

B – Graphs and Statistics, Lesson 6, Correlation Coefficient (r. 2018)

GRAPHS AND STATISTICS

Correlation Coefficient

Common Core Standard	Next Generation Standard
S-ID.C.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.	AI-S.ID.8 Calculate (using technology) and interpret the correlation coefficient of a linear fit.

LEARNING OBJECTIVES

Students will be able to:

- 1) Calculate the correlation coefficient of a linear fit.
- 2) Interpret the meaning of a correlation coefficient.

Overview of Lesson

Teacher Centered Introduction	Student Centered Activities
Overview of Lesson <ul style="list-style-type: none">- activate students' prior knowledge- vocabulary- learning objective(s)- big ideas: direct instruction- modeling	guided practice ←Teacher: anticipates, monitors, selects, sequences, and connects student work <ul style="list-style-type: none">- developing essential skills- Regents exam questions- formative assessment assignment (exit slip, explain the math, or journal entry)

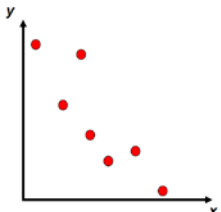
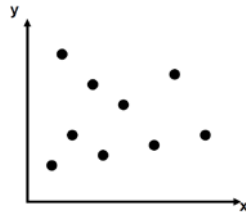
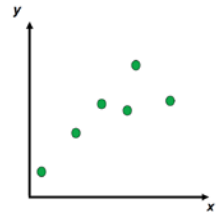
VOCABULARY

correlation coefficient A number between -1 and 1 that indicates the strength and direction of the linear relationship between two sets of numbers. The letter “r” is used to represent correlation coefficients. In all cases, $-1 \leq r \leq 1$.

BIG IDEAS

SIGNS OF CORRELATIONS

The **sign of the correlation** tells you if two variables increase or decrease together (positive); or if one variable increase when the other variable decreases (negative). The **sign of the correlation** also provides a general idea of what the graph will look like.

<div style="text-align: center;">  </div> <p style="text-align: center;"><u>Negative Correlation</u></p> <p>In general, one set of data decreases as the other set increases.</p> <p>An example of a negative correlation between two variables would be the relationship between absenteeism from school and class grades. As one variable increases, the other would be expected to decrease.</p>	<div style="text-align: center;">  </div> <p style="text-align: center;"><u>No Correlation</u></p> <p>Sometimes data sets are not related and there is no general trend.</p> <p>A correlation of zero does not always mean that there is no relationship between the variables. It could mean that the relationship is not <u>linear</u>. For example, the correlation between points on a circle or a regular polygon would be zero or very close to zero, but the points are very predictably related.</p>	<div style="text-align: center;">  </div> <p style="text-align: center;"><u>Positive Correlation</u></p> <p>In general, both sets of data increase together.</p> <p>An example of a positive correlation between two variables would be the relationship between studying for an examination and class grades. As one variable increases, the other would also be expected to increase.</p>
---	---	---

STRENGTH OF CORRELATIONS

	-1	Perfect	+1	
N	-.9		.9	P
E	-.8	Strong	.8	O
G	-.7		.7	S
A	-.6		.6	I
T	-.5	Moderate	.5	T
I	-.4		.4	I
V	-.3		.3	V
E	-.2	Weak	.2	E
	-.1		.1	
		0	None	

- The closer the absolute value of the correlation is to 1, the stronger the correlation between the variables.
- The closer the absolute value of the correlation is to zero, the weaker the correlation between the variables.
- In a perfect correlation, when $r = \pm 1$, all data points balance the equations and also lie on the graph of the equation.

How to Calculate a Correlation Coefficient Using a Graphing Calculator:

STEP 1. Press **STAT** **EDIT** **1:Edit**.

STEP 2. Enter bivariate data in the L1 and L2 columns. All the x-values go into L1 column and all the Y values go into L2 column. Press **ENTER** after every data entry.

STEP 3. Turn the diagnostics on by pressing **2ND** **CATALOG** and scrolling down to **DiagnosticOn**.

