Key

Algebra 1

☆ Turn or Diagnostics ☆

#10.

Date_____

Period_____

Line of Best Fit aka Linear Regression

When data is displayed with a scatter plot, it is often useful to attempt to represent that data with the equation of a straight line for purposes of predicting values that may not be displayed on the plot. Such a straight line is called the "line of best fit." It may also be called a "trend" line.

A line of best fit is a straight line that best represents the data on a scatter plot. This line may pass through some of the points, none of the points, or all of the points.

approximation
estimation

Predicting:

- If you are looking for values that fall within the plotted values when using the line of best fit, you are ***interpolating.***
- If you are looking for values that fall outside the plotted values when using the line of best fit, you are ***extrapolating.*** **Be careful** when extrapolating. The further away from the plotted values you go, the less reliable is your prediction.

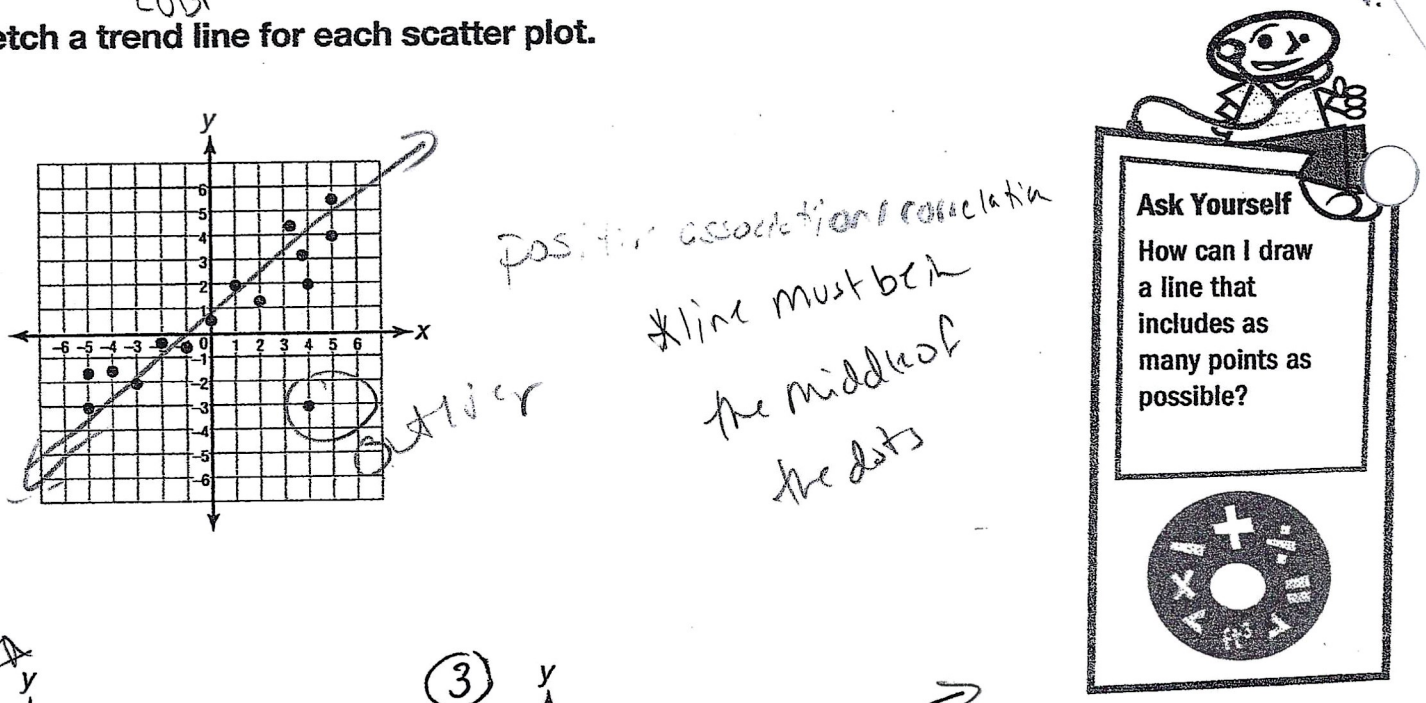
Another term for line of best fit is

Choose:

a) ~~scatter plot~~
b) trend line
c) ~~tangent line~~
d) slope

LOBF

Sketch a trend line for each scatter plot.



