Unit 12H Statistics Study Guide **Name: _____Per:_____**

Assn	Objective	Due		Done
12.1	Just Graphin' It (Representing 1-Variable Data)	4-8	4-9	
12.2	Let's Be Reasonable (Bivariate Data)	4-10	4-11	
	Statistics Study Guide			
12.3	Deviation (Correlation)	4-12	4-15	
12HR	Statistics Review	4-16	4-17	
EMT	Statistics EMT and Study Guide Due	4-18	4-19	

Targets	Sample Question	Ugh	Almost	Got It!
Find and understand Mean, Median, and Mode	Calculate the Mean, Median and Mode. Which is the best to measure the center of the data?			
Lay out data in a histogram, dot plot, or box plot (box and whisker)	Given the data create a histogram, dot plot and/or box and whisker.			
Split data into quartiles	Find the quartiles for {4, 6, 2, 8, 4, 9, 13, 2, 11}			
Find the 5 number summary for a set of data	Using the data, give the 5 number summary			
Find Outliers mathematically	Is there an outlier in the following data? Why or why not?			
Find the line of regression using a calculator	Using the table, find the line of regression and the r-value for (3, 2), (4, 1), (6, -2), (9, -5), (15, -15), & (11, -9) using a calculator.			
Find the r-value				
Understand correlation coefficients	Describe the correlation of the following data with a r-value of 0.8. Positive/Negative? Weak/Strong?			
Interpret Standard Deviation	Find the mean and standard deviation for the following data. What does it tell you about the data.			

Vocabulary

Mean ($\bar{x}$): _____

Median: _____

Mode: _____

5 Number Summary:_____

Quartiles:_____

Inner Quartile Range (IQR):_____

Minimum/Maximum: _____

Outlier: _____

Histograms:_____ Pros: _____ Cons: _____

Dot Plots:_____ Pros: _____ Cons: _____

Box Plots:_____ Pros: _____ Cons: _____

Line of Regression:_____

Correlation Coefficients (r-value): _____

Residuals: _____

Standard Deviation (σx):_____

Measures Of Central Tendency

Use this set {4, 6, 12, 3, 9, 7, 21, 15, 5, 1, 10, 12} to find the following.

Mean (arithmetic average—$\overline{x}$) is the sum of the data points divided by the number of _____. Find $\overline{x}$ ____

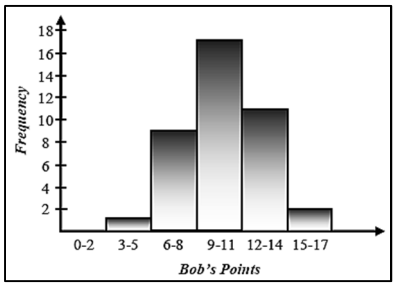
Median is the data point that divides the upper and _____ halves of a data set.

1. Order the numbers least to greatest *x*={1, 3, 4, ___, 6,___, ____, 10, 12, ____, 15, ____}
2. Find the number in the _____. With an even number of elements, find the average (8) of the two middle numbers (__ and __). While 8 is not part of the set, it divides the two middle _____. Explain why the Median of this set has a lower value than the Mean? _____

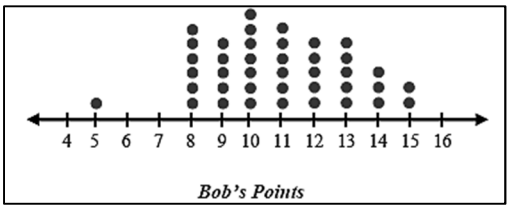
Mode is the number that appears most _____ in your data set. Find the mode: _____

Represent Data in Graphs

Different kinds of graphs have different strengths and weaknesses. The following are for univariate or _____ variable data displays.

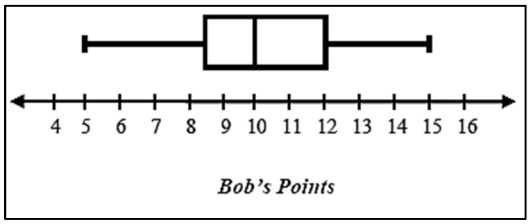


Histograms are like bar graphs except the bars _____ each other. They represent continuous data. The height of a histogram represents the frequency of (how many _____) an event (data point) occurs. List one weakness: _____



Dot Plots are one of the simplest statistical plots, and are suitable for small to moderate sized data sets. They are useful for highlighting clusters and gaps, as well as outliers. List one weakness:

Box Plots (often called "box and whisker plots") can be displayed _____ or horizontally, but they mean the same thing. Box and Whisker plots divide the data into quartiles (or into ____ sets) based on the values of the data from least to _____. The "box" contains the 50% of the data the falls in the middle of the set and the "whiskers" display the first and fourth quarters. List one weakness: _____



Quartiles

Quartiles divide data sets into _____ groups. Given the data set {4, 6, 12, 3, 9, 7, 21, 15, 5, 1, 10, 12}, order the data numerically. {1, 3, __, 5, 6, 7, 9, __, 12, __, 15, 21}. Since there are 12 convenient data points, this can easily be divided into 4 _____. The **Median (__)** divides the first two fourths or Q2 = 8. The first "quartile" falls between 4 and 5 so Q1 is their average and equals _____. Find Q3 = ____.