Then, press **ENTER** **ENTER**. The screen should respond with the message **Done**. **NOTE: If Diagnostics are turned off, the correlation coefficient will not appear beneath the regression equation.**

Step 4. Press **STAT** **CALC** **4:4-LinReg (ax+b)** **ENTER** **ENTER**

Step 5. The r value that appears at the bottom of the screen is the correlation coefficient.

L1	L2	L3	2	CATALOG	DiagnosticOn
186	495	-----		DependAuto	Done
181	477			det(
176	425			DiagnosticOff	
149	322			DiagnosticOn	
184	482			dim(
190	587			Disp	
158	370			DispGraph	
L2(1)=495					

EDIT	CALC	TESTS	LinReg
1:1-Var Stats			y=ax+b
2:2-Var Stats			a=4.587538544
3:Med-Med			b=-346.6018795
4:LinReg(ax+b)			r ² =.8878031316
5:QuadReg			r=.9422330559
6:CubicReg			
7↓QuartReg			

DEVELOPING ESSENTIAL SKILLS

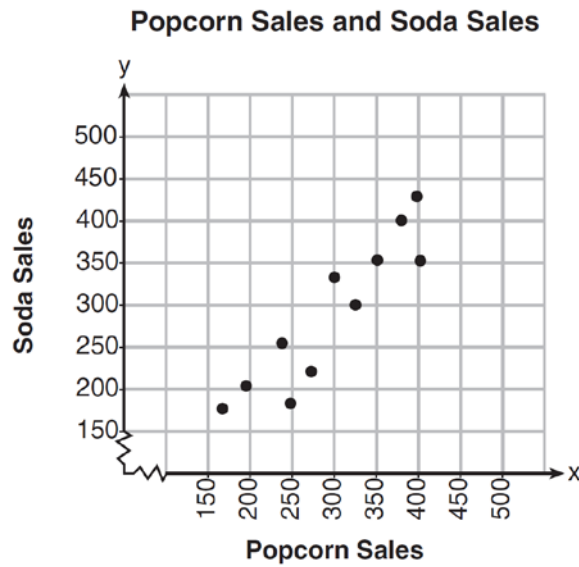
Interpret the following correlation coefficients:

Correlation Coefficient	Interpretation (must include strength and direction)
$r = .5$	Moderate Positive
$r = -.6$	Moderate Negative
$r = -1$	Strong Negative (Perfect)
$r = .7$	Strong Positive
$r = -.9$	Strong Negative
$r = .0$	No Correlation
$r = .2$	Weak Positive

REGENTS EXAM QUESTIONS (through June 2018)

S.ID.C.8: Correlation Coefficients

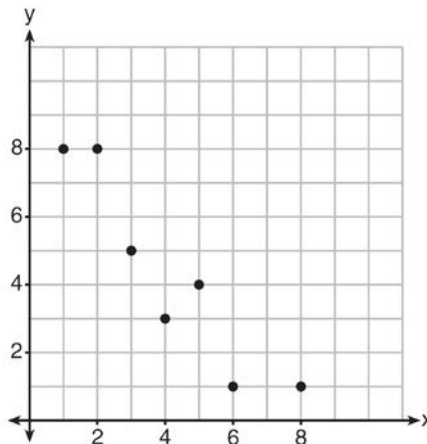
- 32) The scatterplot below compares the number of bags of popcorn and the number of sodas sold at each performance of the circus over one week.



Which conclusion can be drawn from the scatterplot?

- | | |
|--|--|
| 1) There is a negative correlation between popcorn sales and soda sales. | 3) There is no correlation between popcorn sales and soda sales. |
| 2) There is a positive correlation between popcorn sales and soda sales. | 4) Buying popcorn causes people to buy soda. |

33) What is the correlation coefficient of the linear fit of the data shown below, to the *nearest hundredth*?



- | | |
|---------|----------|
| 1) 1.00 | 3) -0.93 |
| 2) 0.93 | 4) -1.00 |

34) Analysis of data from a statistical study shows a linear relationship in the data with a correlation coefficient of -0.524. Which statement best summarizes this result?

