**Unit 5: Investigation 3 (2 Days)**

**Technology and Linear Regression**

***CCSS:*** 8-SP 1; 8-SP 2; 8-SP 3; S-ID 6 a, c; S-ID 7; S-ID 8, S-ID 9

**Overview**

Students will develop a deeper understanding of trend lines and predictions. They will use technology (either a graphing calculator or a spreadsheet) to calculate the linear regression equation and to find the correlation coefficient. The students will be able to interpret the meaning of the correlation coefficient. The students will also be able to explain the difference between correlation and causation. Be creative about collecting and finding data – local newspaper, internet searches, etc.

**Assessment Activities**

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

Students will be able to find the equation for the line of best fit using technology. They should be able to identify the strength and direction of a trend line using the correlation coefficient. Student should be able to explain the difference between one variable being *correlated* to the other and one variable *causing* the other to occur.

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

**Exit Slip 5.3** asks students to match a graph with a possible value of *r*.

Journal Entry probes students’ understanding of correlation and causation.

**Launch Notes**

Try to start off this investigation with a video to get the students attention and interest. If you don’t have the ability to show videos in your classroom bring in objects related to the upcoming lesson that may get them interested (several old cell phones, fake money, college pamphlets, or pictures of basketball players)

**Closure Notes**

Ask students to reflect on how technology is helpful in finding the line of best fit and the correlation coefficient. Then ask them how much information they can obtain from a graph made by hand. They should recognize that without knowing a precise value of *r*, they can obtain a qualitative understanding of the strength and direction of the correlation just by viewing the graph.

**Teacher Notes – Background Information on the Correlation Coefficient**

*What is the correlation coefficient? This information is for the teacher’s background knowledge.*

*The “r” value is Pearson’s Sample Correlation Coefficient or just the correlation coefficient. The word correlation refers to the “co-relation” between the two variables being analyzed. The correlation coefficient is one indication of how well the linear regression fits the data. The value of r is always a number between -1 and 1. It provides two pieces of information: the direction and strength of the linear relationship between the two variables. If r is positive, then the relationship is increasing: as x values increase, so do y. If r is negative, then the data is negatively correlated: as x values increase, y values decrease. The closer r is to 1 or -1, the stronger the linear relationship between the two variables. If all the data points are collinear, non-horizontal, then r = 1 or -1.*

*If there is no linear correlation between the two variables, then r = 0. If r = 0, that does not mean that there is no relationship between the variables, only that the relationship is not linear. For example, data that is in the shape of a circle, a v, or a parabola will have r = 0. Data that is horizontal will also have r = 0, even collinear horizontal data. The magnitude of the correlation coefficient indicates to what extent for any value of x the linear regression is a better estimator of the corresponding value y than is the simple arithmetic mean of all the y data. So, for example, if the slope of the linear regression is zero, then the linear regression can do no better predicting the y variable than the average of the y data, because, for horizontal data, the linear regression is “y = average of the y data.”*

*Some students may ask about the formula for r. If “n” is the number of ordered pairs,* $\overline{x} $*is the mean of the x values,* $\overline{y}$ *is the mean of the y data, sx is the standard deviation of the x values, sy is the standard deviation of the y values in the ordered pairs that are your data, then* $r=\frac{∑(x- ¯x)(y- ¯y)}{(n-1 )Sx∙ Sy }$*. Do not ask students to calculate r by hand except in a statistics course.*

**It is very important that the diagnostics is turned on for all the graphing calculators. If this is not turned on, then you will not get an *r*-value. When they run the linear regression**

|  |  |
| --- | --- |
| **Prepare The Calculator** 1. **Turn on the Diagnostics** so that the correlation coefficient *r* will appear.
2. Clear the home screen.
3. Press **2nd 0 CATALOG** and scroll down to **DiagnosticOn** then press **ENTER**.
4. Home screen now shows **DiagnosticOn**. Press **ENTER**.
5. Home screen shows **Done**.
 |  SCREEN01.JPGSCREEN02.JPG |

There are different sets of graphing calculator directions. The file titled **Unit 5 Complete Graphing Calculator Directions** is generic, in-depth, and useful for all situations. There are other graphing calculator directions that are specific for certain lessons. You have the choice on how you want to distribute the directions. If your students do not have graphing calculators at home, you can always choose an activity from the previous investigation that you did not do.

**Teaching Strategies**

I**. Activity 5.3.1 Fitting Lines with Technology** is intended to piggyback off the previous investigation. It begins with the same set of data for the height and weight of basketball players as in **Activity 5.2.2.** This time students find the line of “best fit” with the linear regression program. Students learn that in this program the slope is represented by *a* (rather than *m*) and the y-intercept, as before with *b*. They may compare the regression equation with the equations they derived by hand in the previous activity.

Technology note:

Screens [1] through [8] on pages 1 and 2 of **Activity 5.3.1** show students the steps needed to find and plot the regression equation. In screen [2] note that you may choose from Plots 1, 2, and 3, and that the data does not necessarily have to be in lists L1 and L2. You also have a choice of which symbol to use for the data points on the graph.

In screen [3] we have instructed students to examine the data to determine the window parameters. The short cut for doing this is Zoom 9 ZoomStat. You may want to avoid introducing this shortcut until you are confident that students can figure out the window parameters on their own.

When you select the LinReg (ax +b) feature as shown on screen [5] you should specify which lists the data are in. On some calculators the default is L1 and L2, but students should be encouraged to name the lists instead of relying on the default option.

