ACTIVITY 2: Scatterplots, Linearity, and Correlation

OVERVIEW: In this activity, students will analyze scatterplots and correlation coefficients for quantitative variables, and use them, when appropriate, to describe covariate relationships. Students will then apply their newfound knowledge to create a scatterplot and calculate the correlation coefficient for their data using Excel. They will create two more slides for their PowerPoint presentation which will contain their scatterplot, a description of the relationship between their chosen variables, and the value of the correlation coefficient with an explanation of its validity.

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 create scatterplots, compute the correlation coefficient, and use them to analyze data
    - Students will understand that when creating scatterplots, they always plot the independent variable on the *x*-axis and the dependent variable on the *y*-axis.
    - Students will understand that scatterplots show the form, direction, and strength of the relationship between two quantitative variables.
    - Students will understand how to differentiate between an outlier in the x and y direction.
    - Students will understand how to create scatterplots using Excel.
- Understand how to calculate and interpret the correlation coefficient.
    - Students will understand that a correlation coefficient close to “-1” represents a strong negative linear relationship, a correlation coefficient close to “1” represents a strong positive linear relationship, and a correlation coefficient close to “0” represents a weak, non-linear relationship or no relationship.
    - Students will understand that the correlation coefficient is only accurate when the relationship is linear.
    - Students will understand that in real-world linear relationships, we look for the rates of change to be approximately constant.
    - Students will understand the importance of first looking at a scatterplot before computing the correlation coefficient.
    - Students will understand outliers in the *y*-direction will decrease the correlation coefficient, and outliers in the *x*-direction can have a strong influence on the correlation coefficient.
    - Students will understand how to find the correlation coefficient using Excel.

INSTRUCTIONAL PLAN:

Have the students work on the first four pages of the activity worksheet. Discuss their answers as a class and make sure they understand the following.

The scatterplot

- always has independent variables on the x-axis and dependent variables on the y-axis.
- shows the form (linear or non-linear), direction (positive for positive slope and negative for negative slope), and strength (e.g. how closely the points are clustered around the regression line) of the relationship.
- shows outliers in the y-direction, which are significantly above or below the other data with similar x-values, and outliers in the x-direction, which are significantly to the left or right of all the other data values.

Have the students work on pages 5-8 of the activity worksheet. Discuss their answers as a class and make sure they understand the following.

A correlation coefficient

- with a value close to “-1” represents a strong negative linear relationship, a value close to “1” represents a strong positive linear relationship, and a value close to “0” represents a weak, non-linear, or no relationship.
- only makes sense when the relationship is linear.
- should not be calculated without also looking at a scatterplot.
- is reduced by an outlier in the y-direction and can be increased by an outlier in the x-direction.

The students are now ready to complete the worksheet. They will input the data they collected in Activity 1 into Excel and find the scatterplot and the correlation coefficient.
Note: Students should save their Excel spreadsheet for use in Activity 3.

After they are done, they can create the next two slides for their PowerPoint presentation. Their slides should contain the following:

Slide 2: The scatterplot with a description of the relationship between their chosen variables.
Note: Make sure they include a description of the form, direction, and strength of the relationship.

Slide 3: The correlation coefficient and an explanation as to why it was or was not appropriate to find this for their data set.

MIP COMPONENTS OF INQUIRY

Active Learning

- In Questions 1 and 2, students select the variables, evaluate the relationship between them, and label them as independent or dependent.
- In Questions 3, 4, and their summary, students select the data they need to make an empirical claim, evaluate whether there is a pattern, and identify that the pattern is that the independent variable goes on the x-axis and the dependent variable goes on the y-axis.
- In Question 5, students must select a definition of linearity and constant rate of change from their prior knowledge, evaluate whether the data meet their definition, and perform a classification of the scatterplots as linear or non-linear.
- In Question 6, students must select a definition of positive and negative slope from their prior knowledge, evaluate whether the data meet their definitions, and classify the slope.
- In Question 7, students select the given definition of “strong” and evaluate whether the given quantities have a strong or weak relationship based on the definition. The mental activity in the evaluation is attending to coordinate points’ distance from the line. Then, they classify the relationships.
- In Question 8, students select the given definitions and evaluate which definition the given points meet by comparing the point’s x- and y-values to the other data. The mental activity is comparison. Students then classify the points.
- In Questions 9-12, students select the data they need (e.g., graphs that have an r value close to -1 and 1), evaluate what is similar in the slope of the lines that would model those graphs, and form a conclusion.
- In Question 13, students select a definition of “strength” and evaluate whether the points’ output value follows a pattern (for strength) and classify the relationship as strong or weak.
- In Question 13, students select a definition of “form” (defined earlier in the activity as linear or nonlinear), evaluate whether the points have a constant or near-constant rate of change and/or look like a line (depending on how they conceptualize linearity), and classify the relationship.
- In Questions 16 and 17, students select a method of thinking about the correlation coefficient (e.g., as how far points are from a line), evaluate how a point at a great distance will change it, and form a conclusion. The mental activity entailed in evaluating it might be considering the distance of non-outlier points from the line as compared to the outlier point.
- In Questions #18-20, students select the appropriate definition of outlier (outlier in the x-direction or outlier in the y-direction), evaluate which definition the given outlier meets, and form a conclusion as to whether the correlation coefficient accurately

reflects the relationship. The mental activity entailed in forming the conclusion involves the student evaluating the meaning and implications of the correlation coefficient.

- On #6 in the Excel directions, students select a definition of linearity (e.g., data with a constant rate of change), evaluate whether a fixed increase in the input is associated with a near-constant change in the output, and classify the relationship as linear or nonlinear.
- On Slide 3 of the PowerPoint presentation, students find the correlation coefficient and then evaluate whether it is appropriate by reflecting on if the data is linear and if it has outliers. In creating the correlation coefficient, students may have had to decide whether to remove an outlier (via selecting a definition of an outlier and comparing any possible outliers to non-outliers based on the input and output values of the non-outliers).

<u>Meaningful Applications</u>

- In Questions 1-2, students generalize their understanding of independent and dependent variables from algebra to statistics.
- In Questions 3-4, students make a generalization about which variable goes on which axis based on empirical examples.
- In the summary of Question 4, students generalize that independent variables go on the x-axis and dependent variables go on the y-axis from viewing empirical examples in which this is the case.
- In Question 5, students generalize linearity from algebra to scatterplots and real-world data.
- In Question 5, students justify their classification of the variables (linear or non-linear).
- In Question 6, students apply their knowledge of slope and expand it to thinking about it as a relationship between variables, if they did not think of it that way already. This extends their covariational reasoning.
- In Question 8, students identify and determine a possible cause of an outlier that has real-life relevance.
- In Questions 9-12, the inequality and the number line the students complete help them connect mathematical representations.
- In Questions 16-17, students engage in critical thinking to make sense of how adding a large output value affects an r value.
- In Questions 18-22, students think about generalization – specifically, to what contexts generalization is appropriate.

<u>Academic Success Skills</u>

- In the summary for Question 4, students generalize that independent variables go on the x-axis and dependent variables go on the y-axis from viewing empirical examples in which this is the case. This supports them in learning the disciplinary practices that help them see themselves as competent doers of mathematics, helping them form an identity as learners.

- In Question 5 and others about linearity, students apply prior knowledge about linearity to a new context. 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.
- In Question 8, students think critically about the potential cause of outliers and explain their reasoning. Critical thinking is a key skill that supports students' mathematical success.
- Many of the questions in this activity give students a definition that they apply to scenarios. Working with definitions is a key mathematical practice and builds students' critical thinking skills.