- | | |
|--|--|
| 1) There is a strong positive correlation between the variables. | 3) There is a moderate positive correlation between the variables. |
| 2) There is a strong negative correlation between the variables. | 4) There is a moderate negative correlation between the variables. |

35) Bella recorded data and used her graphing calculator to find the equation for the line of best fit. She then used the correlation coefficient to determine the strength of the linear fit. Which correlation coefficient represents the strongest linear relationship?

- 1) 0.9
- 2) 0.5
- 3) -0.3
- 4) -0.8

36) The results of a linear regression are shown below.

$$y = ax + b$$

$$a = -1.15785$$

$$b = 139.3171772$$

$$r = -0.896557832$$

$$r^2 = 0.8038159461$$

Which phrase best describes the relationship between x and y ?

- 1) strong negative correlation
- 2) strong positive correlation
- 3) weak negative correlation
- 4) weak positive correlation

37) A nutritionist collected information about different brands of beef hot dogs. She made a table showing the number of Calories and the amount of sodium in each hot dog.

Calories per Beef Hot Dog	Milligrams of Sodium per Beef Hot Dog
186	495
181	477
176	425
149	322
184	482
190	587
158	370
139	322

- a) Write the correlation coefficient for the line of best fit. Round your answer to the *nearest hundredth*.
- b) Explain what the correlation coefficient suggests in the context of this problem.

38) The table below shows 6 students' overall averages and their averages in their math class.

Overall Student Average	92	98	84	80	75	82
Math Class Average	91	95	85	85	75	78

If a linear model is applied to these data, which statement best describes the correlation coefficient?

- 1) It is close to -1.
- 2) It is close to 1
- 3) It is close to 0.
- 4) It is close to 0.5.

39) At Mountain Lakes High School, the mathematics and physics scores of nine students were compared as shown in the table below.

Mathematics	55	93	89	60	90	45	64	76	89
Physics	66	89	94	52	84	56	66	73	92

State the correlation coefficient, to the *nearest hundredth*, for the line of best fit for these data. Explain what the correlation coefficient means with regard to the context of this situation.

- 40) The table below shows the attendance at a museum in select years from 2007 to 2013.

Year	2007	2008	2009	2011	2013
Attendance (millions)	8.3	8.5	8.5	8.8	9.3

State the linear regression equation represented by the data table when $x = 0$ is used to represent the year 2007 and y is used to represent the attendance. Round all values to the *nearest hundredth*. State the correlation coefficient to the *nearest hundredth* and determine whether the data suggest a strong or weak association.

- 41) Erica, the manager at Stellarbeans, collected data on the daily high temperature and revenue from coffee sales. Data from nine days this past fall are shown in the table below.

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
High Temperature, t	54	50	62	67	70	58	52	46	48
Coffee Sales, $f(t)$	\$2900	\$3080	\$2500	\$2380	\$2200	\$2700	\$3000	\$3620	\$3720

State the linear regression function, $f(t)$, that estimates the day's coffee sales with a high temperature of t . Round all values to the *nearest integer*.

State the correlation coefficient, r , of the data to the *nearest hundredth*. Does r indicate a strong linear relationship between the variables? Explain your reasoning.

- 42) The percentage of students scoring 85 or better on a mathematics final exam and an English final exam during a recent school year for seven schools is shown in the table below.

Mathematics, x	English, y
27	46
12	28
13	45
10	34
30	56
45	67
20	42

Write the linear regression equation for these data, rounding all values to the *nearest hundredth*. State the correlation coefficient of the linear regression equation, to the *nearest hundredth*. Explain the meaning of this value in the context of these data.

SOLUTIONS

- 32) ANS: 2
 Strategy: Eliminate wrong answers.

- a) Eliminate choice (1) because a negative correlation is a relationship where the dependent (y) values decrease as independent (x) values increase. A graph showing negative correlation would go down from left to right. The graph in this problem does not go down from left to right.
- b) Select choice (2) because a positive correlation is a relationship where the dependent (y) values increase as independent values (x) increase. A graph showing positive correlation would go up from left to right, like the graph in this problem.
- c) Eliminate choice (3) because no correlation occurs when there is no relationship between the dependent (y) values and independent (x) values. A graph showing no correlation would not appear to go up or down or have any pattern.
- d) Eliminate choice (4) because there is no evidence that buying a bag of popcorn causes someone to buy a soda. Causation only occurs when a change in one quantity causes a change in another quantity. For example, doubling the number of cookies baked causes more cookie dough to be used.

PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient

33) ANS: 3

Strategy #1: This problem can be answered by looking at the scatterplot.

The slope of the data cloud is negative, so answer choices a and b can be eliminated because both are positive.

The data cloud suggests a linear relationship, but the dots are not in a perfect line. A perfect correlation of ± 1 would show all the dots in a perfect line. Therefore, we can eliminate answer choice d .

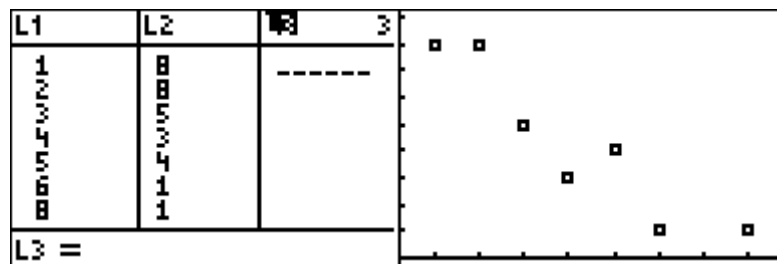
The correct answer is choice c .

DIMS: Does it make sense? Yes. The data cloud shows a negative correlation that is strong, but not perfect. Choice c is the best answer.

Strategy #2: Input the data from the chart in a graphing calculator and calculate the correlation coefficient using linear regression and the diagnostics on feature.

STEP 1. Create a table of values from the graphing view of the function and input it into the graphing calculator.

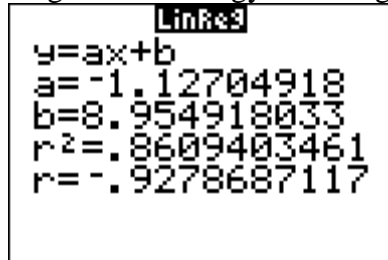
x	y
1	8
2	8
3	5
4	3
5	4
6	1
8	1



- **STEP 2.** Turn diagnostics on using the catalog.



- **STEP 3.** Determine which regression strategy will best fit the data. The graph looks like the graph of an linear function, so choose linear regression.
- **STEP 4.** Execute the appropriate regression strategy with diagnostics on in the graphing calculator.



Round the correlation coefficient to the nearest hundredth: $r = -.93$

- **STEP 4.** Select answer choice c.

DIMS: Does it make sense? Yes. The data cloud shows a negative correlation that is strong, but not perfect. Choice c is the best answer.

PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient and Residuals

34) ANS: 4

A correlation coefficient of -0.524 is both negative and moderate.
A perfect correlation is ± 1 and no correlation is 0.

Strategy: Eliminate wrong answers

- a) There is a strong ~~positive~~ correlation between the variables.
- b) There is a ~~strong~~ negative correlation between the variables.
- e) There is a moderate ~~positive~~ correlation between the variables.
- d) There is a moderate negative correlation between the variables.

PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient

35) ANS: 1

The correlation coefficient with the absolute value closest to 1 indicates the strongest relationship.

$$|.9| > |-0.8| > |0.5| > |-0.3|$$

PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient

36) ANS: 1

The correlation coefficient $r = -0.896557832$ indicates a strong negative correlation.

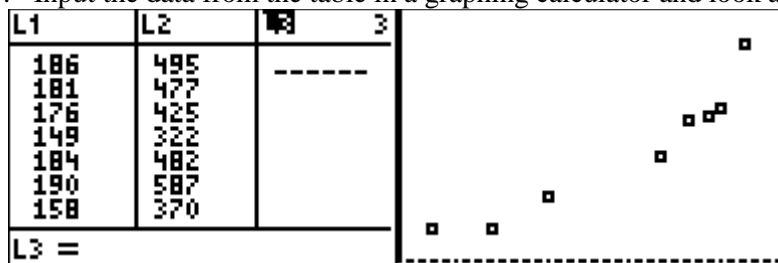
PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient

37) ANS:

$r \approx 0.94$. The correlation coefficient suggests that as calories increase, so does sodium.

