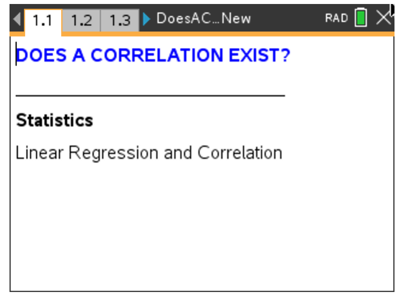
Does A Correlation Exist?

MATH NSPIRED

TEACHER NOTES

IB Aligned



Math Objectives

- Students will construct a scatter plot and determine a linear regression that fits the data.
- Students will calculate a correlation coefficient from data sets.
- Students will describe correlation in data sets and distinguish it from causation.

Vocabulary

- Causation
- correlation coefficient
- outlier
- linear regression

About the Lesson

- This lesson is aligning with the curriculum of IB Mathematics Applications and Interpretations SL/HL and IB Mathematics Approaches and Analysis SL/HL
- This falls under the IB Mathematics Core Content Topic 4 Statistics and Probability:
  4.4a Linear correlation of bivariate data, Pearson's product-moment correlation coefficient, *r*.
  4.4b Scatter diagrams, lines of best fit (by eye), passing through the mean point.
  4.4c Equation of the regression line of y on x, use of the regression line for prediction purposes, and interpret the meaning of the parameters (a and b) in a linear regression $y = ax + b$.
- Students should have prior experience using the TI-Nspire handheld to create scatter plots. In addition, they should have had some discussion about correlation.
- Notes for using the TI-Nspire™ Navigator™ System are included throughout the activity, but is not necessary for completion.

TI-Nspire™ Navigator™ System

- Send out the *DoesACorrelationExist.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software

Tech Tips:

- This activity includes screen captures taken from the TI-Nspire CX II handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:

Student Activity

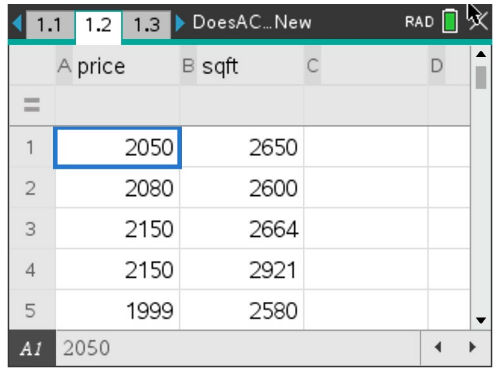
- DoesACorrelationExist_Student.pdf
- DoesACorrelationExist_Student.doc

TI-Nspire document

- DoesACorrelationExist.tns

Discussion Points and Possible Answers

Problem 1: Price of House and Square Footage

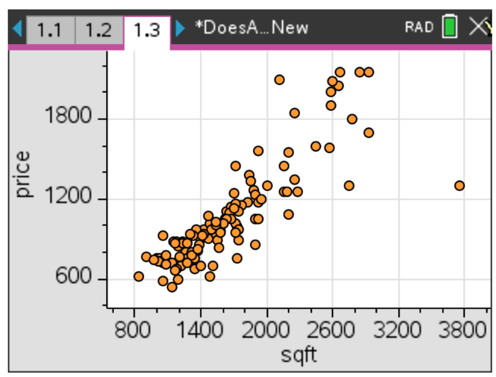


Students are given a data set and are asked to determine the independent and dependent variables. Then, they will graph the data, determine if the correlation is positive or negative, and assess the relative strength of the correlation. Depending on previous discussions, it may be important to help guide the discussion for this first problem.

In the spreadsheet on page 1.2, two columns of data are given. One lists the selling price of houses (given in hundreds of dollars) and the second lists the square footage of the house.

1. Describe how you think the selling price of a house relates to the amount of area of the house or square footage. State if there is any correlation. State which variable is the independent variable. State which is the dependent variable. Explain. Discuss what else the price of a house might depend upon.

   Answer: The price of a house depends on the square footage (among other things), so the dependent variable is the price of the house and the independent variable is the square footage.



On page 1.3, have students create the scatter plot.

