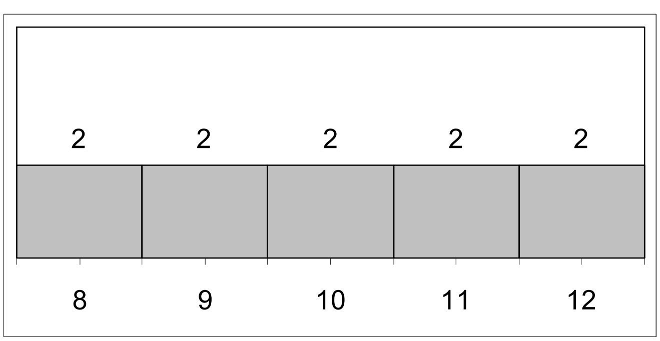
Example 2: Find the range for the grouped data.

Interval	Frequency
1.5–4.5	3
4.5–7.5	4
7.5–10.5	7
10.5–13.5	2

Look at these histograms…







In all:

Mean =

Median =

Range =

While the range is useful, it is dependent on the extreme values of the data set. It doesn't tell you whether the data is close to the mean, far from the mean, or evenly distributed. We need a better measure of dispersion.

Deviation of a data point:

The deviation of a data point is the difference (i.e., the signed distance) between the data point and the mean.

In other words, the deviation of the *i*th data point, x_i is $x_i - \mu$.
(Note that the deviation is positive if $x_i > \mu$; the deviation is negative if $x_i < \mu$.)

Let's average the deviations for a data set.

<u>**Example 3:**</u> $A = \{12, 13, 7, 5, 9\}$

Variance of a population:

<u>Variance of a population</u>:

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for an entire population, with population mean μ. Then the *population variance* σ^2 is

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}.$$

In other words, the variance is the average (mean) of the squared deviations.

<u>Alternative formula for the population variance</u>:
(sometimes known as the computational formula, computing formula or shortcut formula)

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}$$

Example 4: Calculate the variance of the population $A = \{12, 13, 7, 5, 9\}$.

Degrees of freedom:

The quantity known as *degrees of freedom* is the number of scores (data points) in a dataset that are free to vary in the presence of a statistical estimate.

If a sample has *n* data points and the sample mean $\bar{x}$ is specified, then $n-1$ of the data points can theoretically be anything; the *n*th data point is forced to be take on whatever value results in the specified mean $\bar{x}$. In other words, the first $n-1$ of the data points are free to vary; the *n*th data point is not free to vary.

Example 5: Suppose a sample has 5 data points and a mean of 159. Suppose also that the first four data points are 37, 203, 122, and 303. Calculate the fifth data point.

Variance of a sample:

When we calculate the variance of a *sample* (not the whole population), we have no way to calculate the population mean. Therefore, we must use the sample mean (denoted $\overline{x}$) as an estimate of the population mean (denoted μ). Thus, in a sample of *n* data points with sample mean $\overline{x}$, there are $n-1$ degrees of freedom.

When calculating the variance for a sample (not the entire population), we divide by $n-1$ (the degrees of freedom) instead of *n*. Dividing by *n* would underestimate the variance, because the points in the sample will be less spread out than those in the population. Using the degrees of freedom, $n-1$, in the denominator provides an unbiased estimate of the population variance.

Variance of a sample:

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for a sample taken from a larger population, with sample mean $\overline{x}$. Then the *sample variance* s^2 is

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

Alternative formula for the sample variance:
(sometimes known as the computational formula, computing formula or shortcut formula)

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$$

Example 6: Calculate the variance of this sample: {75, 16, 50, 88, 79, 95, 80}.

Standard deviation:

By squaring the deviations, we've changed the units (if there are any). In other words, if we started with "inches", we now have "square inches". This is easily fixed by taking square roots.

<u>Standard Deviation</u>:

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for a sample taken from a larger population, with sample mean $\overline{x}$. Then the *sample standard deviation* s is

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

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for an entire population, with population mean μ. Then the *population standard deviation* σ is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}\ .$$

<u>Example 7:</u> Given the following data sample, calculate the standard deviation to two decimal places.

{70, 39, 54, 84, 68, 93, 75}

<u>Summary: Variance and Standard Deviation (ungrouped data)</u>:

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for a sample taken from a larger population, with sample mean $\overline{x}$.
Then the *sample variance* s^2 and the *sample standard deviation* s are:

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

Suppose the set of values $\{x_1, x_2, \ldots, x_n\}$ are measurements for an entire population, with population mean μ.
Then the *population variance* σ^2 and *population standard deviation* σ are:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \quad \text{and} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

<u>IMPORTANT</u>:

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

$$\left(\text{Standard Deviation}\right)^2 = \text{Variance}$$

Example 8: Suppose a sample has standard deviation 38.92. What is the variance?

Example 9: Suppose the variance for a sample is 142.4. What is the standard deviation?

Variation and standard deviation for grouped data:

Variance (grouped data):

Suppose a data set of n sample measurements is grouped into k classes in a frequency table, where x_i is the midpoint and f_i is the frequency of the ith class interval.

The *sample variance* s^2 for the grouped data (with mean $\overline{x}$) is approximated by:

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n-1} \text{ or, equivalently, } s^2 = \frac{\sum_{i=1}^{k} f_i x_i - n\overline{x}^2}{n-1}.$$

Standard Deviation (grouped data):
The *sample standard deviation* s for the grouped data (with mean $\overline{x}$) is approximated by:

$$s = \sqrt{\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2 f_i}{n-1}} \text{ or, equivalently, } s = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i - n\overline{x}^2}{n-1}}.$$

$n = \sum_{i=1}^{k} f_i$ is the total number of measurements.

Example 10: Suppose the table below summarizes the total wait times for a sample of drive-through visits at a fast-food restaurant Find the mean, median, variance and standard deviation for the sample.

Service Time	Frequency
-0.05–1.45	13
1.45–2.95	24
2.95–4.45	29
4.45–5.95	23
5.95–7.45	26
7.45–8.95	14
8.95–10.45	9
10.45–11.95	8
11.95–13.45	3
13.45–14.95	3

Service Time	Frequency
-0.05–1.45	13
1.45–2.95	24
2.95–4.45	29
4.45–5.95	23
5.95–7.45	26
7.45–8.95	14
8.95–10.45	9
10.45–11.95	8
11.95–13.45	3
13.45–14.95	3