Strategy: Use data from the table and a graphing calculator to find both the regression equation and its correlation coefficient.

STEP 1. Input the data from the table in a graphing calculator and look at the data cloud.

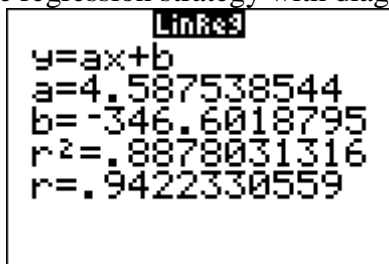


STEP 2. Turn diagnostics on using the catalog.



STEP 3. Determine which regression strategy will best fit the data. The graph looks like the graph of an linear function, so choose linear regression.

STEP 4. Execute the appropriate regression strategy with diagnostics on in the graphing calculator.



Round the correlation coefficient to the nearest hundredth: $r \approx .94$

DIMS: Does it make sense? Yes. The data cloud and the table show a positive correlation that is strong, but not perfect. A correlation coefficient of .94 is positive, but not a perfectly straight line.

PTS: 4 NAT: S.ID.C.8 TOP: Correlation Coefficient and Residuals

38) ANS: 2

A simple inspection of the table shows that high overall student averages are highly correlated in a positive way with math class averages. The actual correlation coefficient for this table is $r \approx .924771\dots$

PTS: 2 NAT: S.ID.C

39) ANS:

$r \approx 0.92$.

The correlation coefficient suggests a strong positive correlation between a student's mathematics and physics scores.

Solution Strategy: Input the data into the stats editor of a graphing calculator and calculate linear regression with diagnostics on.

NORMAL FLOAT AUTO REAL RADIAN MP					NORMAL FLOAT AUTO REAL RADIAN MP					NORMAL FLOAT AUTO REAL RADIAN MP				
L1	L2	L3	L4	L5	DiagnosticOn					LinReg				
55	66	-----	-----	-----	Done					y=ax+b				
93	89									a=.8098804988				
89	94									b=15.18544337				
60	52									r ² =.8492440145				
90	84									r=.9215443638				
45	56													
64	66													
76	73													
89	92													

L2(10)=														

PTS: 2 NAT: S.ID.C.8 TOP: Correlation Coefficient

40) ANS:

$y = 0.16x + 8.27$ $r = 0.97$, which suggests a strong association.

Strategy: Convert the table to data that can be input into a graphing calculator, then use the linear regression feature of the graphing calculator to respond to the question.

STEP 1. Convert the table for input into the calculator.

Attendance at Museum					
Year (L1)	0	1	2	4	6
Attendance (L2)	8.3	8.5	8.5	8.8	9.3

STEP 2. Make sure that STAT DIAGNOSTICS is set to "On" in the mode feature of the graphing calculator. Setting STAT DIAGNOSTICS to on causes the correlation coefficient (r) to appear with the linear regression output.

STEP 3. Use the linear regression feature of the graphing calculator.

L1	L2	L3	EDIT CALC TESTS		LinReg	
0	8.3	-----	1: 1-Var Stats	y=ax+b		
1	8.5		2: 2-Var Stats	a=.1577586207		
2	8.5		3: Med-Med	b=8.269827586		
4	8.8		4: LinReg(ax+b)	r ² =.9496653811		
6	9.3		5: QuadReg	r=.9745077635		
-----			6: CubicReg			
L2(6) =			7: QuartReg			

NOTE: Round the graphing calculator output to the *nearest hundredth* as required in the problem.

STEP 4. Record your solution.

