DATA

$180
$240
$245
$360
$375
$425
$455
$460
$465
$470
$480
$485
$495
$505
$525
$530
$550
$640
$710
$720

DATA ANALYSIS Practice Test #2

#1

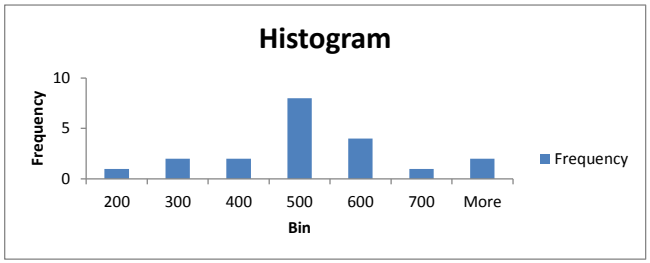
Create a histogram using a bin width of 100 ($1,000s).

Step One: Create A List of Bin Upper Limits. Use the upper bound for each bin. For instance, use 200 for the bin "$200 and under"; 300 for the bin $201-$300 etc.
Note that you do not need to worry about decimals as all values have been rounded to the nearest integer.
Note that Excel will always add one additional bin to account for values above the last bin limit listed. For instance, it will add a bin for "more than 700" below.

200
300
400
500
600
700

Step Two: Create a Frequency Table & draft Histogram using the Data Analysis Toolpak.
Be sure to check off the "Chart Ou

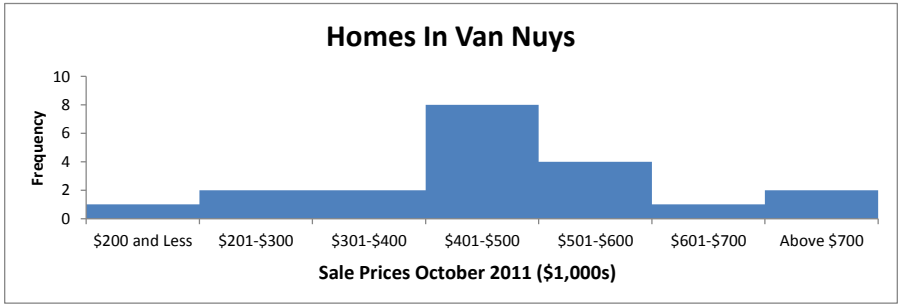
Bin	*Frequency*
200	1
300	2
400	2
500	8
600	4
700	1
More	2



Step Three: Now perform the following modifications.

A. Delete the "frequency" legend.
B. Change generic "Bin" to name of variable being measured on the X-axis ("Sale Prices October 2011")
C. Add the Unit of Measure to the horizontal axis (X-axis) - "$1,000s"
D. Change the bin labels from a single upper limit value to reflect full range of bin, and add $. (300 becomes $201-$300).
E. Change generic chart title "Histogram" to describe the population ("Homes In Van Nuys")
F. Adjust the bars so there is no gap between them.

Bin	*Frequency*
$200 and Les	1
$201-$300	2
$301-$400	2
$401-$500	8
$501-$600	4
$601-$700	1
Above $700	2



#2-#3

For these questions, you need to use the "Descrptive Statistics" in the Data Analysis Toolpak; and the formula QUARTILE.EXC.

Be sure to check the "Summary Statistics" box.

Round all figures to the nearest tenth.

Know that mean and median measure central tendancy.

Know that standard deviation, interquartile range (IQR), and range measure spread.

Column1	
Mean	465.8
Standard Error	31.4
Median	475.0
Mode	#N/A
Standard Deviation	140.5
Sample Variance	19734.9
Kurtosis	0.3
Skewness	-0.2
Range	540.0
Minimum	180.0
Maximum	720.0
Sum	9315.0
Count	20.0

Q1 = 387.5

Q3 = 528.75

IQR = Q3 - Q1 141.25

#4 Skew statistics measure the shape of a distribution.

Use the skew statistic from the Descriptive Statistics. Then you will need to calculate the computed range to interprete the skew statistic value.

In this instance, **skew = -0.2** .

