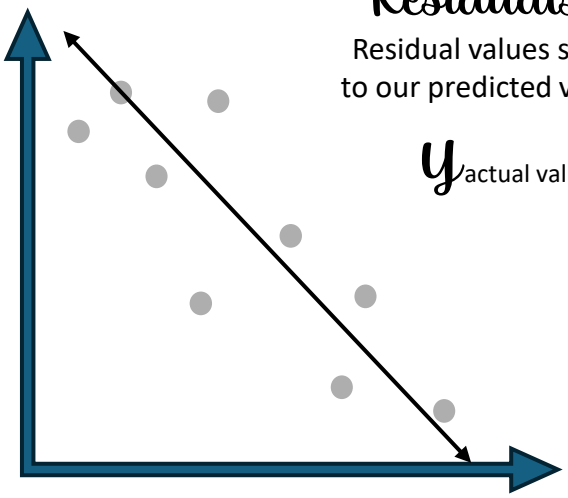
Scatterplots & Linear Regression

Types of Correlation (aka Association)

POSITIVE CORRELATION | NEGATIVE CORRELATION | NO CORRELATION | NONLINEAR CORRELATION

Line of Best Fit (aka the Regression Line) A straight line that best represents the relationship, or trend, between two variables in a scatterplot. Given an input, we can use this linear model to predict the output.

Guess what?
It can also predict inputs when outputs are known!

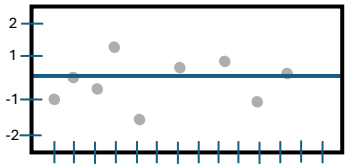


Residuals (think "residue"!)

Residual values show variance in our actual data to our predicted value from the regression model.

$$y_{\text{actual value}} - y_{\text{predicted value}}$$

A look at a Residual Plot



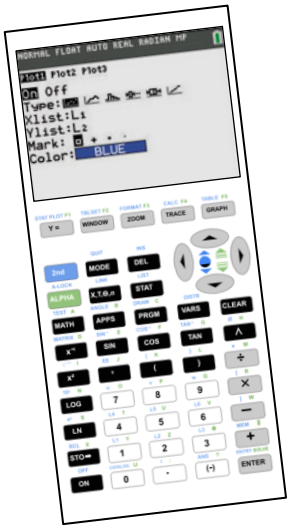
If a data point is...

ON the regression line, the residual value is _____ .

Below the line: _____ **Above** the line: _____

The Bigger the Residual (regardless of sign), the _____ the Correlation!

Let's Get to Calculating!



When working with Statistics, we must make some changes in our calculator before we begin.

1. Turn on Stat Plots 2nd STAT PLOT F1 Y= We only need to turn on ONE stat plot. Press **1 or Enter** to select Stat Plot 1 and then press **Enter** to turn it ON. For the **TYPE** of stat plot, **choose the image that resembles a scatterplot.** Make sure Xlist is L_1 and Ylist is L_2

2. Turn on Stat Diagnostics CE calculators: QUIT MODE then scroll down to Stat Diagnostics and turn ON. Older Caclulators go through the catalog: 2nd CATALOG 0 then scroll to Diagnotics On. Press **Enter** and it will show up on your calculating screen. Press **Enter** again and the calculator will output *Done.*

Flip over for examples →

Study Hrs vs Test Score

Hours	Score
0	62
4.5	96
3	95
2.25	81
4	91
1.5	77
2	81
1.25	74
3	90

a= _____ b=_____
slope y-intercept

$y = ax + b$

Regression Equation:

_____

r = _____

What's this r value?

The Correlation Coefficient is a value between –1 and 1 that measures the direction & strength of the relationship between two variables.

Sketch of Scatterplot and Regression Line

Entering Data

1. STAT Press 1 or Enter for **1:Edit**

2. Enter x-values into **L1** and y-values into **L2**.

Calculating the Regression Equation

Follow Steps 1 and 2 (Entering Data) first.

3. Go back to STAT and use the ▶ to get to the Stat **CALC** menu.

4. Press 4 or scroll to **4:LinReg** & press Enter. Check that **Xlist is L1**, **Ylist is L2**, and that Freq and StoreEq are blank. Press **Enter** or use the down arrow to **CALCULATE**. Press **Enter** to calculate.

Looking at the Scatterplot

Follow Steps 1 and 2 (Entering Data) first.

