11.3: 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.

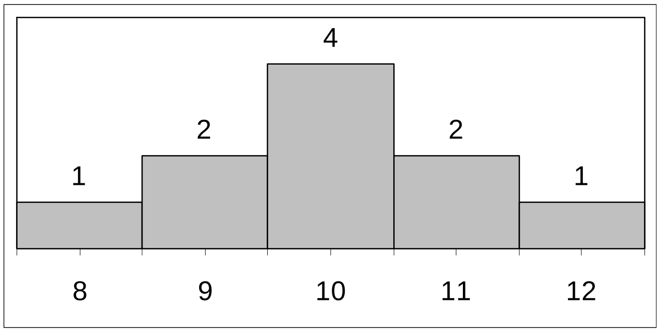
Example 1: Find the range.

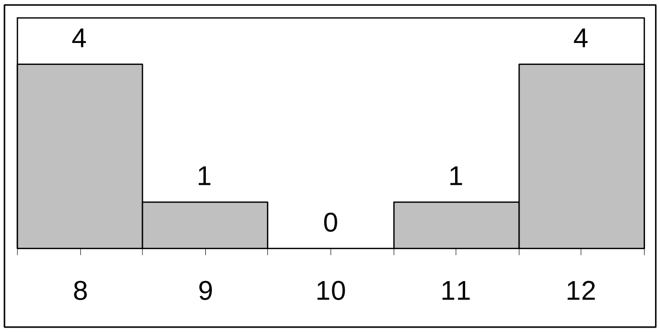
Commute Times							
0.3	0.7	0.2	0.5	0.7	1.2	1.1	0.6
0.6	0.2	1.1	1.1	0.9	0.2	0.4	1.0
1.2	0.9	0.8	0.4	0.6	1.1	0.7	1.2
0.5	1.3	0.7	0.6	1.1	0.8	0.4	0.8

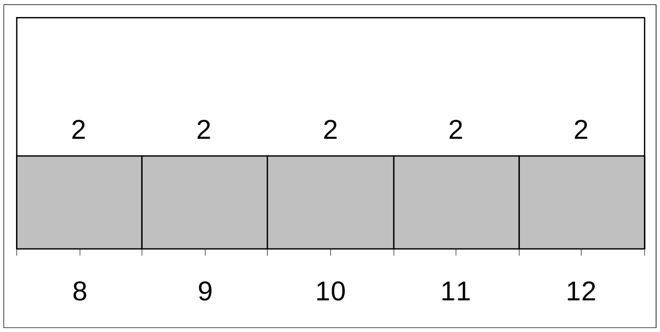
Example 2: Find the range for the grouped data.

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 something else.

Variance and standard deviation:

Variance (ungrouped data):

The *sample variance* s^2 of a set of *n* sample measurements $x_1, x_2, \ldots, x_n$ with mean $\overline{x}$ is given by

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

If $x_1, x_2, \ldots, x_n$ is the whole population with mean μ, then the *population variance* σ^2 is given by

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

Note: For samples (not the entire population), we divide by $n-1$ instead of *n*. Statisticians have found that this gives a better estimate of the population parameters, especially when the sample is relatively small. This has to do with the concept of *degrees of freedom*, which you'll study in more detail if you take an introductory statistics class.

Example 3: Consider this data set: {75, 16, 50, 88, 79, 95, 80}.

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.

Standard Deviation (ungrouped data):

The *sample standard deviation* s of a set of *n* sample measurements $x_1, x_2, \ldots, x_n$ with mean $\overline{x}$ is given by

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

If $x_1, x_2, \ldots, x_n$ is the whole population with mean μ, then the *population standard deviation* σ is given by

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

Example 4: Given the following data sample, calculate the standard deviation to two decimal places.

{7, 6, 10, 8, 9, 5, 9}

IMPORTANT:

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

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 *i*th class interval.

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

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

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

Standard Deviation (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 *i*th class interval.

Then the *sample standard deviation* s for the grouped data (with mean $\overline{x}$) is given by

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

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

Example 5: Suppose that the price-earning ratios of some randomly selected stocks are given in the following table. Find the mean, variance and standard deviation for the sample.

Interval	Frequency
-0.5-4.5	7
4.5-9.5	53
9.5-14.5	22
14.5-19.5	14
19.5-24.5	0
24.5-29.5	4
29.5-34.5	3