

**B – Graphs and Statistics, Lesson 5, Regression (r. 2018)**

**GRAPHS AND STATISTICS**

**Regression**

Common Core Standard	Next Generation Standard
<p><b>S-ID.B.6</b> Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.</p> <p><b>S-ID.B.6a</b> Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use given functions or choose a function suggested by the context. Emphasize linear, quadratic, and exponential models.                      PARCC: Tasks have real world context. Exponential functions are limited to those with domains in the integers.                      NYSED: Includes the regression capabilities of the calculator.</p>	<p><b>AI-S.ID.6</b> Represent bivariate data on a scatter plot, and describe how the variables’ values are related.  <b>Note: It’s important to keep in mind that the data must be linked to the same “subjects,” not just two unrelated quantitative variables; being careful not to assume a relationship between the actual variables (correlation/causation issue).</b></p> <p><b>AI-S.ID.6a</b> Fit a function to real-world data; use functions fitted to data to solve problems in the context of the data.                      (Shared standard with Algebra II)  <b>Note: Algebra I emphasis is on linear models</b> and includes the regression capabilities of the calculator.</p>

**LEARNING OBJECTIVES**

Students will be able to:

- 1) Draw an approximate line of best fit through a scatterplot.
- 2) Use a graphing calculator to find the equation of the line of best fit for a given set of data.
- 3) Make a prediction using a linear regression equation.

**Overview of Lesson**

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

**VOCABULARY**

Regression  
Line of Best Fit

Scatterplot  
Data Cloud

## BIG IDEAS

**Regression Model:** A function (e.g., linear, exponential, power, logarithmic) that fits a set of paired data. *The model may enable other values of the dependent variable to be predicted.*

