**BIVARIATE REGRESSION ACTIVITY GUIDANCE**

**Activity learning goals:**

* Understand and apply the Ordinary Least Squares (OLS) estimation method
* Understand and apply the Least Absolute Deviation (LAD) estimation method
* Recognize situations where these two methods work well and do not work well

**Introducing the activity:**

1. Write down a bivariate regression model.
2. Give examples of what it could describe (e.g., wages as function of education and other stuff)
3. How do we estimate the beta’s?
4. If we ignore the error term, we have a line (draw one).
5. The model says that the observed data will be random deviations from this line (draw some dots).
6. But we don’t know what the line is (erase the line).
7. How do we go from the data (the dots) to an estimate of the line they came from?

**Guiding students during the activity:**

1. *How do the above scatter plots differ from each other?*

The first plot is the simplest one, and students should be encouraged to compare the other figures to it. Plots 2 and 3 are identical but with the addition of a few outliers. Plot 4 is exactly like the first except with a negative slope. Plot 5 has the same general slope as the first, but contains more noise, and the last plot is the same as the fifth but with a negative slope. We have found that students are quite good at identifying these differences.

1. *Write down a procedure (i.e., sets of steps) for fitting a line () through data (i.e., a set of n points xi, yi).*Students will often initially write down procedures that are not well-defined. For example, we’ve seen many groups include a step calling for outliers to be removed. Instructors circulating around the classroom should ask for clarification in these cases.
2. *Write down another procedure for fitting a line through the data.*The students who remembered the method of Ordinary Least Squares from another class are forced to be creative here.
3. *How do you think the results of each procedure compare in each of the above data sets?*  
     
   This is the most important question in the whole activity as students learn to identify the contexts where their method works well and where it does not. Often a method that works well when there is a strong positive correlation (e.g., “Connect the bottom left point to the upper right point”) works poorly when there are outliers or a strong negative correlation.
4. *Which of your procedures better represents the average linear relationship between x and y?*This is difficult and motivates the idea that there isn’t a single method that is the “best” in all contexts. It can also lead to a good discussion of how one might quantify the uncertainty in our estimates using standard errors or confidence intervals.

**Wrapping up the activity:**

Select 2-4 examples of student work, take pictures of them, and share them with the class. Point out where procedures are well-defined and ill-defined, and show cases (scatter plots) where procedures give good and poor results. Now that the students have identified several important features of bivariate data and have practiced evaluating their own algorithms, they are ready to be taught the methods of Ordinary Least Squares (OLS) and Least Absolute Deviations (LAD). The last question (about which procedure is best) can be used to motivate a presentation of the Gauss-Markov Theorem that says OLS is the Best Linear Unbiased Estimator (BLUE).