① positive association/correlation

*line must be in the middle of the dots

outlier

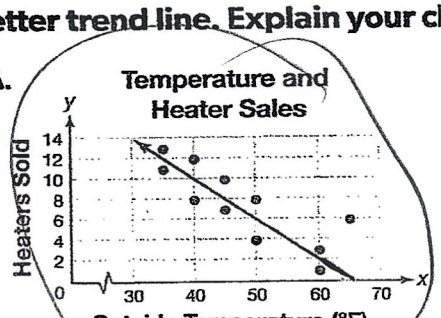
② positive correlation

③ positive correlation

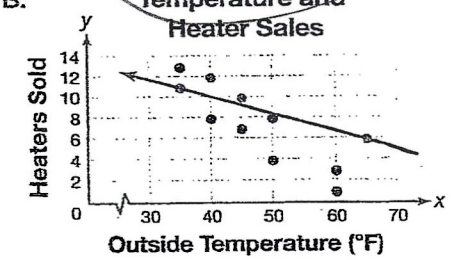
Cluster

Consider each pair of identical scatter plots. Circle the letter of the plot that shows the better trend line. Explain your choice.

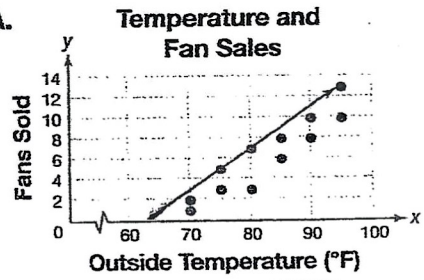
① A.



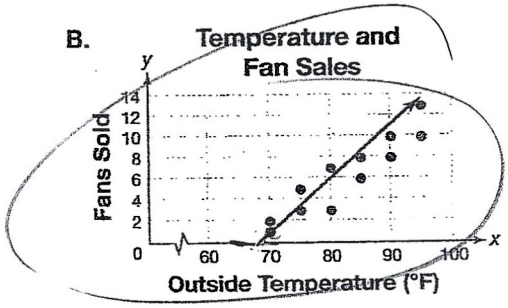
B.



② A.



B.



The line is in the middle of the dots

Correlation Coefficients

We know that the graphing calculator can find a "best fit" regression equation that can be used to predict new values. But, how reliable will these prediction be? Is there a way to determine how well our regression equation fits our data?

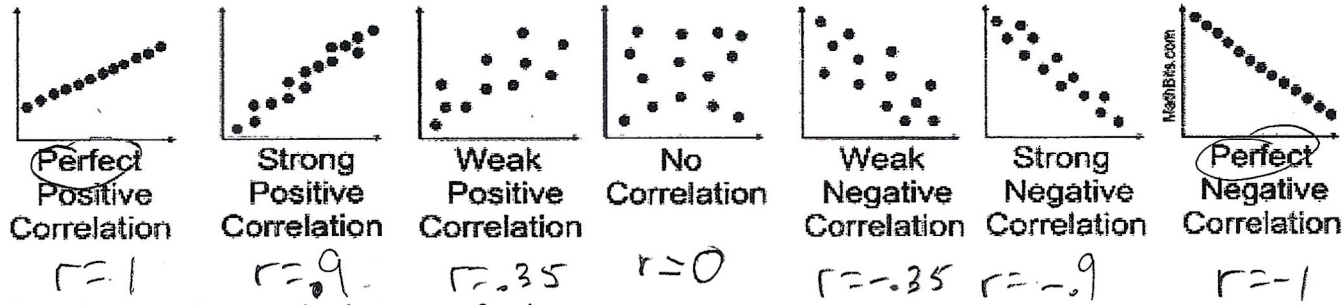
Yes! There is a way of measuring the "goodness of fit" of the best fit line (least squares line), called the correlation coefficient. It is a number between -1 and 1, inclusive, which indicates the measure of linear association between the two variables, and also shows whether the correlation is positive or negative.