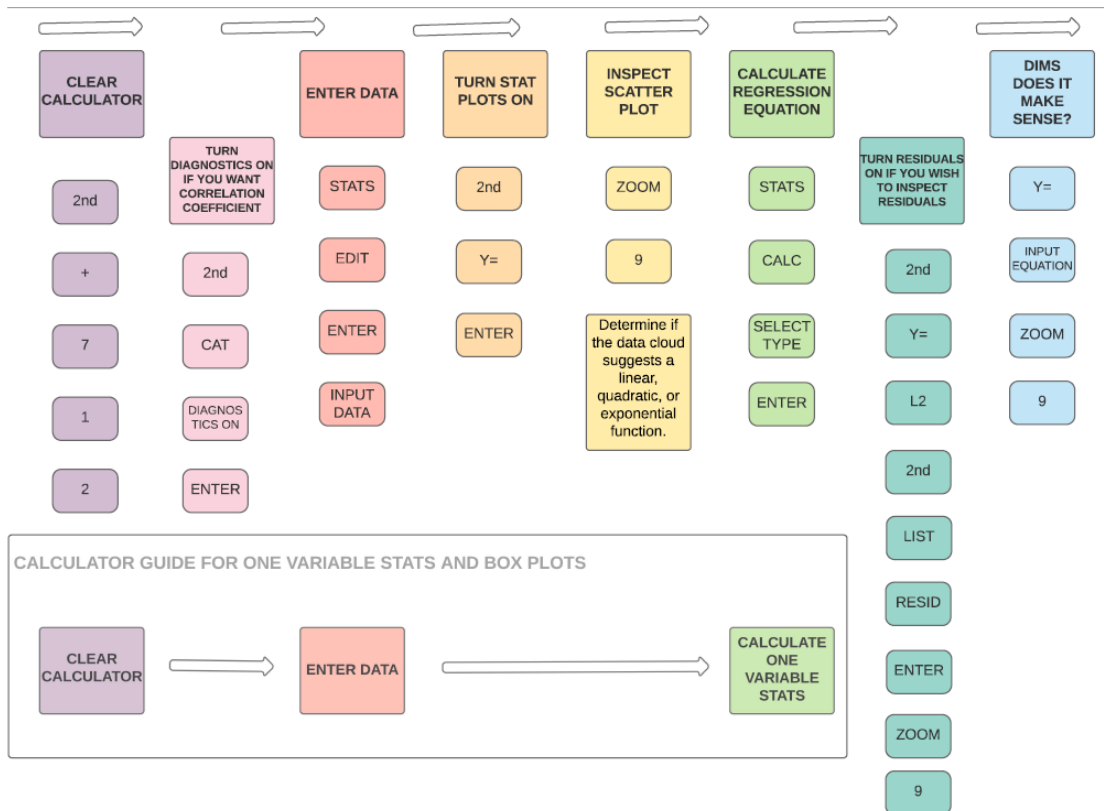
## Big Ideas

The **individual data points** in a **scatterplot** form **data clouds** with shapes that suggest relationships between dependent and independent variables.

A **line of best fit** divides the data cloud into two equal parts with about the same number of data points on each side of the line. A line of best fit can be a straight line or a curved line, depending on the shape of the data cloud.

### Overview of Regression Using TI 83/83 Family of Graphing Calculators

#### CALCULATOR GUIDE FOR REGRESSION

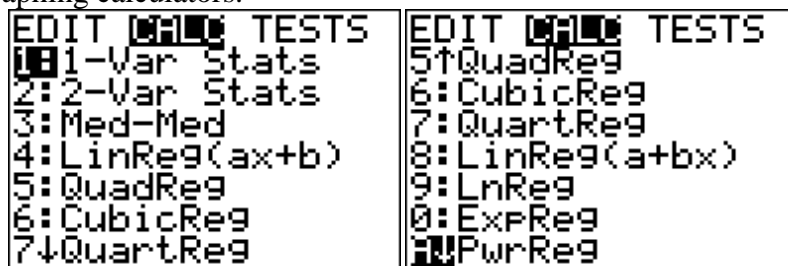


**Calculating Regression Equations.** Technology is almost always used to calculate regression equations.

- **STEP 1.** Use STATS EDIT to Input the data into a graphing calculator.
- **STEP 2.** Use 2nd STAT PLOT to turn on a data set, then ZOOM 9 to inspect the graph of the data and determine which regression strategy will best fit the data.
- **STEP 3.** Use STAT CALC and the appropriate regression type to obtain the regression equation.
- **STEP 4.** Ask the question, “Does it Make Sense (DIMS)?”

### DIFFERENT TYPES OF REGRESSION

The graphing calculator can calculate numerous types of regression equations, but it must be told which type to calculate. All of the calculator procedures described above can be used with various types of regression. The following screenshots show some of the many regressions that can be calculated on the TI-83/84 family of graphing calculators.



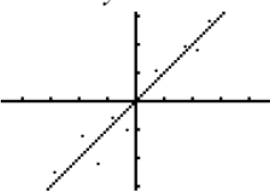
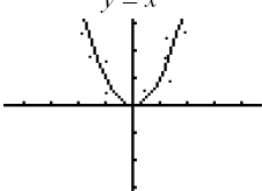
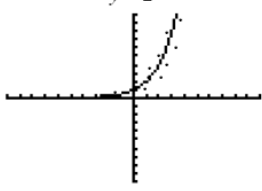
The general purpose of linear regression is to make predictions based on a line of best fit.

### Choosing the Correct Type of Regression to Calculate

There are two general approaches to determining the type of regression to calculate:

- The decision of which type of regression to calculate can be made based on visual examination of the data cloud, or.
- On Regents examinations, the wording of the problem often specifies a particular type of regression to be used.

### Using the Data Cloud to Select the Correct Regression Calculation Program

<p style="text-align: center;"><math>y = x</math></p>  <p>If the data cloud takes the general form of a straight line, use <b><u>linear regression</u></b>.</p>	<p style="text-align: center;"><math>y = x^2</math></p>  <p>If the data cloud takes the general form of a parabola, use <b><u>quadratic regression</u></b>.</p>	<p style="text-align: center;"><math>y = 2^x</math></p>  <p>If the data cloud takes the general form of an exponential curve, use <b><u>exponential regression</u></b>.</p>
<p><b>Note:</b> The general forms of some data clouds are difficult to interpret. In difficult to interpret cases, the strength of the correlation coefficient can be used to determine which type of regression best fits the data. See lesson for standard S.ID.C.8,</p>		

### Drawing a Line of Best Fit on a Scatterplot

A line of best fit may be drawn on a scatterplot of data by using values from the regression equation.

