Does A Correlation Exist?

MATH NSPIRED

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
- linear regression
- outlier

About the Lesson

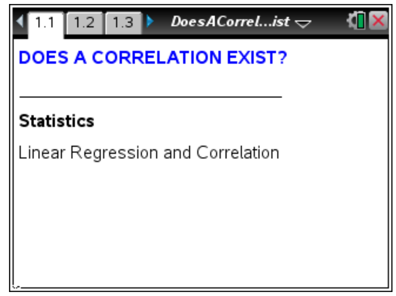
- In this activity, students will determine, by examining a graph, if a data set has a positive or negative correlation coefficient. Then, they will find the linear regression equation and calculate the correlation coefficient. They will use this line to predict the value of *y* for a given *x* and vice-versa.
- Students should have prior experience using the TI-Nspire handheld to create scatter plots. In addition, they should have had some discussion about correlation. Students should also know how to solve an equation using the solver.
- Notes for using the TI-Nspire™ Navigator™ System are included throughout the activity. The use of the Navigator System is not necessary for completion of this activity.

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 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

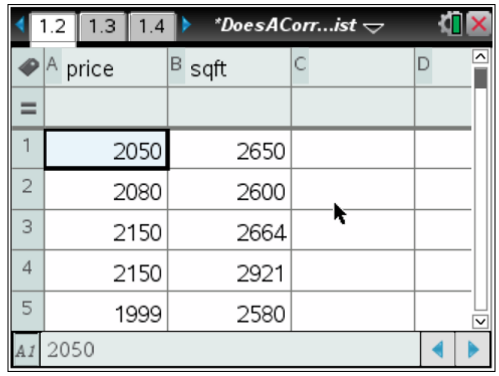
TI-Nspire document

- DoesACorrelationExist.tns
- DoesACorrelationExist_Soln.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.



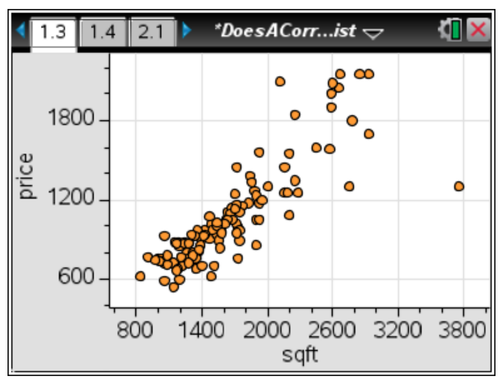
	A price	B sqft	C	D
=				
1	2050	2650		
2	2080	2600		
3	2150	2664		
4	2150	2921		
5	1999	2580		

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. Which variable is the independent variable? Which is the dependent variable? Explain.

   <u>Answer:</u> 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. Choose the type of correlation (one from each row).

   Positive Negative

   Very strong Moderately strong Moderately weak Very weak

   <u>Answer:</u> Positive; moderately strong

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

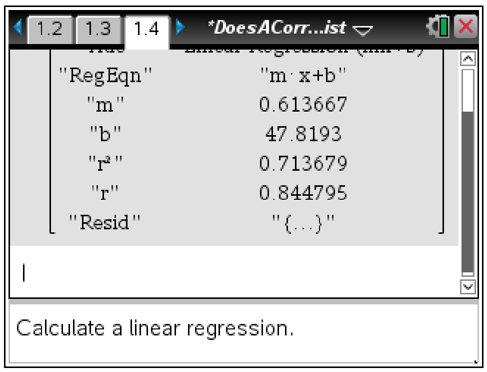
   Sample Answer: I predict that the correlation coefficient is 0.75 because the data seem to have a positive linear relationship for lower square foot values. This linear relationship seems weaker at higher square foot values.

> **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?