Definition: A correlation coefficient, designated by r, is a number in the range $-1 \le r \le 1$, that indicates how well a regression equation truly represents data being examined.

If r is close to 1 (or -1), the model is considered a "good fit".

- If r is close to 0, the model is "not a good fit".
- If $r = \pm 1$, the model is a "perfect fit" with all data points lying on the line.
- If $r = 0$, there is no linear relationship between the two variables.

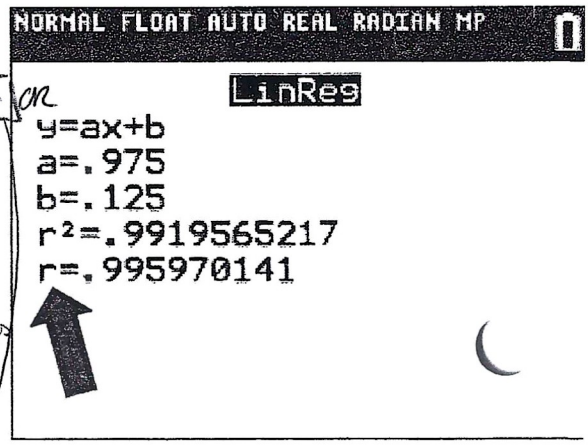
A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*.



Using the graphing calculator to find r

Be sure the TI-84+'s "Diagnostics" are turned on.
If not, you will not see the r-value.

When you choose a regression equation on the calculator, the correlation coefficient will be displayed on the screen with the regression equation information (assuming the Diagnostics are turned on).



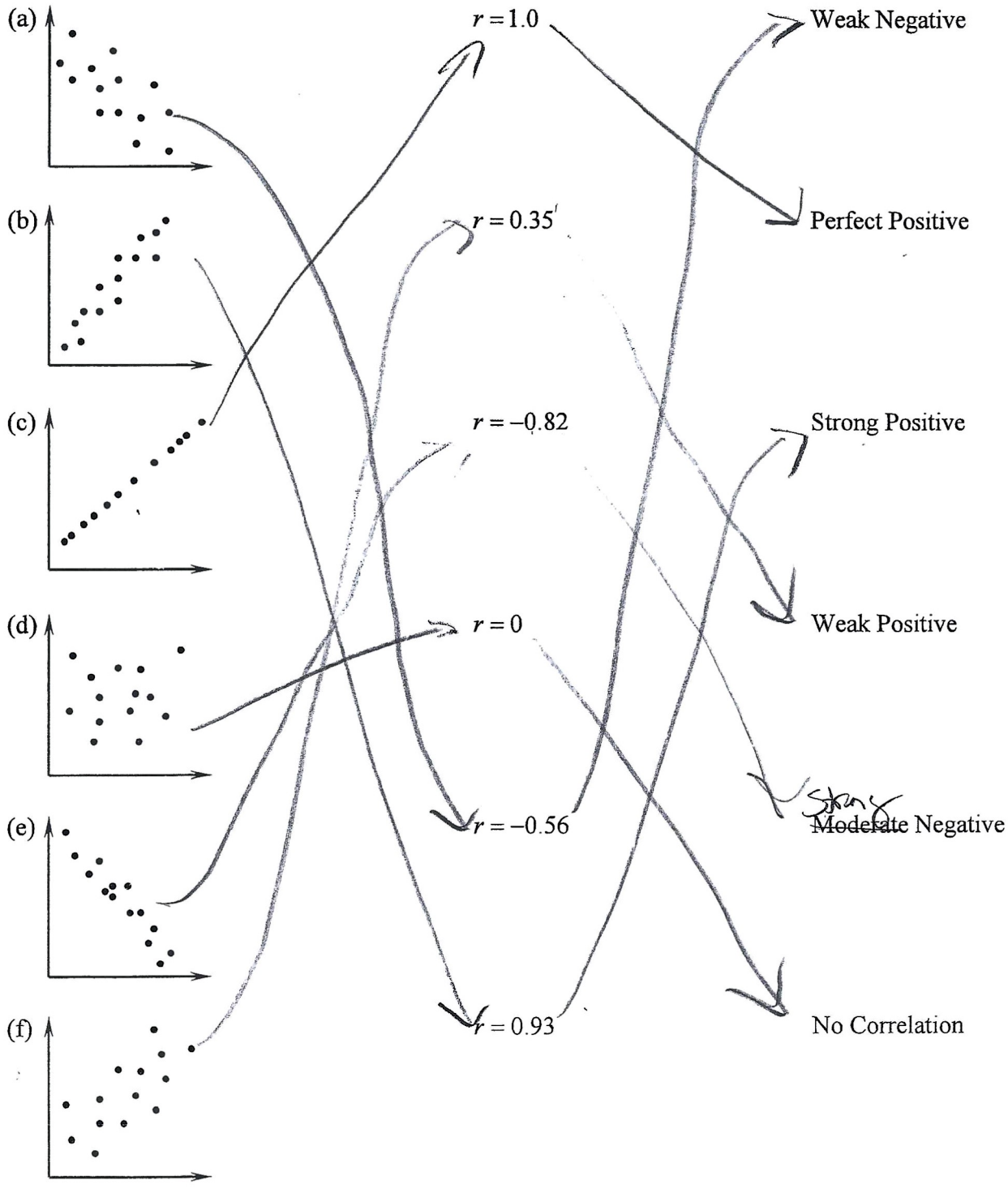
The linear regression screen shown at the right shows an "r" value of 0.995970141, which implies a strong correlation.