PTS: 4 NAT: S.ID.C.8 TOP: Regression KEY: linear

NOT: NYSED classifies as S.ID.B.6a

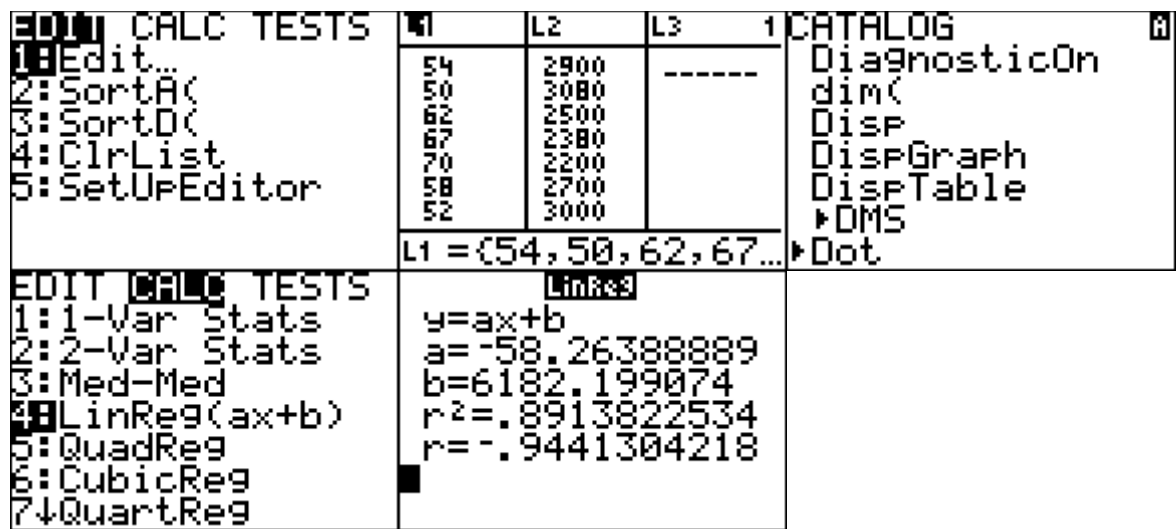
41) ANS:

$$f(t) = -58t + 6182$$

$$r = -.94$$

The correlation coefficient indicates a strong linear relationship because the absolute value of r is close to 1.

Strategy: Input the table of values into the stats-editor of a graphing calculator, then use the stats-calc-linear regression with “diagnostics on” to obtain both the linear regression equation and the correlation coefficient (r). The following screenshots illustrate the solution using a TI-84 family graphing calculator.



PTS: 4 NAT: S.ID.C.8 TOP: Regression KEY: linear
NOT: NYSED classifies as S.ID.B.6

42) ANS:

$$y = 0.96x + 23.95, 0$$

A correlation coefficient value of .92 indicates a strong positive correlation between scores 85 or better on the math and English exams for the seven schools.

Strategy: Use the linear regression function of a graphing calculator to determine the equation and the correlation coefficient for the data in the table.

STEP 1. Input the values from the table into the stats editor of a graphing calculator.	STEP 2. Turn diagnostics on to calculate the correlation coefficient with the regression equation.	STEP 3. Calculate the regression equation and correlation coefficient.																								
<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> </tr> </thead> <tbody> <tr> <td>27</td> <td>46</td> <td>-----</td> </tr> <tr> <td>12</td> <td>28</td> <td></td> </tr> <tr> <td>13</td> <td>45</td> <td></td> </tr> <tr> <td>10</td> <td>34</td> <td></td> </tr> <tr> <td>30</td> <td>56</td> <td></td> </tr> <tr> <td>45</td> <td>67</td> <td></td> </tr> <tr> <td>20</td> <td>42</td> <td></td> </tr> </tbody> </table> <p>L3(1)=</p>	L1	L2	L3	27	46	-----	12	28		13	45		10	34		30	56		45	67		20	42		<p>DiagnosticOn</p> <p>Done</p>	<p>LinReg</p> <p>y=ax+b a=.9577039275 b=23.94864048 r²=.8473127129 r=.9204959059</p>
L1	L2	L3																								
27	46	-----																								
12	28																									
13	45																									
10	34																									
30	56																									
45	67																									
20	42																									

PTS: 4 NAT: S.ID.B.6 TOP: Regression KEY: linear with correlation coefficient