MATH 1342

Section 1.3

Another important question we want to answer about data is about its spread or dispersion. Roughly speaking, the **population standard deviation**, σ, tells the average distance that data values fall from the mean. The standard deviation is the square root of the **population variance**, σ^2. So, what is the variance? The variance is the average of the squared differences of the data values from the mean.

If N is the number of values in a population with mean μ, and x_i represents each individual value in the population, then the variance is found by:

σ : population standard deviation

σ^2 : population variance

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

And the population standard deviation is $\sigma = \sqrt{\sigma^2}$

Standard deviation is a measure that tells how far data points (on average) from the center point.
Large standard deviations mean that the data is more spread out (distance from the mean is greater)
Small stardard deviations mean that data is more concentrated around the mean (distance from the mean is lesser)

Most of the time we are not working with the entire population. Instead, we are working with a sample.

- Sample variance - $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$
- Sample standard deviation - $s = \sqrt{s^2}$

s: sample stnd dev

s^2: sample variance

standard deviation = $\sqrt{\text{variance}}$

Example:

1. A statistics teacher wants to decide whether or not to curve an exam. From her class of 300 students, she chose a sample of 10 students and their grades were:

72, 88, 85, 81, 60, 54, 70, 72, 63, 43

Find the mean, variance and standard deviation for this sample.

```
> assign("grades",c(72, 88, 85, 81, 60, 54, 70, 72, 63, 43))
> mean(grades)
[1] 68.8
> var(grades)
[1] 199.7333
> sd(grades)
[1] 14.1327
> sqrt(199.7333)
[1] 14.1327
> 14.1327^2
[1] 199.7332
```

$\sqrt{\text{variance}}$

$(\text{snd dev})^2$

To copy: 72, 88, 85, 81, 60, 54, 70, 72, 63, 43

In R Studio:

var(list) will give the variance

and

sd(list) will give the standard deviation.

2. Suppose the statistics teacher decides to curve the grades by adding 10 points to each score. What is the new mean, variance and standard deviation?

Theory: Adding 10 points to every score will increase the mean.
Since the variance and standard deviation of measures of dispersion (distance from the center) adding 10 points to each value will not change these measures.

```
> newGrades=grades+10
> newGrades
 [1] 82 98 95 91 70 64 80 82 73 53
> mean(newGrades)
[1] 78.8
> var(newGrades)
[1] 199.7333
> sd(newGrades)
[1] 14.1327
```

To summarize: adding or subtracting to every data point in your set will have no influence on the variance or the standard deviation.

We can see from example 2 that adding the same value to all elements does not affect the variance (or standard deviation) of a set of data. What about multiplying?

3. Find the variance and the standard deviation for the following set of data (whose mean is 4.5)

3, 6, 2, 7, 4, 5

To summarize: multiplying or dividing all your data points by the same value will have an influence on the mean, standard deviation, and variance.

Now, multiply each value by 2. What is the new variance and the new standard deviation?

```
> assign("data",c(3,6,2,7,4,5))
> mean(data)
[1] 4.5
> var(data)
[1] 3.5
> sd(data)
[1] 1.870829
```

```
> newData=2*data
> mean(newData0)
Error in mean(newData0) : object 'newData0' not found
> mean(newData)
[1] 9
> var(newData)
[1] 14
> sd(newData)
[1] 3.741657
```