The linear regression equation, in this case, will be a reliable model for future forecasts or predictions.

Name: _____ Date: _____

QUANTIFYING PREDICTABILITY
COMMON CORE ALGEBRA I HOMEWORK

1. Below there are six scatter plots, six correlation coefficients, and six terms. Match the appropriate r-value with the scatter plot it most likely corresponds to. Then match the term you think is most appropriate to the r-value as well (not to the graph).



More Examples ★ By hand: USC the smallest x-value & the largest x-value 7786 -

1) Given the data in the chart below.

x	4	5	6	7
y	8	10	12	14

★ must put #'s in the calc in same order that you see

$y = mx + b$

$m = 2$ Determine a line of best fit. (equation)

$B = 0$ (4, 8) (7, 14)

$x_1\ y_1$ $x_2\ y_2$

$y = 2x$

$m = \frac{y_2 - y_1}{x_2 - x}$

$m = \frac{14 - 8}{7 - 4}$

$m = \frac{6}{3}$

$m = 2$

(4, 8) $m = 2$

xy

$y = mx + b$

$8 = (2)(4) + b$

$8 = 8 + b$

$-8 \quad -8$

$0 = b$

$r = 1$ ★ perfect pos. dots create line

2) The chart below shows the number of minutes studied and the grade received on a test.

L_1 Minutes Studied (x)	L_2 Test Grade (y)
15	50
40	67
45	75
60	75
70	73
75	89

★ must put #'s in the calc in the same order as you see

Determine a line of best fit for this data.

★ Round to the nearest thousandths

$y = .519x + 45.115$

$r = .903$ → strong pos. close to pos. straight line

(15, 50) (75, 89)

$x_1\ y_1$ $x_2\ y_2$

$m = \frac{y_2 - y_1}{x_2 - x_1}$

$m = \frac{89 - 50}{75 - 15}$

$m = \frac{39}{60}$

$m = .65$

(15, 50) $m = .65$

xy

$y = mx + b$

$50 = .65(15) + b$

$50 = 9.75 + b$

$-9.75 \quad -9.75$

$40.25 = b$

$y = mx + b$

$m = .65$

$b = 40.25$

$y = .65x + 40.25$

equations are similar but the calculator ... more ... short

3) James uses data that he collected in a science experiment to calculate a line of best fit. He determines the equation of the line to be $y = 7x + 2.25$.

