ACTIVITY 3: Modeling and Prediction with Least Squares Regression

OVERVIEW: In this activity, students will understand that the least squares regression line is a linear equation that models the relationship between two quantitative variables. They will understand when it is appropriate to use the least squares regression line for prediction through critical thinking questions on data provided by the instructor. Lastly, they will calculate the least squares regression line and evaluate whether it will accurately predict the value of the dependent variable from the value of an independent variable based on the scatterplot and the correlation coefficient they found and determine what would be the best predictor for their data.

MATERIALS NEEDED TO CARRY OUT THE LESSON:

- In-Class Activity Worksheet
- Laptop with Excel and PowerPoint
- Access to the internet

CONCEPTS TO BE LEARNED/APPLIED:

- Understand how to calculate the least squares regression line
    - Students will understand that the least squares regression line is a linear equation that models the covariational relationship between linearly related variables.
    - Students will understand that the least squares regression line is only appropriate when the relationship between the two variables shows a linear trend.
    - Students will understand that outliers in the x-direction can have a strong influence on the least squares regression line.
      Note: Outliers in the y-direction do not typically impact the regression line as severely as outliers in the x-direction. More specifically, outliers in the y-direction do not normally change the direction or slope of the line as much as outliers in the x-direction could. This is especially true for larger data sets.
- Understand when to use the regression line for prediction
    - Students will understand that the regression line should only be used for prediction when the scatterplot is showing a linear relationship.
    - Students will understand that the least squares regression line should not be used to predict outside the range of the independent variable.

PREREQUISITE CONTENT KNOWLEDGE:

- Students need to understand linear equations.

INSTRUCTIONAL PLAN:

Before handing out the activity worksheet, explain to students that when they have two quantitative variables that are linearly correlated, they can calculate the least squares regression line. The least squares regression line is a linear equation that best fits the data. Excel and other software packages will calculate the slope and y-intercept of the linear equation that would best fit the data. Often, we want to predict the value of the dependent variable based on the value of the independent variable. We cannot always use the regression line for this prediction. Below is a list of the best predictors depending on the relationship.

Relationship	**Best predictor**
Linear	The best predictor would be the least squares regression line.
Relationship is nonlinear	The best predictor would be a nonlinear equation that fits the curve of the relationship.
No relationship	The average of the dependent variable would be the best predictor.
Data has an outlier in the x-direction that impacts the slope of the regression line	It would be best to remove the outlier in the x-direction and find the best model to fit the remaining points.

Note: Typically, outliers in the y-direction do not change the direction or slope of the regression line. Once the outlier is verified and not a mistake in the data, it is often included in the data when finding the regression line.

MIP COMPONENTS OF INQUIRY

<u>Active Learning</u>

- In Questions 1-4, students must select a definition of slope and y-intercept from their prior knowledge, evaluate how to utilize this knowledge to create an equation for the least squares regression line, and draw the line on a graph.
- In Questions 6-8, students select the definition of “least squares regression line” and evaluate whether it is appropriate to find a least squares regression line for the data displayed graphically. The mental activity in the evaluation is attending to whether the data appears to be linear with a near-constant rate of change.
- In Question 9, students select the definition of “least squares regression line” and evaluate why a least squares regression line is not appropriate for prediction.
- In Question 9, students encounter a “problematic situation” as they need to determine a method which they would use for prediction.

- In Questions 10-16, students encounter “problematic situations” as they encounter situations whose resolution focuses on the content to be learned, which is how to handle outliers in either the x-direction or y-direction.

The following bullet points pertain to the questions where students use a given Excel spreadsheet. Note that the numbering starts again at 1.

- In Question 3, students select the appropriate information for Excel, perform the computation in Excel by utilizing the correct formula, and use their prior knowledge of linear equations to write the equation of the least squares regression line.
- In Questions 4-5, students select the equation and input value to make a prediction using their least squares regression line equation. Then, they evaluate its accuracy using their critical thinking skills.
- In Questions 7-8, students select the equation and input value to make a prediction using their least squares regression line equation. Then, they evaluate its accuracy using their critical thinking skills.

<u>Meaningful Applications</u>

- In Questions 2-3, students interpret the meaning of their slope and y-intercept as they generalize their knowledge to this new context.
- In Question 5, students use their critical thinking skills and justify whether they believe the correlation is, in fact, causation.
- In Questions 6-8, students justify their claim on whether it is appropriate to find a least squares regression line for the given data.
- In Question 9, students use their critical thinking skills to justify why a least squares regression line is not appropriate for prediction.
- In Question 11, students justify whether they would get an accurate prediction if they used the scatterplot with an outlier.
- In Question 13, students justify whether the regression line in a graph would be useful in a certain scenario.
- In Question 14, students generalize their findings regarding the best way to deal with outliers in the x-direction.
- In Question 16, students justify whether they would still get a good prediction using the regression line that was computed using an outlier in the y-direction.
- In the reflection/summary after Question 16, students generalize their findings on how to deal with outliers.

The following bullet points pertain to the questions where students use a given Excel spreadsheet. Note that the numbering starts again at 1.

- In Question 1, students analyze and describe the relationship between two variables after they create a scatterplot in Excel with the given data.

- In Question 2, students use their critical thinking skills to justify whether it would be appropriate to find the least squares regression line in this situation.
- In Question 10, students analyze and describe the relationship between two variables after they create a scatterplot in Excel with the given data.
- In Question 11, students use their critical thinking skills to justify whether extrapolation would be advisable.
- In Question 12, students generalize their findings throughout the activities regarding a least squares regression line and when to use it for prediction.

Academic Success Skills

- In Question 9, students have to come up with alternate prediction methods whenever a linear equation is not a good fit. This engages students in critical thinking and highlights the fact that they have mathematical knowledge that is relevant and useful.
- Many of the questions in this activity give students a definition (e.g. least squares regression line or outliers) that they apply to real-world scenarios. Working with definitions is a key mathematical practice and builds students' critical thinking skills.
- Many of the questions in this activity require students to use prior knowledge regarding the characteristics of a line (e.g. slope, constant rate of change, and y-intercept) and apply it to a real-life example. This positions them to see themselves as people who know some relevant mathematics that they can use to make sense of new mathematics. This will support them in seeing themselves as members of an academic community.