2. Explain the meaning of the point (2650, 2050). Include units.

   Answer: A house that has 2,650 square feet is priced at $205,000.

3. Choose the type of correlation (one from each row).

   Positive Negative

   Very strong Moderately strong Moderately weak Very weak

   Answer: Positive; moderately strong

4. Predict the value of the correlation coefficient. Explain your reasoning.

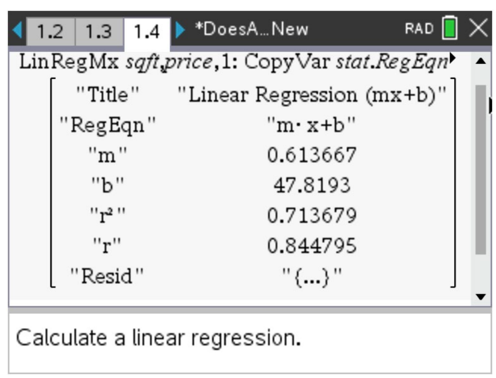
Sample Answer: I predict that the correlation coefficient is 0.75 because the data seem to have a positive linear relationship for lower square foot values. This linear relationship seems weaker at higher square foot values.

> **Teacher Tip:** Have students consider the following questions:
> - Does there have to be an independent and dependent variable in each relation?
> - Will the correlation change if the variables are switched? (This can be investigated easily by switching the variables and performing the calculations again.)
> - How does one know if the correlation is positive or negative?
> - What would the scatter plot look like if the correlation coefficient is 1? –1? 0?
> - Consider correlation versus causation. If the data are correlated, does this mean that one variable causes the other? Does an increase in the square footage of a house cause the price of the house to increase?

TI-Nspire Navigator Opportunity: *Class Capture*

See Note 1 at the end of this lesson.

Have students determine the linear regression equation on page 1.4. Select **MENU (or 🔧) > Statistics > Stat Calculations > Linear Regression (mx+b)**. In the dialog box, choose your independent variable for the **X list** and your dependent variable for the **Y list**. Save your regression equation to **f1.** (This should be in the box already.) Ignore the other boxes.



Tech Tip: To determine the linear regression, have students select **MENU > Statistics > Stat Calculations > Linear Regression (mx+b)**. In the dialog box, have them choose the independent variable for the **X list** and the dependent variable for the **Y list**. They should save the regression equation to **f1.** (This should be in the box already.) Instruct students to ignore the other boxes.

Tech Tip: To determine the linear regression, have students select **> Statistics > Stat Calculations > Linear Regression (mx+b)**.

Teacher Tip: Have students consider the following questions:

- Can we make a prediction for any *x*-values?
- Can we make a prediction for any *y*-values?

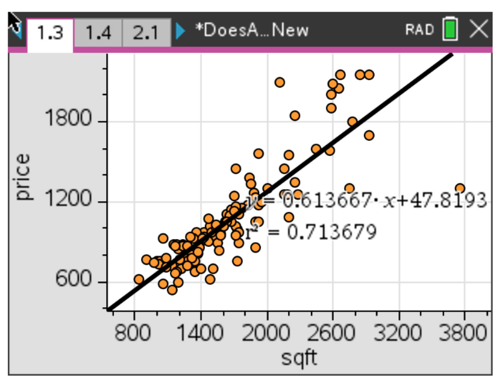
5. Find the correlation coefficient, *r.* Describe how the coefficient compares with your description of the correlation. Explain how your prediction compares.

Answer: The correlation coefficient is 0.84. This is consistent with the description of the correlation, and it is close to the predicted value.

6. Write down the regression equation.

Answer: $f(x) = 0.61x + 47.82$

Have students return to the scatter plot on page 1.3 and graph the regression equation. Have them select **MENU (or** **) > Analyze > Regression > Show Linear (mx+b)**. The line and its equation will appear.



7. State the sign of the slope. Explain how this relates to the sign of the correlation coefficient. Describe the meaning of the slope in the context of the data. Also explain the *y*-intercept in the context of the data.