Use this equation to calculate the value of y when $x = 6$.

A) 15.25
B) 39.75
C) 44.25
D) 42

$y = 7x + 2.25$

$y = 7(6) + 2.25$

$y = 42 + 2.25$

$y = 44.25$

CALC: 1) STAT 1: Edit put #'s in L_1, L_2

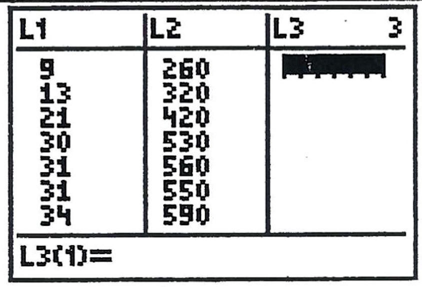
2) ZOOM (make sure stat plot is on) 9: STATS to graph

3) STAT → CALC 4: LinReg (ax+b) ↓ Calculate enter to get the equation

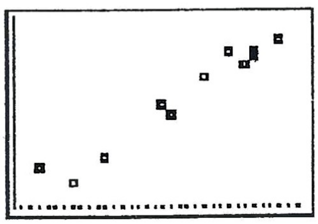
Line of best fit

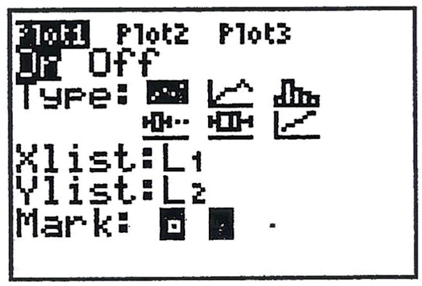
Can we predict the number of total calories based upon the total fat grams?

1. Enter the data in the calculator lists. Place the data in L_1 and L_2.
STAT, #1Edit, type values into the lists



2. Prepare a scatter plot of the data.
Set up for the scatterplot.
2nd StatPlot - choices shown at right.
Choose ZOOM #9 ZoomStat. Graph shown below.

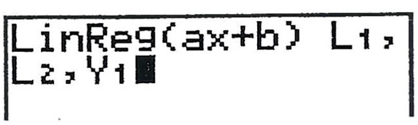




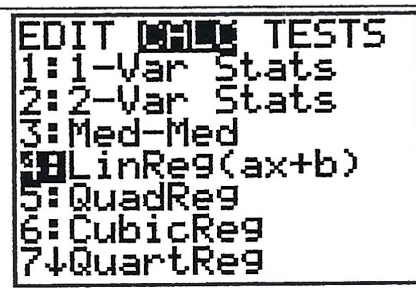
CALC Steps→

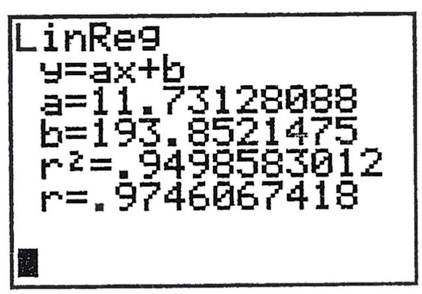
3. Have the calculator determine the line of best fit.
STAT → CALC #4 LinReg(ax+b)

Include the parameters L_1, L_2, Y_1.
(Y_1 comes from VARS → YVARS, #Function, Y_1)

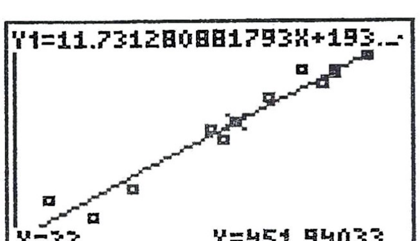


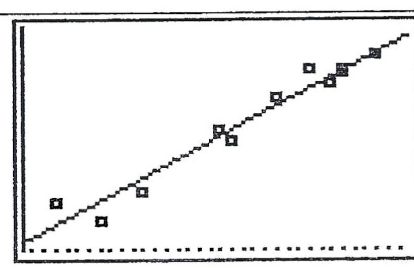
You now have the values of a and b needed to write the equation of the line of best fit. See values at the right.
$y = 11.73128088x + 193.8521475$





4. Graph the line of best fit.
Simply hit GRAPH. To get a predicted value **within the window**, hit TRACE, up arrow, and type the desired value.





