

## Administrivia:

1. Class roster...initial it!
2. Solms206
3. Review

### INDEPENDENCE (in the statistical sense)

#### a. Sex and Last digit of phone number

Phone\Sex	M	F
0-6		
7-9		

#### b. Shoe size and Height

Hgt\Shoe	Small, 7- or 9-	Large
Short 5'2"- 5'7"-		
Tall		

## Stem and Leaf

counts=frequency

0 0336	4	
10 022222333345566789	18	22
20 112689	6	
30 0002223445666789	16	22
40 1133445566678888	16	
50 0056	4	20
60 00012223344588888	17	
70 022336	6	23
80 38	2	
90 26	2	4
100 04	2	
110 5	1	3
120 0	1	
130	0	2
140 0	<u>1.</u>	
	96	

Note: Skip box-plots

Q. Why the ‘see-saw’-ing of counts?

DEFN: The case=subject=participant who marks the 80<sup>th</sup> percentile is the one who is 80/100 above the bottom of the sorted data.

EX1. In our 96 EX2. The 40<sup>th</sup> percentile

EX3. In those with 59 hours or less

Q. How would you represent the ‘center’?

DEFN: The median is the subject who marks the 50<sup>th</sup> percentile (or 2<sup>nd</sup> quartile).

EX1. In our 96

EX2. In our 96—1<sup>st</sup> and 3<sup>rd</sup> quartile

Q. What’s a decile?

Note: Be aware of different methods (p.67)

Q. Where would the balance point be for our stem and leaf?

DEFN: The mean (or average or expected value) of a data set is the sum of all values divided by the number of values.

Q. What is the label for our 96?

$$\sum_{i=1}^N x_i / N$$

Q. Must someone in the set have the exact value of the mean?

EX1. Mean of our 96

EX2. Mean of 5 randomly selected

Q. Should our mean = median? Can this happen?  
Does 140 influence the mean or median more?

Summary of hours	
<b>Count</b>	96
<b>Mean</b>	45.323
<b>Median</b>	43.5
<b>Std Dev</b>	27.996
<b>Variance</b>	783.758
<b>Range</b>	140
<b>Min</b>	0
<b>Max</b>	140
<b>IQR</b>	41.5
<b>25th%</b>	21.5
<b>75th%</b>	63

Note: I use the mean as a global statistic and the median as a personal statistic. What happens to our 96 if I add 4 students with 140 credit hours?

▷ | Summary of hours

<b>Count</b>	100
<b>Mean</b>	49.11
<b>Median</b>	44.5
<b>Std Dev</b>	33.163
<b>Variance</b>	1099.776
<b>Range</b>	140
<b>Min</b>	0
<b>Max</b>	140
<b>IQR</b>	40.5
<b>25th%</b>	24
<b>75th%</b>	64.5

Q. How could you describe the ‘spread’ of our credit hours? Drawback of range = largest minus smallest (or maximum – minimum)

XXXXXXXXXXXXXXXXX  
 XXXXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXX140

XXXXXX  
 XXXXXX  
 XXXXXX  
 XXXXXX  
XXXXXXXXXXXXXXXXX140

Q. How about a number that tells on the average how far the set of credit hours is from the center?

DEFN: The variance is the average squared distance of each value from its mean.

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

Notation:  $s^2$  or  $\sigma^2$   
 Q. Label?  
 EX1. Use a computer  
 EX2. Our random 5

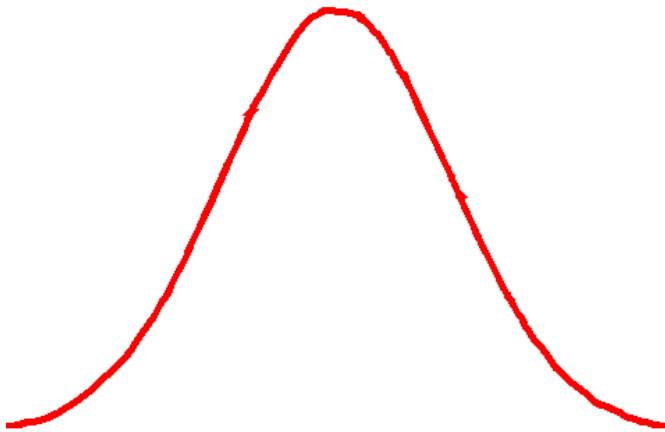
DEFN: The standard deviation is the square root of the variance. Notation:  $s$  or  $\sigma$

Q. What happens to the mean and variance if I add 10 to each value?

Q. What happens to the mean and variance if I double each value?

Q. Should we think of 140 hours as an outlier?

Fact: If a set of data forms a stem and leaf that looks bell-shaped we may be able to refer to it as normally distributed data.



1.

Symmetric

2. Mean=Median

3. symbol  $N(\mu, \sigma)$

4.

68% of data is

within one standard deviation of the mean

5. 95% of the data is within two standard deviations of the mean

6. 99.7% of the data is within three standard deviations of the mean

EX1. Is 140 credit hours an outlier?

How many standard deviations away from the mean credit hours is it?

DEFN: A z-score is computed via  $(\text{value} - \text{mean}) / \text{st.dev.}$  It tells how many standard deviations a value is away from the mean. Note: negative indicates below or less than and positive shows it above or greater than the mean.

Q. Let us assume that credit hours for all Math2200 students have our mean and standard deviation and follow a normal distribution. How many students in Dr. K's class of 40 will have

- more than 101 hours?
- between 17 and 101 hours?
- less than 10 hours?