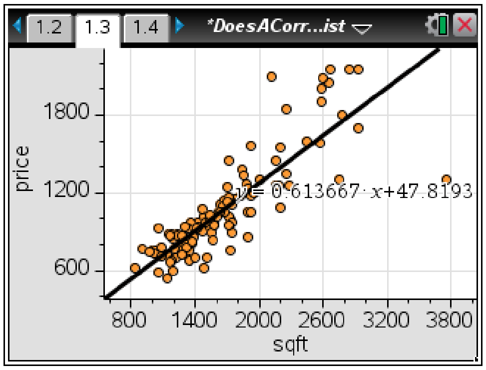
4. What is the regression equation?

   <u>Answer:</u> f(*x*) = 0.61*x* + 47.82

5. What is the correlation coefficient? How does it compare with your description of the correlation? How does your prediction compare?

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

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.



6. Determine the slope and *y*-intercept of the linear regression and describe what each means in the context of the data.

   <u>Answer:</u> The slope is 0.614 and 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,

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

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

8. 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.

9. Predict the price of a house with 50,000 sq. ft. Does this prediction seem 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.

10. Predict the number of square feet for a house costing $5.2 million. Does this prediction seem 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: Teacher Salary and Student Spending

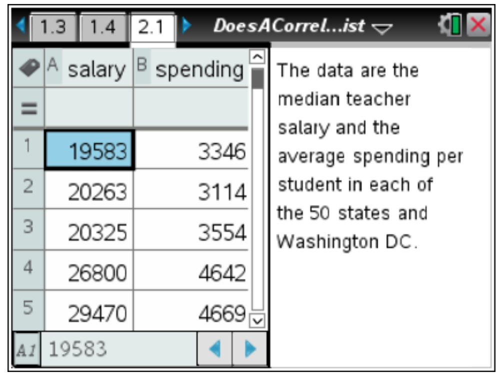
Teacher Tip: Problems 2 and 3 are similar to Problem 1, except they have different data sets. Students are asked to make predictions about correlation and to make extrapolations based on regression equations. For each of these problems, have students consider the following:

- 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 median teacher salary and the second lists the average spending per student in each of the fifty states.

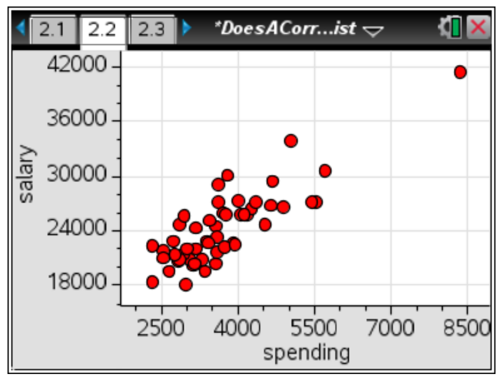


11. Which variable is the independent variable? Which is the dependent variable?

    Answer: The average spending per student in each of the fifty states is the independent variable and the teacher salary is the dependent variable. The median teacher salary depends upon the independent variable of the average spending per student.

12. Create the scatter plot on page 2.2. Describe the type of correlation.

    Answer: There is a very strong positive correlation.



Teacher Tip: Ask students what they notice about the scatter plot on page 2.2. Is there an extreme point in the data, and if so, what is this data point? (The extreme data point represents data from Alaska.)

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

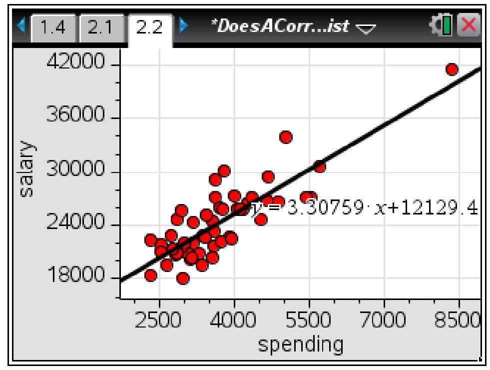
    Answer: The correlation coefficient will be around 0.8 because there appears to be a strong positive linear relationship between the variables.

