**Unit 5: Investigation 5 (2 Days)**

**Outliers**

***CCSS:*** *S- ID 6; S-ID 8*

**Overview**

Students will work with data sets that contain outliers to identify the influence that outliers have on the calculation and interpretation of the slope, *y*-intercept, linear regression equation, and correlation coefficient.

**Assessment Activities**

**Evidence of Success: What Will Students Be Able to Do?**

Students will be able define an outlier, identify whether a potential outlier is present on a scatter plot and name the coordinates of the outlier, draw regression lines and provide a general description of the influence that outliers have on the slope as well as the direction and strength of the relationship between two variables; and describe the impact that outliers have on linear regression equations, their related components (i.e., slope, *y-*intercept, correlation coefficient), and the conclusions drawn from an analysis of a data set in which they are included.

**Assessment Strategies: How Will They Show What They Know?**

**Exit Slip 5.5** requires students to identify if an outlier exists and then make a prediction using the data set.

**Journal Entry** asks students to explain how to determine whether or not a data point is an outlier.

**Launch Notes**

Begin with a whole class discussion about outliers. Have students think about what they learned in Investigation 1. Ask students how they would identify an outlier from a dot plot, a histogram, or a box-and-whiskers plot. Then ask them to think about what an outlier would look like if they had a scatter plot. They may suggest that an outlier would appear at a greater distance from a regression line than the other points on the graph. This leads into **Activity 5.5.1**, which introduces outliers in two variable data.

**Closure Notes**

In a wrap up discussion you may ask the class, “Why are outliers important?” Students should realize that outliers can affect the accuracy of a prediction made with a regression line. **Exit Slip 5.5** may be used to assess their ability to take the effect of an outlier into account when making a prediction.

**Teaching Strategies**

1. After reviewing the concept of outlier from Investigation 1, use **Activity 5.5.1** **Outliers** to introduce outliers in the context of two-variable data. Students may have difficulty identifying an outlier from the table; some may select “Crazy, Stupid, Love” since it has been out the longest time from its release. Others may focus on “The Help” since it leads in gross sales. Suggest that making a scatter plot will help resolve the issue. You may choose to have them make a scatter plot by hand or use the calculator. (Note that these data present an opportunity to discuss the degree of precision that is appropriate to this situation. In making a graph and using it to make predictions, it makes sense to round the sales figures to the nearest million dollars.)

Once students have seen a scatter plot of the data they are likely to agree that “The Help” is indeed an outlier. “Crazy, Stupid Love” has the largest value of *x* but it appears to be close to a line passing through the other data points. An explanation for the outlier is that “The Help” was an extremely popular movie. When students address question 5, there is no correct answer. Including the outlier takes into account that a few of the future movies may also turn out to be big hits. Excluding the outlier will give a more accurate prediction for the sales of a *typical* movie.

Students may then follow up with the questions on page 2, which ask them to make predictions based on the regression line they choose in question 6.

In **Activity 5.5.2** **Barry Bonds’ Home Runs**, students are asked to analyze a table and a graph to identify and explain two possible outliers. They are asked to create a regression line and find the equation and correlation coefficient for the data. Students should use a graphing calculator; however, if that is not an option, instead of having them find the correlation coefficient, have them describe the type of correlation between the data. Students will then remove the outlier and repeat the process. Students should then look at the two different sets of information they have and determine the differences between them.

**Activity 5.5.2** can lead to a rich discussion about how much weight statistical evidence can be given in determining a legal case. When looking at his previous performance, Bonds’ 73 home runs in 2001 do appear to be an outlier, which requires some explanation. However, the fact that it is an outlier does not prove that the cause was his use of performance enhancing drugs.

**Activity 5.5.3 Home Prices** involves a less complex set of data than Barry Bonds’ homeruns and is suitable for individual work either in class or for homework. In this activity, students are asked to identify an outlier and then find a trend line. They can find a trend line by hand or find the line of best fit using a graphing calculator. Students will then be asked to make predictions and determine whether the prediction involved interpolation or extrapolation.

1. The second day of this lesson gives you a chance to consolidate students’ understanding of outliers. **Activity 5.5.4 Chicago Bulls** may be assigned for work in groups followed by a whole class discussion. Students will discover that one super star (Michael Jordan) does not fit the regression line relating number of minutes played in a season to total points scored in a season. This activity also reinforces the importance of the correlation coefficient and how to interpret the slope of the line in a real world context (in this case, points per minute).

**Group Activity**

Students may play the “Outlier Game.” Students work in pairs, each with a calculator. Each student enters 10 pairs of data in L1 and L2. They should try to find a set with strong positive correlation, but must not place all the points on a single line. Each student then computes the value of *r*.

Students then switch calculators. Each student now enters an eleventh ordered pair. The object is to make this pair an outlier. They now re-compute the value of *r*. The winner is the student whose outlier reduces the opponent’s value of *r* the most. You may want to restrict the values they can choose to lie within a specified window.

**Differentiation (Enrichment)**

**Activity 5.5.5 Crickets Chirping** may be assigned. This is activity has a minimal amount of structure, asks students to use their judgment concerning outliers, and then has them research other formulas relating temperature and cricket chirps to the one they have derived.

**Differentiation (For Learners Needing More Help)**

**Activity 5.5.4 Sea Glass** may be assigned. This is a highly scaffolded activity that will help students master the steps needed to analyze as set of data with an outlier.

**Exit Slip 5.5** may be used to assess students’ understanding of outliers and their effect on regression lines.

**Journal Entry**

In your own words, describe how you would determine if an ordered pair is an outlier. You may use examples from class to support your explanation.

**Resources and Materials**

* **Activities 5.5.1** Outliers
* **Activities 5.5.2** Barry Bonds’ Home Runs
* **Activities 5.5.3** Home Prices
* **Activities 5.5.4** Chicago Bulls
* **Activities 5.5.5** Crickets Chirping
* **Activities 5.5.6** Sea Glass
* **Exit Slip 5.5** Cooling Off
* Instructions for the Outlier Game
* Bulletin board for key concepts
* Calculators