STEP 1. Input the regression equation in the y-editor of a graphing calculator

STEP2. Use ordered pairs of coordinates from the table of values to plot the line of best fit.

- In linear regression, the line of best fit will always go through the point  $(\bar{x}, \bar{y})$ , where  $\bar{x}$  is the mean of all values of  $x$ , and  $\bar{y}$  is the mean of all values of  $y$ . For example, the line of best fit for a scatterplot with points  $(2,5)$ ,  $(4,7)$  and  $(8,11)$  must include the point  $\left(x = \frac{14}{3}, y = \frac{23}{3}\right)$ , because these  $x$  and  $y$  values are the averages of all the  $x$ -values and all the  $y$ -values.

- If the regression equation is linear and in  $y = mx + b$  form, the y-intercept and slope can be used to plot the line of best fit.

### Making Predictions Based on a Line of Best Fit

Predictions may be made based on a line of best fit.

STEP 1. Input the regression equation in the y-editor of a graphing calculator

STEP 2. Use ordered pairs of coordinates from the table of values to identify expected values of the dependent (y) variable for any desired value of the independent (x) variable.

### DEVELOPING ESSENTIAL SKILLS – Class Assignment

Nazmun and Daniel came to America from two different parts of the world. Nazmun measures temperature in degrees Celsius, while Daniel measures temperature in degrees Fahrenheit. They want to understand the relationship between these two different ways of measuring temperature. They each know the temperature when water freezes, when water boils, the temperature outside today, and the temperature inside their very warm classroom. They record these temperatures in the following table.

<u>Comparison Table</u>	Fahrenheit Degrees	Celsius Degrees
Water Freezes	32	0
Water Boils	212	100
Today's Outdoor Temperature	41	5
Temperature in Classroom	77	25

Use linear regression to find the mathematical relationship between degrees Fahrenheit and degrees Celsius. Then, use your regression equation to add three new rows to the comparison table.

The image shows three screenshots from a TI-84 calculator. The first screenshot shows data being entered into lists L1 and L2: L1 contains 32, 212, 41, 77 and L2 contains 0, 100, 5, 25. The second screenshot shows the 'TESTS' menu with '4: LinReg(ax+b)' selected. The third screenshot shows the regression results:  $y = ax + b$ ,  $a = .5555555556$ , and  $b = -17.77777778$ .

The linear regression equation is  $y = .55x - 17.77$   
 $C = .55F - 17.77$

This regression equation can be transformed to a more familiar formula as follows:

$$C = .55F - 17.77$$

$$C = \frac{5}{9}F - 17\frac{7}{9}$$

$$C = \frac{5}{9}F - \frac{160}{9}$$

$$C = \frac{5}{9}F - \left(\frac{5}{9}\right)\left(\frac{\cancel{9}}{5}\right)\frac{160}{\cancel{9}}$$

$$C = \frac{5}{9}F - \left(\frac{5}{9}\right)\frac{160}{5}$$

$$C = \frac{5}{9}F - \left(\frac{5}{9}\right)32$$

$$C = \frac{5}{9}(F - 32)$$

To add three new rows to the comparison table, input the regression formula into the y-editor of a graphing calculator and use the table of values.