   Answer: The slope is a positive 0.614, just as the correlation coefficient was positive. The *y*-intercept is 47.819. Students may think the *y*-intercept should be zero since this relates to the price of a house with zero square feet, however the cost of a home also includes the land. On average each square foot costs 0.614 hundred dollars, or it costs $61.4 per square foot.

Have students use the regression equation on page 1.4 to answer the following questions.

> **Teacher Tip:** The final part of this problem asks students to predict values based on the regression equation. Students can use the **nSolve** command to numerically solve some problems. This can provide an opportunity to discuss an appropriate domain and range to make predictions within and the difference of interpolating (inferring within the data) and extrapolating.

8. Predict the price of a house that has 3,500 square feet.

   Answer: **f1**(3500) = 2195.65. The house would cost $219,565.

9. Predict the number of square feet for a house costing $150,000.

   Answer: nSolve(**f1**(*x*) = 1500,*x*) = 2366.4. The house would have approximately 2,366.4 sq. ft.

10. Predict the price of a house with 50,000 sq. ft. State if this prediction seems reasonable based on the data given. Explain.

    Answer: **f1**(50000) = 30731.20. The house would cost $3,073,120. 50,000 is outside of the domain of the given data, and so the prediction is not trustworthy.

11. Predict the number of square feet for a house costing $5.2 million. State if this prediction seems reasonable based on the given data. Explain.

    Answer: nSolve(**f1**(*x*) = 52000,*x*) = 84658.6. The house would have approximately 84,658.6 sq. ft. This prediction is probably too high since the data appear to have a non-linear relationship at higher square footage values.

Problem 2: S.A.T Verbal and Math Scores

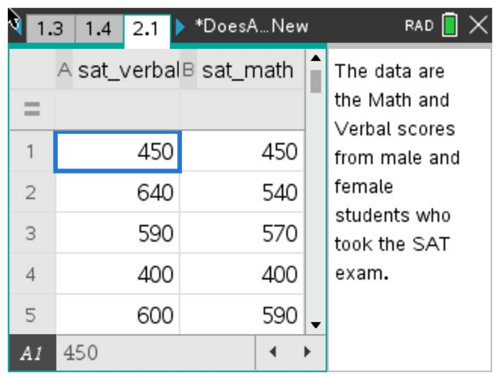
Teacher Tip: Problems 2 and 3 are similar to Problem 1, except they have different data sets. Students are asked to make predictions about correlation and to make extrapolations based on regression equations. For each of these problems, have students consider the following:

- Discuss the relationship between the two variables.
- Ask students: Is one variable dependent on the other? (Have students consider correlation vs. causation.) That is, does an increase in student spending in a state mean that teacher salaries necessarily increase?

TI-Nspire Navigator Opportunity: ***Quick Poll***

See Note 2 at the end of this lesson.

On page 2.1, the spreadsheet contains two columns of data. One lists the SAT verbal scores and the second lists the SAT math scores from male and female students who took the SAT exam.



12. State if you think the students who score well on the Verbal section of the SAT exam also score well on the Math section. Discuss and record your thoughts on which variable is the independent and dependent variable. State if you think there will be a correlation.

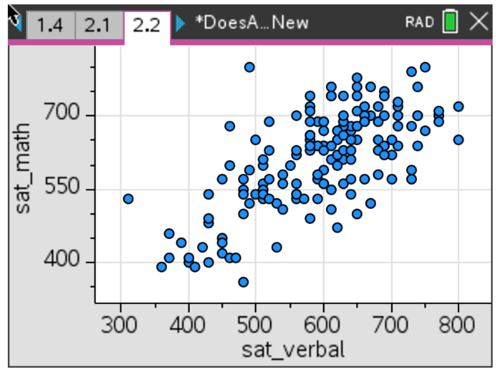
    <u>**Answer:**</u> Students may realize that a student who does well in one section often does well in both sections. However, there is no clear dependent or independent variable. Verbal scores do not depend on Math scores or vice versa.

13. Create the scatter plot on page 2.2. Choose the type of correlation (circle your answer).

a. positive negative

State if you think the correlation is stronger, weaker, or about the same as the data set from Problem 1.

Answer: The data show a positive correlation that looks similar to the data set from Problem 1.



