

Linear Regression and using lines to make predictions

The table shown lists the average global temperature for each decade from 1950 to 2009.



Decade Number	Decade	Average Temperature (°F)
0	1950-1959	56.878
1	1960-1969	56.642
2	1970-1979	56.732
3	1980-1989	56.822
4	1990-1999	57.038
5	2000-2009	57.238

a. Identify the independent and dependent quantities and their units of measure.

I: time (Decades)

D: temp (°F)

b. Why is the first decade numbered 0?

It represents the starting decade - which would be the y-intercept on the graph.

c. Does the data represent a linear function?

No, Avg. temp does not increase at a constant rate.

there are two decades where it decreases.

d. Does the data have an increasing or decreasing trend?

the data generally is increasing.

e. Is it possible to predict the average global temperature for 2070-2079 from the table? Graph?

the Avg. temp will most likely rise, but it is impossible to approx. what the avg. temp will be for this decade.



Making a Prediction using a "Trend Line"

Vocabulary:

Linear Regression: Models the relationship between two variables in a data set by producing a line of best fit.

Line of Best Fit ("Trend Line"): The line that best approximates the linear relationship between two variables in a data set.

Linear Regression Equation: The equation of the line of best fit.



1. Find and graph the Linear Regression Equation for the global average temperature on a calculator.

use calc instructions on last page

2. Re-write the Linear Regression Equation as a function. How should you round the slope and y-intercept? Explain your reasoning.

$$f(x) = 0.110x + 56.572$$

Round to 3 decimal places because that is how the temps are rounded in the table.

3. Does the data show a positive correlation or a negative correlation? How can you tell?

positive because the trend line has a positive slope.



4. Do you think the line fits the data well? Explain your reasoning.

most of the points are close to the trend line, so I would say the line fits the data reasonably well.



Vocabulary:

Correlation Coefficient: Indicates how closely the data points fit a straight line.
(The **r-value** on your calculator)

Positive Correlation: When r is a number between 0 and 1.
(The closer to 1 the more in line the data points are)

Negative Correlation: When r is a number between 0 and -1.
(The closer to -1 the more in line the data points are)

No Correlation: When r is 0. (Data points show no linear relationship)

1. What is the correlation coefficient, or r-value, for your line of best fit? Interpret its meaning.

$$r = 0.889$$

Since this value is close to 1, that means the data has a strong linear relationship.

2. About how much was the average global temperature changing each decade from 1880 to 2009 according to the data? Explain how you know.

about 0.11 degrees per decade.

this is the slope of the trend line.

3. Compare the y-intercept from the table with the y-intercept from the linear regression equation. What do you notice? Does this make sense in terms of the problem situation? Why or why not?

They are not the same.

yes, because the trend line is only an approximation.

4. Use your equation to predict the average global temperature for the years 2070-2079. Show your work and explain your reasoning.

2070-2079 is the 19th decade so:

$$\begin{aligned} f(19) &= 0.110(19) + 56.572 \\ &= 58.662 \text{ degrees} \end{aligned}$$

Check for Understanding

Use your previous notes to find the Line of Regression equation for the following data:

The table shows the Dow Jones average on the NYC Stock Exchange over a period of 6 consecutive months (August through January) in 1987. Black Monday refers to Monday, October 19, 1987, when stock markets around the world crashed. The Black Monday decline was the largest one-day percentage decline in stock market history to date.

Time (months)	Dow Jones Average (points)
1	2700
2	2600
3	1750
4	1950
5	1900
6	2000

1. Determine the slope (a), the y -intercept (b), the correlation coefficient (r), and the equation for the linear regression line for the month versus the Dow Jones Industrial average the month.

- a. The value of " a " is -154.29
b. The value of " b " is 2690
c. The value of " r " is -0.726
d. The Regression Equation is $y = -154.29x + 2690$
Round coefficients to the nearest hundredth where needed

2. What does a represent in this problem situation?

Avs. drop in (points) over 6 months.

3. What does b represent in this problem situation?

the Avs. points of the Dow just before August 1987.

4. How well does the regression line fit the data?

Since $r = -0.7$ it fits moderately well.

5. When should you not use the regression line to predict future data?

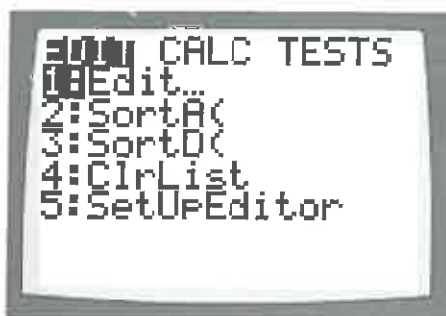
When the data does not show a linear trend.