NORMAL FLOAT AUTO REAL RADIAN MP				NORMAL FLOAT AUTO REAL RADIAN MP			
Plot1 Plot2 Plot3				PRESS + FOR $\Delta$ Tb1			
$\blacksquare$ Y1 =	.5555X-17.7777			X	Y1		
$\blacksquare$ Y2 =				0	-17.78		
$\blacksquare$ Y3 =				1	-17.22		
$\blacksquare$ Y4 =				2	-16.67		
$\blacksquare$ Y5 =				3	-16.11		
$\blacksquare$ Y6 =				4	-15.56		
$\blacksquare$ Y7 =				5	-15		
$\blacksquare$ Y8 =				6	-14.44		
$\blacksquare$ Y9 =				7	-13.89		
				8	-13.33		
				9	-12.78		
				10	-12.22		
				X=0			

**REGENTS EXAM QUESTIONS (through June 2018)**

**S.ID.B.6: Regression**

- 26) Emma recently purchased a new car. She decided to keep track of how many gallons of gas she used on five of her business trips. The results are shown in the table below.

Miles Driven	Number of Gallons Used
150	7
200	10
400	19
600	29
1000	51

Write the linear regression equation for these data where miles driven is the independent variable. (Round all values to the *nearest hundredth*.)

- 27) About a year ago, Joey watched an online video of a band and noticed that it had been viewed only 843 times. One month later, Joey noticed that the band’s video had 1708 views. Joey made the table below to keep track of the cumulative number of views the video was getting online.

Months Since First Viewing	Total Views
0	843
1	1708
2	forgot to record
3	7124
4	14,684
5	29,787
6	62,381

a) Write a regression equation that best models these data. Round all values to the *nearest hundredth*. Justify your choice of regression equation.

b) As shown in the table, Joey forgot to record the number of views after the second month. Use the equation from part *a* to estimate the number of full views of the online video that Joey forgot to record.

28) The table below shows the number of grams of carbohydrates,  $x$ , and the number of Calories,  $y$ , of six different foods.

Carbohydrates ( $x$ )	Calories ( $y$ )
8	120
9.5	138
10	147
6	88
7	108
4	62

Which equation best represents the line of best fit for this set of data?

1)  $y = 15x$

3)  $y = 0.1x - 0.4$

2)  $y = 0.07x$

4)  $y = 14.1x + 5.8$

29) An application developer released a new app to be downloaded. The table below gives the number of downloads for the first four weeks after the launch of the app.

Number of Weeks	1	2	3	4
Number of Downloads	120	180	270	405

Write an exponential equation that models these data. Use this model to predict how many downloads the developer would expect in the 26th week if this trend continues. Round your answer to the nearest download. Would it be reasonable to use this model to predict the number of downloads past one year? Explain your reasoning.

30) The data table below shows the median diameter of grains of sand and the slope of the beach for 9 naturally occurring ocean beaches.

Median Diameter of Grains of Sand in Millimeters ( $x$ )	0.17	0.19	0.22	0.235	0.235	0.3	0.35	0.42	0.85
Slope of Beach in Degrees ( $y$ )	0.63	0.7	0.82	0.88	1.15	1.5	4.4	7.3	11.3

Write the linear regression equation for this set of data, rounding all values to the *nearest thousandth*. Using this equation, predict the slope of a beach, to the *nearest tenth of a degree*, on a beach with grains of sand having a median diameter of 0.65 mm.

- 31) Omar has a piece of rope. He ties a knot in the rope and measures the new length of the rope. He then repeats this process several times. Some of the data collected are listed in the table below.

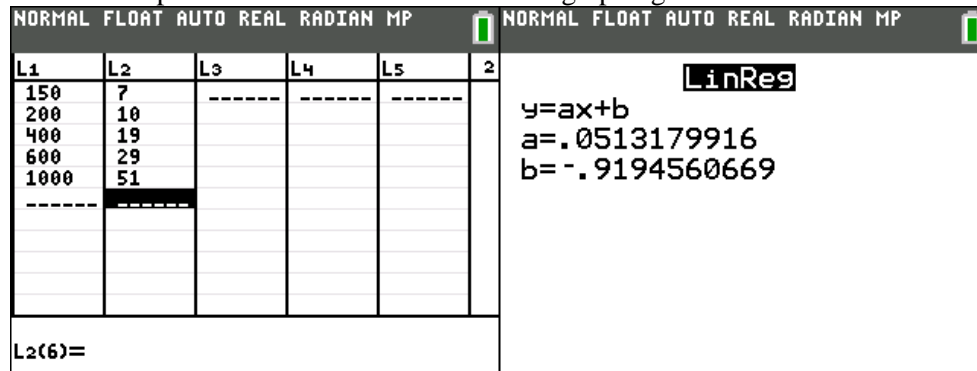
Number of Knots	4	5	6	7	8
Length of Rope (cm)	64	58	49	39	31