The **Minimum** value of the above data set is 1 and the **Maximum** value is 21. These are your upper and lower extremes unless there is an outlier. The **Range** is the max minus the _____. In this case, Range = _____. The **Inner Quartile Range (IQR)** is the range between the quartiles or Q3 to Q1. In this case 12 – 4 = ___.

- **5 Number Summary** is the above data values. Use…{3, 5, 2, 7, 9, 1, 4} find the 5 number summary:
  1. the sample minimum (smallest observation) = _____
  2. the lower quartile or *first quartile (Q1)* = _____
  3. the median (middle value) = _____
  4. the upper quartile or *third quartile (Q3)* = _____
  5. the sample maximum (largest observation) = _____

Outliers are data points that lie "outside" of the normal distribution. (If an outlier falls outside of the whiskers in a box plot, it is represented as a dot.) They are easily observed if they are extreme. Statisticians often use the common ratio of 1.5 to calculate whether a data point is a true outlier.

To determine an outlier, find the **Inner Quartile** _____ and multiply it by 1.5. Add and subtract that from Q1 and to Q3 respectively. An **Outlier** is any point that falling _____ of those values.

The data set {4, 6, 12, 3, 9, 7, 21, 15, 5, 1, 10, 12} has an **IQR** of 8. Since 8(1.5) = 12 an outlier would be any data point 12 below Q1 (4) or any number less than – 8. Likewise, an outlier would be any number 12 greater than Q3 (12) or any number greater than 24.

If a number 24 was in the data set, would it be an _____? Explain:

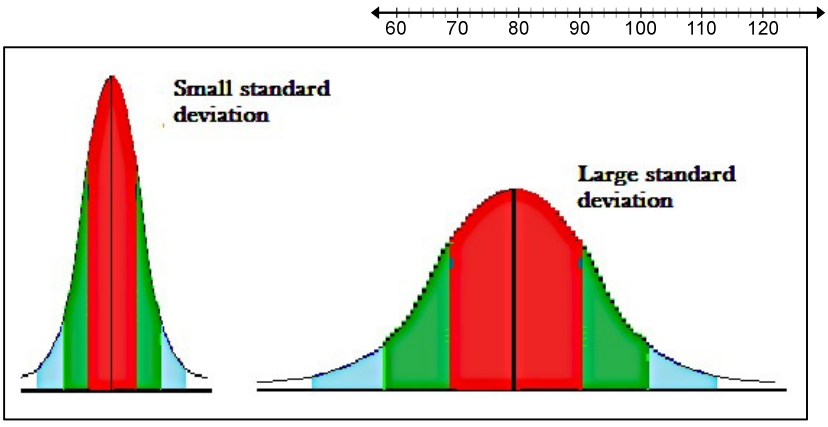
- **Given the data set {71, 70, 73, 70, 70, 69, 70, 72, 71, 125, 71, 69}**

Find the five number summary:

1. the minimum = _____
2. the lower quartile or *Q1* = _____
3. the median = _____
4. the upper quartile or *Q3* =_____
5. the maximum = _____

- Find the IQR = _____
- Determine whether the value 35 is an outlier. SYW.

- Make a Dot Plot
- Make a Histogram
- Make a Box Plot:



Standard Deviation σx

Standard Deviation applies to "univariate data" and shows how far the data are from the **Mean**.

If Joe has test scores of 60, 68, 69, 78, 90, 95, and 100 and Sam has test scores of 78, 78, 79, 79, 82, 82, and 82, the mean will not reveal the characteristics of the test score data. A standard deviation close to 0 indicates that all the data are **very close** to the **Mean**. A high standard deviation indicates that the data points are more _____ out.

- To calculate the Standard Deviation, enter the data in L_1. (See page 4.)
- Calculate the 1-variable data.
- Find σx at the bottom. (*Sx* is also a Standard Deviation but refers to all the data for an entire population.

Joe has a standard deviation (σx) of 14.1118. Find the standard σx for Sam's test scores. _____

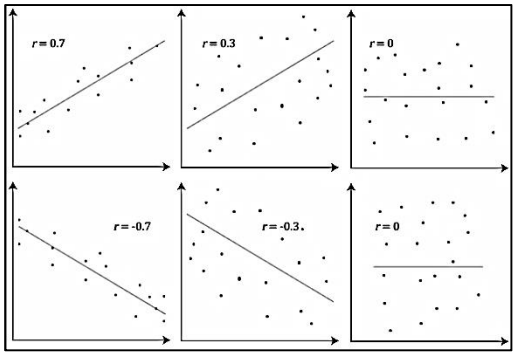
Bivariate Data

Scatter Plots show non-linear bivariate data and may have a linear correlation. A **trend line** reveals the "trend" or correlation and helps predict other _____. It is an estimate of the line representing the data.

A **Line of Regression** is the equation that gives the **exact** trend line or "_____ of best fit".

Correlation tells how closely the data points relate to each other. Two factors describe the general correlation: 1) **positive/negative** and 2) **strong/**_____. Positive/_____ indicates the general direction of the slope of the line representing the data.

