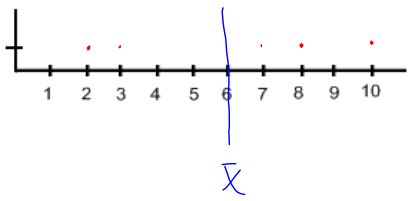
Section 1.2 Summary Statistics

Summarize the data we sampled so we can describe and draw conclusions about the population.

Pay close attention to notation and know what symbols are used for which statistics.

Sample Observations: 2, 7, 3, 8, 10

Sample Mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+7+3+8+10}{5} = \frac{30}{5} = 6$



$\bar{x}$ is used for the sample mean

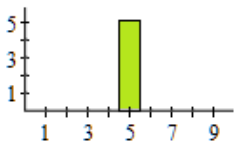
μ is used for the population mean

The Mean is a measure of "central tendency."
It is also the center of mass or the balance point of a set of data.
We also need a way to quantify the spread of the data.

This is where the variance comes in.

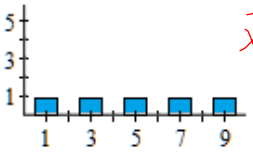
Measure of Spread or Variability:

Sample 1: 5, 5, 5, 5, 5



$\bar{x} = 5$

Sample 2: 1, 3, 5, 7, 9



$\bar{x} = 5$

Sample Variance:

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

$$s^2 = \frac{1}{4}\left[16 + 4 + 0 + 4 + 16\right]$$

$$= 10 \text{ units}^2$$

An easier way to calculate the variance when you have to do it by hand:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

Sample 2: 1, 3, 5, 7, 9

$= \frac{1}{4}\left[165 - 125\right]$

$= \frac{1}{4}(40) = 10 \text{ units}^2$

Yet another option...

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	-4	16
3	-2	4
5	0	0
7	2	4
9	4	16
		$\Sigma = 40$

$s^2 = \frac{40}{4} = 10 \text{ units}^2$

$s = \sqrt{10} = 3.16 \text{ units}$

Drawback of using the Sample Variance:

The units are inconvenient.

Sample Standard Deviation:

Or in general, n.

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{5}(x_i - \bar{x})^2}$$

σ = population standard deviation.

Reconsider Sample 2 and use your calcuator:

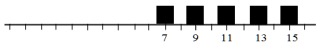
1, 3, 5, 7, 9 $\quad \bar{x} = 5$

$s_x^2 = 10$

$s_x = \sqrt{10} \approx 3.16$

How will the mean and variance change if we add 6 to each data value?

7, 9, 11, 13, 15 $\quad \bar{y} = \bar{x} + 6$



$s_y^2 = s_x^2$

$s_y = s_x$

In General:

$\bar{y} = \bar{x} + c$

$s_y^2 = s_x^2$

$s_y = s_x$

Reconsider Sample 2:

$\bar{x} = 5$

1, 3, 5, 7, 9

$s_x^2 = 10$

$s_x = \sqrt{10}$

How will the mean and variance change if we multiply each value by 2?

2, 6, 10, 14, 18

$\bar{y} = 2\bar{x}$

$s_y^2 = 4s_x^2$

$s_y = 2s_x$

In General:

$\bar{y} = c\bar{x}$

$s_y^2 = c^2 s_x^2$

$s_y = |c| s_x$

*Note: Because spread in general is positive (it is a distance after all) note that we should have absolute values in the equation for standard deviation.

Measures of Center and Measures of Variation

Median

3, 6, 8, 10, 12, 20, 28, 34, 40

Median = 12

$.5(n+1) = .5(10) = 5^{th}$ position

$\Rightarrow 12$.

3, 6, 8, 10, 12, 20, 28, 34

Median: 11

$.5(n+1) = .5(9) = 4.5$

Avg of #'s in the 4th and 5th position.

$\Rightarrow 11$.

Quartiles and Percentiles

1st quartile: 25th percentile!

$.25(n+1) = .25(10) = 2.5$

Avg #'s in the 2nd and 3rd position,

$\Rightarrow$ 25th percentile

OR

1st quartile.