State, to the *nearest tenth*, the linear regression equation that approximates the length,  $y$ , of the rope after tying  $x$  knots. Explain what the  $y$ -intercept means in the context of the problem. Explain what the slope means in the context of the problem.

### SOLUTIONS

- 26) ANS:

STEP 1: Input the data in the stats editor of a graphing calculator and calculate linear regression.



$$y = 0.05x - 0.92$$

PTS: 2      NAT: S.ID.B.6a      TOP: Regression

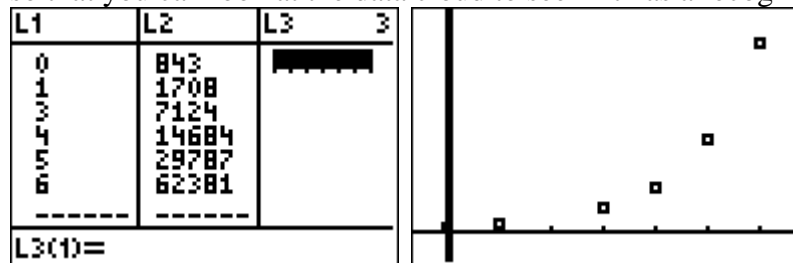
- 27) ANS:

Part a:  $f(x) = 836.47(2.05)^x$  The data appear to grow at an exponential rate.

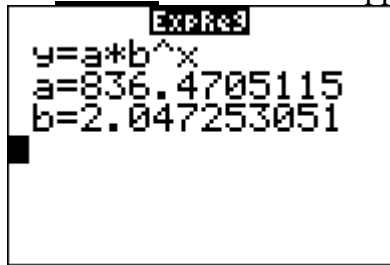
Part b:  $f(2) = 836.47(2.05)^2 = 3515$

Strategy: Input the data into a graphing calculator, inspect the data cloud, and find a regression equation to model the data table, input the regression equation into the y-editor, predict the missing value.

- **STEP 1.** Input the data into a graphing calculator or plot the data cloud on a graph, if necessary, so that you can look at the data cloud to see if it has a recognizable shape.



- **STEP 2.** Determine which regression strategy will best fit the data. The graph looks like the graph of an exponential function, so choose exponential regression.
- **STEP 3.** Execute the appropriate regression strategy in the graphing calculator.



Round all values to the nearest hundredth:  $y = 836.47(2.05)^x$

- **STEP 4.** Input the regression equation into the y-editor feature of the graphing calculator and view the associated table of values to find the value of y when x equals 2.

X	Y1	
0	836.47	
1	1714.8	
2	3515.3	
3	7206.3	
4	14773	
5	30284	
6	62083	

Press + for  $\Delta/|b|$

Round 3515.3 to 3515.

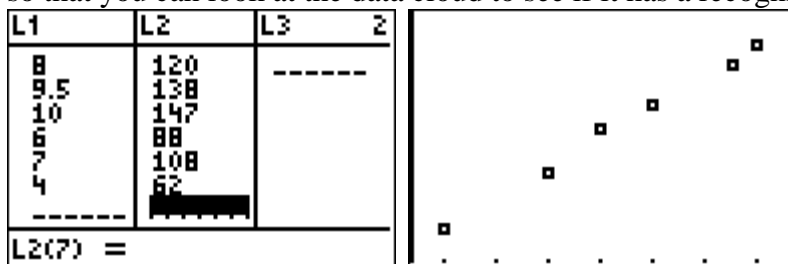
- **STEP 4.** In Ask the question, “Does it Make Sense (DIMS)?” that the missing total number of views in month 2 would be around 3515 views?

