Name ________________________________ Period ________ Date______________

Activity 11:1 Regression Lines:

Materials needed for this activity:
Rulers or meter sticks
Calculators
Paper and Pencil

Preliminary Instructions: The TI-83x family of calculators have the built in capability to calculate the equation for both the MED-MED line and the Least Squares Regression Line.

You would be shortchanging the learning process if you started using the calculator too soon. **SO you are only allowed to use the calculator for basic operations; addition, subtraction, multiplication, division, square roots, list management, graphing, etc. You may not use the calculator to get the equations of the regression lines until after you have done the calculations "by hand". Then you may use the calculators to check your answers. If your answers are incorrect you are expected to go back and re-do the arithmetic until you get it right.**

A regression line is a line that represents the relationship between two variables. It is used to predict the behavior of the dependent or response variable over the interval being explored. It may or may not be valuable as a predictor outside of the interval that was originally explored, depending on the validity of the original assumptions over a larger interval.

There is more than one regression line. We can also calculate the Least Squares Regression Line. The major property of this line, which makes it desirable, is that it minimizes the sum of the squares of the residuals.

Name ______________________________ Period ________ Date______________

Activity 11:1 Regression Lines:

1. In this exercise we are going to explore another regression line called the MED-MED line which utilizes the median of a set of points rather than the mean. Take a moment and write down four properties of the mean and four for the median.

Mean	Median
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________

2. Now take another moment and **evaluate each** of the points you noted above and decide which might be the most important with respect to the process of regression. Explain how you decided on the relative importance of each of the attributes listed above.

____

____

____

____

____

Name ________________________________ Period ________ Date ______________

Activity 11:1 Regression Lines:

3. How do you calculate equation for the MED-MED line? Consider the following points:

x	3	5	7	12	14	16	20	21	22
y	10	17	25	41	49	47	62	65	70

- Sort the data points into ascending order based upon the x values as in the example
- Create a scatter plot for this data.
- Does the data seem to have a strong linear relationship? ______________________
- Are there any departures from the pattern of the relationship
- Divide the data points in three equal subsets. If the number of elements in not exactly divisible by three keep the first and third sets the same size and let the middle set be larger or smaller as needed.
- Find the median of the first set of x and y values. These will become the coordinates of the first reference point for the MED-MED line.
- Find the median of the third set of x and y values. These will become the coordinates of the third reference point for the MED-MED line. **Yes, they are out of sequence.**
- Write the equation of the line that goes through these two points. This will be the reference line.
- Find the median of the middle set of x and y values. This will be the second reference point.
- Calculate the predicted value of this line at the median x value of the middle set.

Name ______________________________ Period ________ Date______________

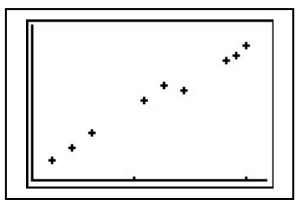
Activity 11:1 Regression Lines:

- Calculate the difference (Observed value – Predicted value) and **divide** this answer by three. Be sure to use the median y value for the observed value.
- Add this answer to the intercept of the line calculated earlier. This is now the equation for the MED-MED line.

Here's how these steps play out with the data given at the start of these instruction.

x	3	5	7	12	14	16	20	21	22
y	10	17	25	41	49	47	62	65	70

The scatter plot looks like this. There does seem to be strong linear relationship. If there were not a strong indication of the linear relationship we would stop here and look for a different model to explain the data. There also does not appear to be any major departures from the overall pattern.



The data would get separated in three groups
X values for group 1 (3, 5, 7) and the median is 5
Y values for group 1 (10, 17, 25) and the median y value is 17.
Medians of the x and y values in group 3 are Median x=21, Median Y=65

The equation of the line that connects these two points is $\hat{y} = 3x + 2$.
The medians of the x and y values in the middle group are (14, 47). The predicted value $\hat{y} = 3(14) + 2 = 44$. The observed value = 47. The difference 47-44 =3.

So we add $2 + \frac{3}{3} = 3$. The equation for the **MED-MED line = y=3x+3**. **Note that the median x and the median y for each group do not necessarily come from the same data point.**

This is a regression line so we have residuals. The residuals are (Observed Value-Predicted Value). Using this definition the residual when x=3 is 10-12=-2.

Name ________________________________ Period ________ Date______________

Activity 11:1 Regression Lines:

4. You should now find the equation for the MED-MED line for the following data.

X	2	4	5	8	9	12	14	18	20
Y	8	19	17	43	47	62	72	89	87

____

____

____

____

5. Collecting Data: Find 9 friends and measure the distance from the tip of their middle finger to the tip of their elbow. The distance from the tip of the middle finger to the tip of the elbow is called a cubit. Record each person's cubit length. This will be the x-value. Also measure their hand span. To do this have the person stretches his/her fingers apart as far as possible. The hand span is the length from the tip of the pinky to the tip of the thumb. Record the length. This will be the y-value.

Now for each of the data points find the residual using the equation of the MED-MED line.

	Observed x	Observed Y	Predicted Y Using MED-MED	Residual $(y-\hat{y})$	Residual $(y-\hat{y})^2$
1					
2					
3					
4					
5					
6					
7					
8					
9					

Name ______________________________ Period ________ Date______________

Activity 11:1 Regression Lines:

6. Is the relationship linear? ____

What is the equation of the MED-MED line? ____________________________________

7. What is the sum of the residuals squared?

____

____

8. Do you think the MED-MED line is a good predictor of the length of the hand span? Explain your reasoning.

____

____

____

____

The second regression line we are exploring is the Least Squares Regression Line. The French mathematician Adrien Legendre developed the method of finding the equation for this line.

$$\boxed{Slope = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sum\left(x-\bar{x}\right)^2}, \text{ intercept is } \bar{\mathrm{y}} - b_1\bar{x}}$$

As you know from elementary Algebra, the slope is a ratio (fraction) of the change in the vertical axis divided by the change in the horizontal axis.

Name ________________________________ Period _________ Date_______________

Activity 11:1 Regression Lines:

9. Interpret in words the meaning of the numerator of the fraction

____

____

10. Interpret in plain language the meaning of the denominator.

____

____

____

____

____

11. What is the sign of the denominator? ____________________________________

12. The equation of the Least Squares Regression Line for the data you collected.

____

Name ______________________________Period ________ Date______________

Activity 11:1 Regression Lines:

13. Repeat the analysis that you performed for the MED-MED line.

	Observed x	Observed Y	Predicted Y Using LSRL	Residual $\left(y-\hat{y}\right)$	Residual $\left(y-\hat{y}\right)^2$
1					
2					
3					
4					
5					
6					
7					
8					
9					

14. What is the sum of the residuals squared?

____

15. Which regression line has the smaller sum of the residuals squared?

____

16. Explain why this difference is important.

____

____

____

____

Name ______________________________ Period ________ Date______________

Activity 11:1 Regression Lines:

17. Do you think the Least Squares Regression line is a good predictor of the length of the hand span? Explain your reasoning.

____

____

____

____

18. If you use the regression line as a predictor of the behavior of the response variable, what assumptions are you making about the data?

____

____

____

____

____