Teacher Tip: Have students consider the correlation if the variables were switched. This can be investigated easily by switching the variables and performing the calculations again. If students explore this they will discover that, while the regression equation changes, the correlation coefficient does not change.

On page 2.3, have students determine the linear regression equation.

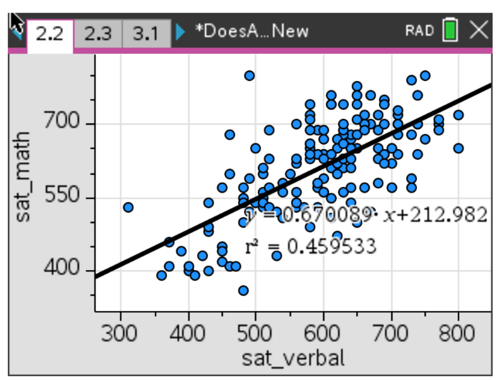
14. State the correlation coefficient.

Answer: The correlation coefficient is 0.83.

15. Record the regression equation and explain the meaning of the slope.

Answer: $y = 0.67x + 213$. For every increase of 100 in the Verbal score, the Math section improves by 67.

Have students return to the scatter plot and graph the regression equation. Have them use the regression equation to answer the following questions.



16. Predict the Math score if the Verbal is 500.

Answer: The Math score would be 548.

17. Predict the Verbal score if the Math score is 620.

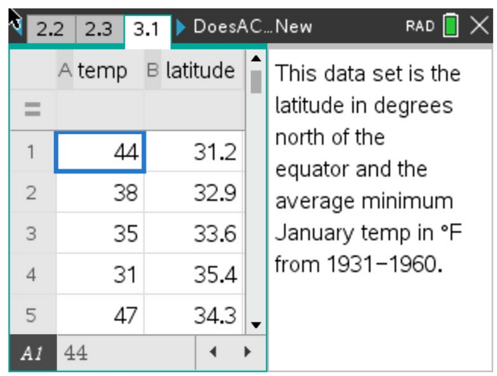
    Answer: The verbal score would be 607.

18. State if there is a relationship between these two variables. State if one is dependent on the other. State if an increase in one means an increase in the other. In other words, while there is correlation, discuss if there is causation.

    Answer: Yes, there is a relationship. As the one increases, the other increases, however one is not dependent on the other. There is a correlation, but not necessarily causation.

Problem 3: Latitudes and Temperatures in January

On page 3.1, the data set is the latitude in degrees north of the equator and the average minimum January temp in °F (1931–1960).

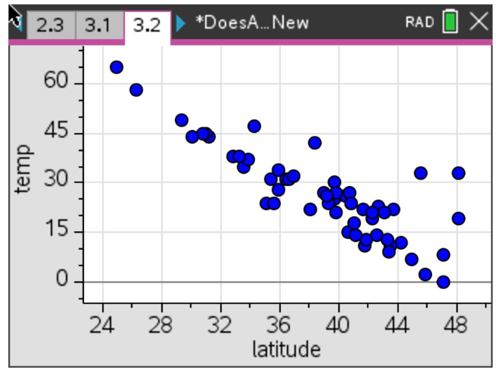


Teacher Tip: In Problem 3, have students consider the following:
- Discuss the change in units from Fahrenheit to Celsius.
- Does a change in unit affect a regression line?
- Have students consider correlation versus causation. Does a change in latitude cause a change in temperature at a particular location?

19. State if you think the latitude of a location is related to the temperature at that location. Discuss and record you thoughts. State the independent and dependent variables. Discuss the other variables that affect the temperature of a location.

    Answer: Latitude is the independent variable and temperature is the dependent variable. This is because the temperature in a given location depends on the latitude of that location on Earth. As you go further North, the temperature decreases. As the latitude increases (relative to the equator), the temperature decreases, but there are other factors to consider, such as mountains or elevation. Also ocean currents keep some coastal cities warmer – Anchorage, Alaska is an example.