In Screen [7] students have entered the regression equation from the information from screen [6] into the Y= menu, with the parameters rounded off. A short cut is to tell the calculator where to store the regression equation. On some calculators this is done by adding the location to the home screen with the command LinReg(ax + b) L1, L2, Y1. On the Ti-84 Plus there is a prompt on the screen for the regression equation. In both cases you access Y1 by pressing VARS🡪Y-VARS-🡪 1 Function🡪 1: Y1.

Students are then introduced to the meaning of *r*. The scatter plots shown on page 2 are designed to develop an intuitive understanding of how the value of *r* indicates the strength and direction of the correlation. Students should then classify the correlation between basketball players’ heights and weights as strong and positive.

If time permits, students may then look at two other data sets and again find the regression equation and the correlation coefficient. Note that in both of these examples the correlation is again strong although it is positive in problem 7 and negative in problem 8. Point out that the direction of the correlation (positive or negative) is related to the slope of the regression line.

You may then introduce **Activity 5.3.2 Evolution of the Telephone** by showing a video showing the history of the cell phone: <http://www.youtube.com/watch?v=JcnXOhrmDB8>. If you can project videos, but you are blocked from YouTube, then search for “evolution of the cell phone” on another website that your school does allow. If you cannot show videos, then print out some pictures of old cell phone and put them around the room. Have volunteer students read about the history of the telephone.

Have students circle or highlight the row for corded phone the sales in the first table before starting the table at the bottom of page 1 and then transferring the data to lists in the STAT menu. Go through the accompanying slideshow with the students. If you do not have access to PowerPoint, all the information can be done as board notes. The students may get hung up on question 11 in which they write the equation of the trend line by hand. If the students are struggling, then give them the steps to writing the equation as guidance as was done in the previous investigation.

**Differentiated Instruction (For Learners Needing More Help)**

Have students make a procedure card that lists the keystrokes for plotting data and calculating the regression. Allow students to use the card.

As in the previous activity students are asked to interpret the value of the correlation coefficient *r*. And as in the previous investigations students continue to make predictions based on the trend line. The last two pages of Activity 5.3.2 ask students to pick another product and perform and analysis similar to the one done for corded phones. Questions 14–20 may be assigned as homework even if students do not have access to calculators at home. They may then use calculators to answer questions 21–24 in class the next day.

II. As students continue to find regression lines and interpret correlation coefficients they can now begin to think about the distinction between correlation and causation. **Activity 5.3.3 Correlation and Causation** provides and opportunity to develop this concept. In the first example one can argue that higher temperatures cause an increase in sales of ice cream, since in warm weather more people are outdoors and looking for cool refreshments. In the second example sales of sunglasses and ice cream may both be related to the weather, but one does not cause the other. In the third example Target sales increase in the later months of the year (probably due to holiday shopping) but in no way to the high sales cause the decrease in sunlight (nor does the decrease in sunlight cause sales to rise). In the final example, however, the decrease in sunlight with the changing seasons does cause a drop in average temperature.

**Group Activity**

Do the first example in **Activity 5.3.3** with the entire class and then assign the remaining examples to small groups. Each group can then report their findings to the class. They should report the correlation coefficient and their conclusion as to whether causation is present. The courtroom scene for Shark Attacks (**Activity 5.3.4)** described below is also suitable for students working in groups.

An alternative way to introduce the distinction between correlation and causation found in **Activity 5.3.4 Shark Attacks.** There *is* a strong positive correlation (r ≈ 0.93) between ice cream sales and the number of shark attacks, but confusing correlation with causation could lead to the absurd conclusion that ice cream sales should be banned! There are two variations of this activity. If you want to set up a courtroom scene, use **Activity 5.3.4a**. Otherwise you can use the worksheet format **Activity 5.3.4b**, possibly as a homework assignment.

**Differentiated Instruction (For Learners Needing More Help)**

The wording for **Activity 5.3.4 Shark Attacks** can be simplified. Vocabulary and key words can be defined in the students’ notes. Instead of grouping by math ability, group the students by reading ability.

**Differentiated Instruction** **(Enrichment)**

To make questions 7-10 more challenging, ask students to come up with their own examples and have other groups give their opinions of them.

**Activity 5.3.5 Regression Equation Practice** may be used as needed for additional class work or homework. The examples include the relationship between the area of a home and its sale price, the battle against malaria, and how the cost of attending college has increased over time.

**Differentiated Instruction (Enrichment)**

Have the students research the cost of the “top choice” college they want to attend. Is the college they like private or public? Have the students compare their prediction to price of this school.

**Journal Entry**

How do you determine the difference between correlation and causation? Give an example of two variables that will be highly correlated and one variable causes the other to happen. You may come up with your own variables or you may use examples from previous classes (but not today’s class). Explain why these variables demonstrate correlation and causation. Your response should be no less than 5 sentences.

**Resources and Materials**

* **Activity 5.3.1 Fitting Lines with Technology**
* **Activity 5.3.2 Evolution of the Telephone**
* **Power point for Activity 5.3.2**
* **Activity 5.3.3 Correlation and Causation**
* **Activity 5.3.4a Shark Attacks Courtroom scene**
* **Activity 5.3.4b Shark Attacks Worksheet**
* **Activity 5.3.5 Regression Equation Practice**
* **Exit Slip 5.3 - Scatter Plots and Correlations**
* Bulletin board for key concepts
* Graphing Calculators
* Student Journals
* Projector
* Computers
* Rulers