Entering Stats Data into a Calculator:

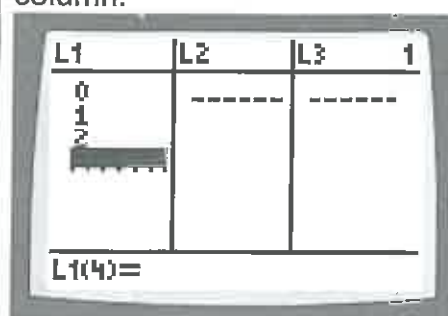
1. Press **STAT** button



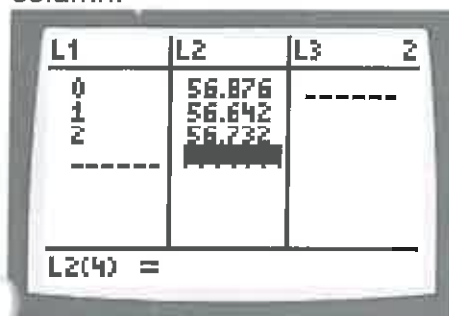
2. Press **ENTER** for choice 1.



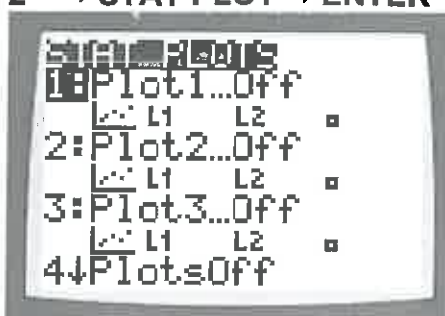
3. Enter independent data into L1 column.



4. Enter dependent data into L2 column.



5. Press:
2nd → **STAT PLOT** → **ENTER**



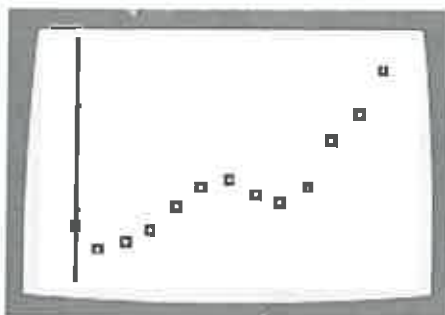
6. Set **PLOT1** options to these:



7. Press **ZOOM**
Arrow down to choice **9:ZOOM STAT**
Press **ENTER**



8. Press **TRACE**
use arrow keys to move between the data points



Finding & Graphing the Linear Regression Equation using a Calculator

1. Turn on Diagnostics:
Press 2^{nd}
Press **CATALOG** (on 0 key)
Scroll to **DiagnosticOn**



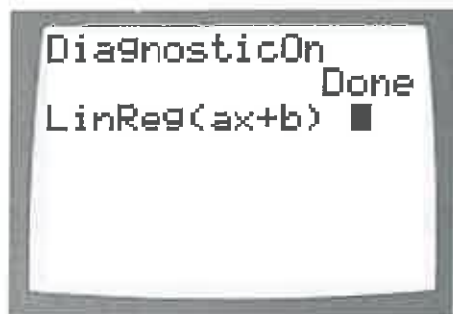
2. Press **ENTER** twice
(until you see **DONE**)



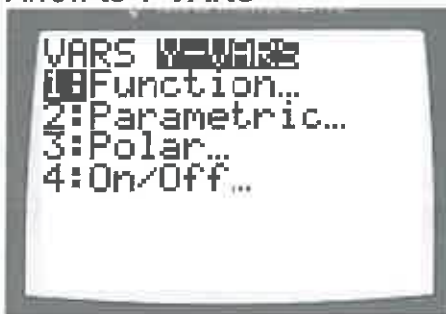
3. Press **STAT**
Arrow to **CALC**
Arrow down to **4:LinReg(ax+b)**



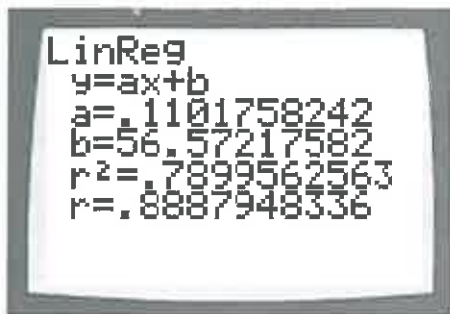
4. Press **ENTER**



5. Press **VARs**
Arrow to **Y-VARS**



6. Press **ENTER** (3 times)



7. Press **GRAPH** to see the line of best fit.