Create a scatter plot on page 3.3



20. Predict the type of correlation (circle your answer).

    a. positive negative

    b. very strong moderately strong
    moderately weak very weak

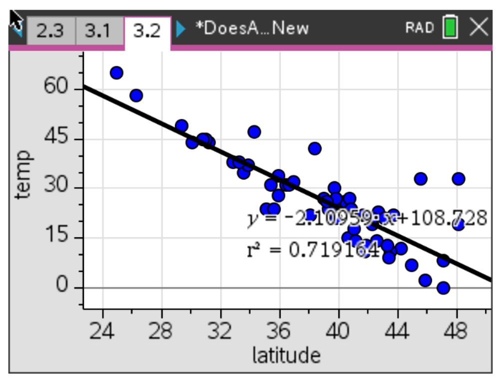
    Answer: Students should circle negative and very strong, but since this is a prediction before they even see the data, answers may vary.

On page 3.3, find the linear regression equation.

21. State the correlation coefficient.

    Answer: The correlation coefficient is −0.85.

Return to the scatter plot and graph the regression equation. Use the regression equation to answer the following questions:



22. Record the regression equation and explain the meaning of the slope and *y*-intercept.

    Answer: $y = -2.11x + 109$. This means that for every degree north of the equator, the temperature drops about 2.3 ºF.

    The *y*-intercept implies that the temperature at the equator is approximately 109 ºF.

23. Predict the temperature for a city with latitude 28.3.

    Answer: **f1**(28.3) = 49.29. The temperature for the city would be 49.29°F.

24. Predict the latitude for a city with an average minimum temperature of 46°F.

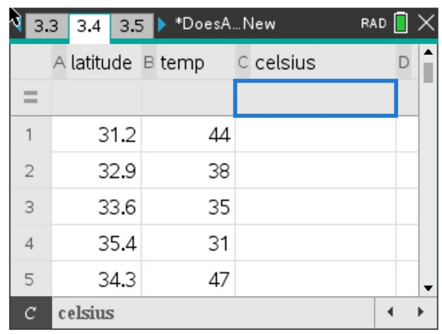
    Answer: nSolve(**f1**(*x*) = 46,*x*) = 28.57. The latitude of the city would be 28.57.

25. Let's investigate what would happen if the temperatures were changed from Fahrenheit to Celsius. If you know that 0 ºC is 32 ºF and 100 ºC is 212 ºF, state the formula to convert the temperature in degrees Fahrenheit to a temperature in degrees Celsius.

Answer: Since $C = 5/9\ (F – 62)$,

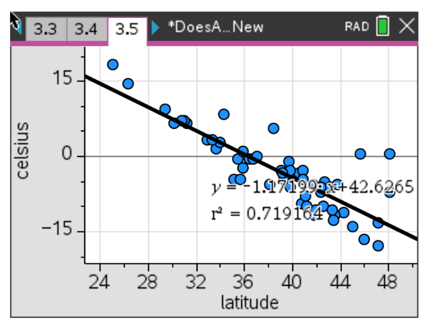
students should enter 5/9(temp – 32)

26. On page 3.4, create a third list that converts the temperatures to Celsius by entering a formula in the grey cell of Column C. Draw a new scatter plot on page 3.5 using the "Celsius" variable instead of "temp" and find a new regression line on page 3.6.



Answer: *y = -1.17 x + 42.6*

The correlation is still -0.85.



27. Describe what happened to the plot of Celsius vs. Latitude compared to the Fahrenheit vs. Latitude. Explain.

Answer: The correlation is the same. Because the conversion from Fahrenheit to Celsius is a linear conversion, the correlation between the temperature and latitude variables remains constant. The slope decreased and the data were shifted down vertically.

TI-Nspire Navigator Opportunities

Note 1
Problem 1, *Class Capture*

This would be a good place to do a class capture to verify students have entered the equation correctly. Viewing different class captures of student work throughout the lesson can be used to help students verify they have the right information.

Note 2
Problems 1–3, *Quick Poll*

You may choose to use *Quick Poll* to assess student understanding. The worksheet questions can be used as a guide for possible questions to ask.

***Note: This activity has been developed independently by Texas Instruments and aligned with the IB Mathematics curriculum, but is not endorsed by IB™. IB is a registered trademark owned by the International Baccalaureate Organization.*