Strong/weak describes how close the points are to the line.

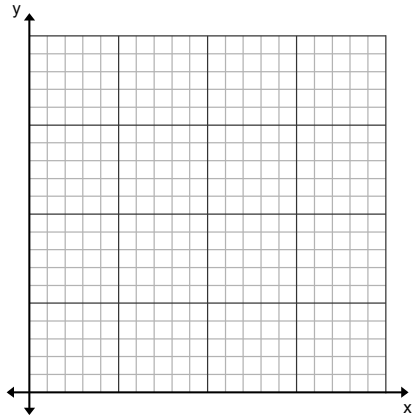


Residuals tell the vertical distance between the data points and the line of best _____. The **r-value** or "**correlation coefficient"** falls between -1 and 1. An r-value close to 1 or -1 shows a strong correlation (the points are very close to the line). An r-value of 0 shows no correlation. (A positive r-value says there is a positive slope; a negative r-value indicates there is a _____ slope.)

For example, if the same math test was given to a group of students and taller students scored better on the test, there is correlation between the height of the _____ and the results of the _____ (the 2 variables). This indicates that a student 6' tall would score _____ on the test than a student 5'2".

- Make a **Scatter Plot** for the data below. Follow the instructions at the end of the study guide to enter the points, graph the data, and find line of regression using the graphing calculator.

Pupil	A	B	C	D	E	F	G	H
Height "	46	72	42	56	74	52	46	66
Score	56	99	42	70	85	70	60	80



- Is there correlation of your data? _____
- Line of regression Equation: _____
- r-value (Correlation Coefficient): _____

Causation indicates one thing **causes** the other. In this example, causation would mean that being _____ generates higher test scores. If the students taking the test were different ages, taller students would logically test better as they would probably be the older students. While there is correlation, there is not _____.

- Find the line of regression and the r-value for the following table.
  - Line of regression Equation: _____
  - r-value (Correlation Coefficient): _____
  - Describe the correlation of your data._____

X	Y
3	5
4	7
5	10
8	12

Stats with the TI-84 (Quick Guide)

Calculators must be set to **DIAGNOSTICS ON.** (Go to MODE and then scroll to STATDIAGNOSTICS.)

Entering Data Points from the table to the right.

1. STAT > EDIT
2. Enter the data in the columns. (To plot points on a coordinate grid, you need two columns of data representing _____ variable (x and y).

X (L_1)	Y (L_2)
2	4.7
2.3	4
3.3	4.2
3.7	3.9
4.2	2.8
4.6	3.2
4.5	4.5
5	3.7
5.5	3.2
5.7	4.8
6.1	5
6.4	4.4

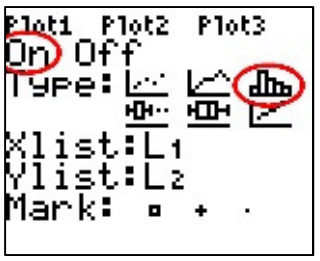
Line of Regression after entering data values:

1. Return to STAT go to CALC. Press either #4 or #8
   a. By default X will be L_1 and Y is L_2. (You can change it with the blue key & numbers.)
   b. Scroll down to CALCULATE to reveal the data for the line of regression.
   c. You should recognize $y = ax + b$. Find Line of Reg: _____
   d. $r - value$ is the correlation coefficient. _____
   e. Describe the correlation._____
   f. If Diagnostic is not set to ON, the r-value will not show. (See above.)

Five Number Summary and more from Y or L2 of the sets of data above.

1. Start with STAT > CALC.
2. Select #1 (1-Var Stats)
3. Select 2ND #2 for L_2. (This time we **only** want single variable data from _____ 2 or the y-values.)
4. Scroll down to Calculate. (Now there is a bunch of stuff so scroll down some _____.)
5. At the bottom, you'll find the 5-number summary: minX, the Q_1, Med (Median), Q_3, and maxX.
6. List the 5 Number Summary for the data in L_2.
   - Minimum = _____
   - Q1: lower quartile or *first quartile* = _____
   - Median = _____
   - Q3:upper quartile or *third quartile* =_____
   - Maximum = _____
7. Calculate the IQR. _____. Mathematically show if there is an outlier of not.
8. Find the mean ($\bar{x}$)= _____
9. Find the standard deviation (σx)**:** _____ What does the standard deviation tell you about the spread of the data?_____

Graphs--For univariate graphs for the y-values or from L_2



- STAT PLOT (2ND Y=) (Note: Line graphs and scatter plots see below.). Arrow down to desired stat plot (default is 1). It doesn't matter which you _____.
- Press ENTER and turn On the plot. Press ENTER
- Arrow to the desired kind of graph and press ENTER
- Set Xlist to L_2 by pressing 2ND #2. ENTER
- Use WINDOW to set x and y values to see the data entered. Select GRAPH.

For bivariate graphs

- Set Xlist to L_1 and Ylist to L_2 if these are the data you want to graph.
- Select the kind of graph: scatter plot, line graph, etc as in step 2 for univariate _____.