To get the "Computed Range" for interpretation, use this formula:

$$+/- 2 * \sqrt{\frac{6}{n}}$$

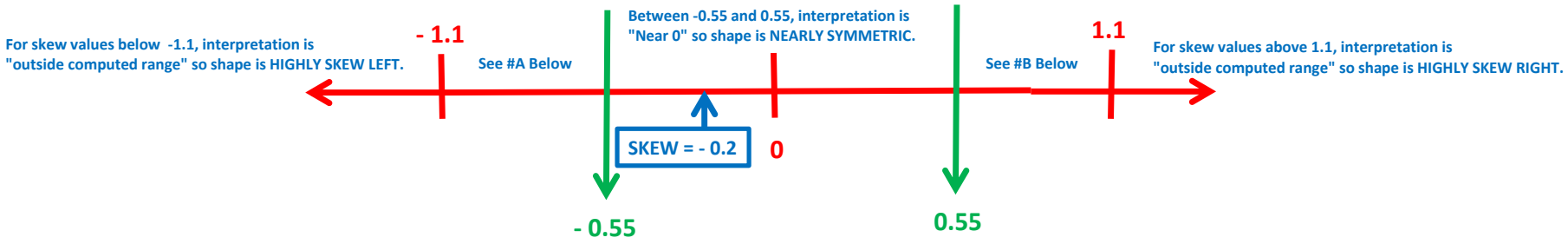
where n = sample size (number of pieces of data)

Since n=20 in this case, the computed range is:

-1.095445 to 1.095445

Rounded **-1.1 to 1.1**

For interpretation, draw the following scale.



Note: -0.55 is halfway bewteen 0 and -1.1; while .55 is halfway between 0 and 1.1.

Don't Be Intimidated, Remember: You are just plotting numbers on a number line.

#A Between, - 1.1 and -0.5, interpretation is "within the computed range" and shape slightly skew left.

#B Between, 1.1 and 0.5, interpretation is "within the computed range" and shape slightly skew right.

NOTE: Negative skew values indicate shape of distribution is skew left.

NOTE: Positive skew values indicate shape of distribution is skew right.

In this case, SKEW = - 0.2 which places it "near zero", and distribution shape is "nearly symmetric"

#5 To find outliers using the "Standard Deviation Method", first get the standard deviation and mean from the Descriptive Statistics above.

In this case, the standard deviation is $140.5 (millions), and the mean is 465.8.

Multiplying 140.5 times 2 provides the value you will add and subtract from the mean. In this case, that's 281 .

Just add this product of 281 to the mean. That gives you the upper limit of where outliers start. Upper Limit = 465.8 + 281 = **746.8**

Any value above this number is designated an outlier by this method.

Just subtract this product of 281 from the mean. That gives you the lower limit of where outliers start.

Any value below this number is designated an outlier by this method. Lower Limit = 465.8 - 281 = **184.8**

Note: Some problems may ask you to multiply the standard deviation by 3, but the calcualtions will otherwise be the same. 2 or 3 will always be provided in the question.

To find outliers using the "IQR Method", first get the IQR, median, Q1 and Q3 from the Descriptive Statistics above.

In this case, the IQR is $141.3 (millions), the median is 475, Q1 is 387.5, and Q3 is 528.8.

Multiplying the IQR times 1.5 provides the value you will add to the third quartile (Q3), and subtract from the first quartile (Q1).

In this case, 1.5 times the IQR of 475 equals 212.

Just add this product of 212 to Q3. That gives you the upper limit of where outliers start. Upper Limit = 528.8 + 212 = **740.8**

Any value above this number is designated an outlier by this method.

Just subtract this product of 212 from Q1. That gives you the lower limit of where outliers start. Lower Limit = 387.5 - 212 = **175.5**

#6 Just remember the the following connections.

For skew distributions, median is the best measure of central tendancy (center for short); IQR and range are the best measures of spread;

and IQR Method is best to determine outliers.

For symmetric distributions, mean is the best measure of central tendancy (center for short); standard deviation is the best measure of spread;
and Standard Deviation Method is best to determine outliers.

ANSWERS: **MEAN; STANDARD DEVIATION; STANDARD DEVIATION METHOD**

#7 Standardized Test Scores have no zero value but "differences" are consistent so it is "INTERVAL"

Number of Siblings in a Family has a zero value and "differences" are consistent so it is "RATIO"

These variables are categorical so they are reported in a "NOMINAL" scale.

Ranking system reflects an ordering of history knowledge but lacks a zero and the "differences" between values do not have meaning so it is "ORDINAL"

#8 See #4 above for more explanation.

NOTE THAT THE FORMULA DETERMINING COMPUTED RANGE MUST BE DONE SEPARATELY FOR EACH PROBLEM BECAUSE THE SAMPLE SIZE CHANGES.

$$+/- 2 * \sqrt{\frac{6}{n}}$$

	n =	Calculated Value	Computed Range	SKEW	Interpretation
8a.	9	1.633	-1.6 to 1.6	-1.3	Inside the computed range, so SLIGHTLY SKEW LEFT
8b.	49	0.6999	-0.7 to 0.7	0.2	Near zero so NEARLY SYMMETRIC
8c.	36	0.8165	-0.8 to 0.8	-0.9	Outside the Computed Range so HIGHLY (or Slightly OK) SKEW LEFT

You will be provided the formula for the final.

#9 - #12 BONUS QUESTIONS THAT WE WILL REVIEW IN CLASS.