PTS: 4 NAT: S.ID.B.6a TOP: Regression

28) ANS: 4

Strategy: Input the data into a graphing calculator, inspect the data cloud, and find a regression equation to model the data table, input the regression equation into the y-editor, predict the missing value.

- **STEP 1.** Input the data into a graphing calculator or plot the data cloud on a graph, if necessary, so that you can look at the data cloud to see if it has a recognizable shape.



- **STEP 2.** Determine which regression strategy will best fit the data. The graph looks like the graph of an linear function, so choose linear regression.
- **STEP 3.** Execute the appropriate regression strategy in the graphing calculator.



```
LinReg
y=ax+b
a=14.11239669
b=5.833057851
```

Write the regression equation in a format that can be compared to the answer choices:  $y = 14.11x + 5.83$

- **STEP 4.** Compare the answer choices to the regression equation and select choice d.

PTS: 2      NAT: S.ID.B.6a      TOP: Regression

29) ANS:

- a)  $y = 80(1.5)^x$
- b)  $80(1.5)^{26} \approx 3,030,140.$
- c) No, because the prediction at  $x = 52$  is already too large.

Strategy: Use data from the table and exponential regression in a graphing calculator.

STEP 1: Model the function in a graphing calculator using exponential regression.

L1	L2	L3	Z
1	120	-----	
2	180		
3	270		
4	405		
-----			

```
ExpReg
y=a*b^x
a=80
b=1.5
```

L2(5) =

The exponential regression equation is  $y = 80(1.5)^x$

STEP 2. Use the equation to predict the number of downloads when  $x = 26$ .

```
80(1.5)^26
3030140.195
```

Rounded to the nearest download, the answer is 3,030,140.

STEP 3. Determine if it would be reasonable to use the model to predict downloads past one year.

```
80(1.5)^53
1.721578051E11
```

It would not be reasonable to use this model to make predictions past one year. The number of predicted downloads is more 170 billion downloads, which is more than 20 downloads in one week for every person in the world.

DIMS? Does It Make Sense? For near term predicitions, yes. For long term predictions, no.

PTS: 4 NAT: S.ID.B.6a TOP: Regression NOT: NYSED classifies this as A.CED.A.2

30) ANS:

$$y = 17.159x - 2.476$$

$$y = 17.159(0.65)x - 2.476$$

$$y \approx 8.7$$

$$y = 17.159x - 2.476. y = 17.159(.65) - 2.476 \approx 8.7$$

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 the linear regression equation, then use the linear regression equation to calculate the value of y when  $x = 0.65$ .

PTS: 4 NAT: S.ID.B.6

31) ANS:

STEP 1.

Input values from the table in the stats editor of a graphing calculator, then calculate linear regression.

The screenshot shows a TI-84 Plus calculator interface. On the left, a data table is displayed with columns L1 through L5 and a cursor in L2(6). The table contains the following data:

L1	L2	L3	L4	L5
4	64	---	---	---
5	58	---	---	---
6	49	---	---	---
7	39	---	---	---
8	31	---	---	---
---	---	---	---	---

On the right, the LinReg screen shows the following results:

```

LinReg
y=ax+b
a=-8.5
b=99.2
r²=.9940836544
r=-.9970374388
  
```

At the bottom left, the cursor is on the line  $L2(6)=$ .

Regression equation:  $y = -8.5x + 99.2$

STEP 2. Explain what the y-intercept means in this equation.

When there are no knots, the rope is 99.2 cm. long.

STEP 3. Explain what the slope means in the context of this problem.

The slope is -8.5 and means there is a negative relationship between the number of knots and the length of the rope. Each knot makes the rope 8.5 cm. shorter.

$y = -8.5x + 99.2$  The y-intercept represents the length of the rope without knots. The slope represents the decrease in the length of the rope for each knot.

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