Press ZOOM and select **9: ZoomStat**.

This guarantees that I can see my data. Pressing the GRAPH button won't always show the correct window.

Viewing the Regression Line on the Scatterplot

Follow Steps 1-4 first.

To graph the regression equation, we need to select Y= . Enter the regression equation into **Y1**. Press ZOOM and select **9: ZoomStat**.

Let's put our model to the test!

In the data I collected, studying 2 hours resulted in an 81. That can be represented at (2, 81). Find this point on the *scatterplot*.
Using my regression model, I can predict a test score after studying 2 hrs:

Rounded to a whole number, this can be represented as (2, _____). Find this point of the *line of best fit*.

Let's go back to 2 hours of studying. Find the difference between my actual value for studying 2 hours and my model's predicted value:

This is my residual value *for this one data point.*

Calculating Residuals on the TI-84:

Follow Steps 1-4 above first.

1. STAT Press 1 or Enter for **1:Edit**

2. You should see your L1 and L2 already. Since we used L3 in a previous step, move your arrow until L4 is highlighted. We are going to choose to make L4 our Residual values. Select 2nd STAT to bring up all the different list names. Choose **7:RESID**. Press Enter. The bottom of the screen should say "L4=RESID". Press **Enter**.

Using your TI-84 to show Predicted Values:

Follow Steps 1-4 above first.

1. STAT brings up our entered data. Use the **arrow button** to go to the very top to highlight **L3**. At bottom of the screen, it should read "L3="

2. Remember L1 are our x-values, and x-values are plugged into our regression equation to find the *predicted* y-values. To have the calculator do this for us, I am going to type the regression equation in, but replace the x in the equation with an L1. To type L1, I need to access my list names under 2nd STAT . (Example: y = 2.3x + 6 should become L3 = 2.3L1 + 6)

3. Once the equation is in for L3, select **Enter**. The predicted values will be generated.

Hours	ACTUAL Score	Predicted (from Model)	Residual
0	62		
4.5	96		
3	95		
2.25	81		
4	91		
1.5	77		
2	81		
1.25	74		
3	90		

LINEAR REGRESSION: TRY IT OUT ON YOUR OWN!

The table below shows the recorded amount of time for a popsicle to melt at observed temperatures.

Temp in ºF (x)	72	76	78	84	88	92
Time in min. (y)	20	15	14	13	11	9

1. Using technology, find the linear regression equation that best fits the given data. *Round to hundredths.*
2. Correlation Coefficient:

   a) What is the value of the correlation coefficient?

   b) Using the correlation coefficient, tell me if the data has positive, negative, or no correlation:

   c) Using the correlation coefficient, tell me if the relationship of the data is weak or strong:
3. Using your regression model, predict how long it would take the popsicle to melt if the temperature was 80ºF.
4. Using your regression model, predict how long it would take the popsicle to melt if the temperature was 60 ºF
5. Question 3 is an example of _____ because we are making a prediction **within** the boundaries of our data.
6. Question 4 is an example of _____ because we are making a prediction **outside** the boundaries of our data.

The table below shows the population of single-celled organisms growth in a Petri dish over a period of 16 hours. The number of organism at a given time is record in the table.

Time, hrs (x)	**Number of Organisms (y)**
0	25
2	36
4	52
6	68
8	85
10	124
12	260
16	350

1. Using technology, run a Linear Regression model. Write the value of the correlation coefficient (round to thousandths)
2. With your data still entered into the calculator, now run an EXPONENTIAL REGRESSION by choosing 0:ExpReg instead of 4:LinReg. Write the value of the correlation coefficient (round to thousandths)
3. Based off these values, which model better fits this data?
4. Write the equation that best fits this model:

Calculator Troubleshooting

You Deleted one of your lists: L1, L2, etc – Click Stat and instead of 1:Edit, choose 5:SetUpEditor then Enter. Once the calculator says Done, go back to Stat, 1:Edit and you will see them!
Calculator Error says DIM MISMATCH: This means your L1 and L2 lists are uneven. Make sure all the data was entered correctly.
You can't see the Line of Best Fit: A lot of times we forget to go to the y= button and type it in!
You can't see your scatterplot: Make sure that your Stat Plot 1 is turned on (2^{nd} and y= button), or that you selected Zoom, 9:ZoomStat