Question: Predict the total calories based upon 22 grams of fat.
ANS: 451.940 calories

Y= → then VARS/STATS → EQ 1: RegEQ ZOOM 9: ZoomStat

34 The data given in the table below show some of the results of a study comparing the height of a certain breed of dog, based upon its mass.

*must put the #'s in the calc in the order they are in in the table

x L_1	**Mass** (kg)	4.5	5	4	3.5	5.5	5	5	4	4	6	3.5	5.5
y L_2	**Height** (cm)	41	40	35	38	43	44	37	39	42	44	31	30

LOBF

Write the linear regression equation for these data, where x is the mass and y is the height. Round all values to the *nearest tenth*.

STAT 1 put #'s into L_1 + L_2. then go to STAT CALC 4 to get the equation

$y = 1.9x + 29.8$

State the value of the correlation coefficient to the *nearest tenth*, and explain what it indicates.

MODE ↓ to Stat diagnostics ON to get the "r"

$r = 0.3$

The correlation coefficient suggests a weak positive linear relationship between the mass and height of a certain breed of dog Bc the correlation coefficient is not close to +1.

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

must put the #'s in the calc in the order they are in the table

L_1 x	**Mathematics**	55	93	89	60	90	45	64	76	89
L_2 y	**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.

$y = .81x + 15.19$

$r = 0.92$

move ◀ to stat diagnostics on to get the r"

STAT ◀ put #'s in L_1 & L_2 then go to STAT CALC (4) to get the equation

Explain what the correlation coefficient means with regard to the context of this situation.

The correlation coefficient suggests a strong positive linear relationship between the math scores and physics scores of nine students because the correlation coefficient is close to +1.

35 Stephen collected data from a travel website. The data included a hotel's distance from Times Square in Manhattan and the cost of a room for one weekend night in August. A table containing these data appears below.

*must put the #s in the calc in the order they are in the tab

L_1	**Distance From Times Square** (city blocks) (x)	0	0	1	1	3	4	7	11	14	19
L_2	**Cost of a Room** (dollars) (y)	293	263	244	224	185	170	219	153	136	111

Write the linear regression equation (LOBF) for this data set. Round all values to the *nearest hundredth*.

$y = -7.76x + 246.34$

STAT 1 put #s in L_1 & L_2 + then go to STAT CALC (4) to get the equation

State the correlation coefficient (r) for this data set, to the *nearest hundredth*.

$r = -0.88$

mode √ to stat diagnostics on to get the r

Explain what the sign of the correlation coefficient suggests in the context of the problem.

you can right this → BUT you MUST have → Blc it's asking about the sign

The correlation coefficient suggests a strong negative linear relationship between the distance from Times Square and the cost of a room Blc the correlation coefficient is close to -1.

*The negative sign suggests a negative correlation and means as the distance from times square increases, the cost of a room decreases.

↑ MUST have Blc it asks about the sign

36 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.

*must put the #'s in the calc in the order they are in the table

Percentage of Students Scoring 85 or Better	
Mathematics, x L_1	English, y L_2
27	46
12	28
13	45
10	34
30	56
45	67
20	42

Write the ~~linear~~ LOBF regression equation for these data, rounding all values to the *nearest hundredth*.

$y = 0.96x + 23.95$

STAT ▷ put #'s in L_1 & L_2 ↑ then go to STAT CALC 4 to get the equation.

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.

$r = 0.92$

mode ↓ to Stat diagnostics ON to get the "r"

The correlation coefficient suggests a strong positive linear relationship between the percentage of students scoring 85 or better on a math final exam and an English final exam because the correlation coefficient is close to +1.