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

14. What is the correlation coefficient? How does your prediction compare with the calculated correlation coefficient?

    Answer: The correlation coefficient is 0.83. This is very close to the predicted correlation coefficient.

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



15. Predict the amount that a state will spend per student if the median teacher salary is $40,000.

    Answer: nSolve(**f1**(x) = 40000,x) = 8426.28. The state would spend approximately $8,426.28 per student.

16. Predict the teacher salary for a state that spends, on average, $1,500 per student.

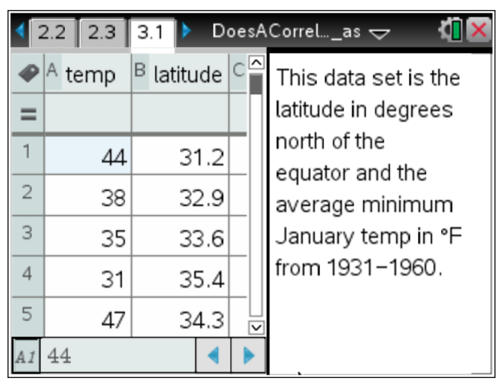
    Answer: **f1**(1500) = 17090.7. The teacher salary would be $17,090.70.

17. Is there a relationship between these two variables? Is one dependent on the other? Does an increase in one mean an increase in the other? In other words, while there is correlation, is there causation?

    Answer: While the data are correlated, there is not necessarily a causation between the variables. It is not necessarily the case that increasing the state spending per student would cause a direct increase in teacher salary.

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?

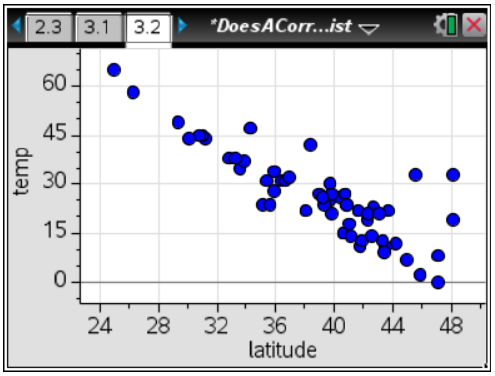
18. Which variable is the independent variable? Which is the dependent variable? Explain.

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.

Create a scatter plot on page 3.3

19. Describe the type of correlation between the variables.

Answer: The data show a strong negative correlation.



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

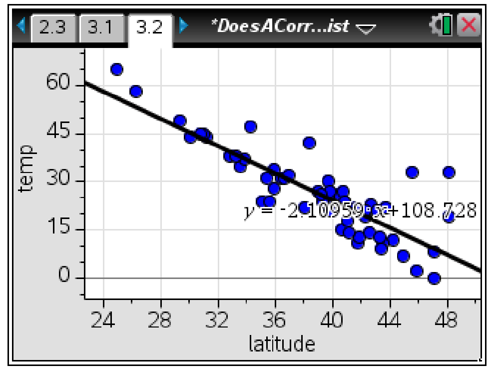
Answer: −0.9. Because the slope is negative, the correlation is negative. With the exception of some outliers, he data show a strong linear correlation.

On page 3.3, find the linear regression equation.

21. What is the correlation coefficient? How does your prediction compare? How does it compare with your description of the correlation?

Answer: The correlation coefficient is −0.85. This is consistent with the description of the correlation because it represents a strong negative correlation. This is also very close to the prediction of the correlation coefficient.

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



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

    <u>Answer:</u> **f1**(28.3) = 49.03. The temperature for the city would be 49.03°F.

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

    <u>Answer:</u> nSolve(**f1**(*x*) = 46,*x*) = 29.73. The latitude of the city would be 29.73.

24. Let's investigate what would happen if the temperatures were changed from Fahrenheit to Celsius. 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. Label the column "Celsius." 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. What happened?

    <u>Answer:</u> 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.

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.