

### Line of Best Fit

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.

**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

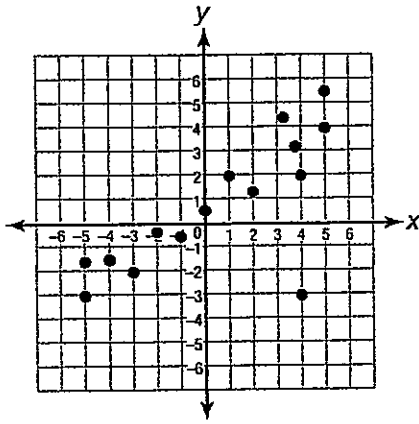
**Choose:**

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



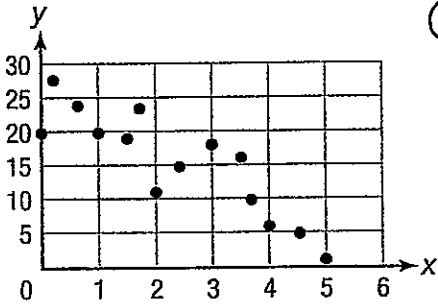
Sketch a trend line for each scatter plot.

①

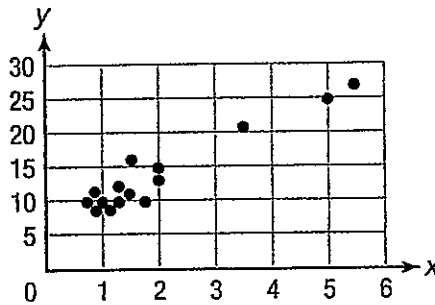


**Ask Yourself**  
How can I draw a line that includes as many points as possible?

②

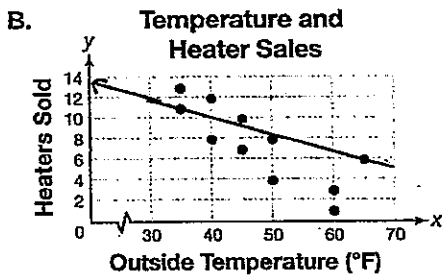
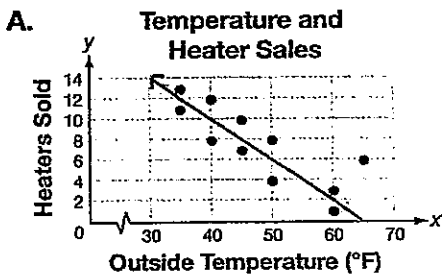


③

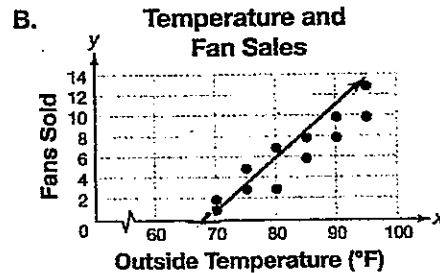
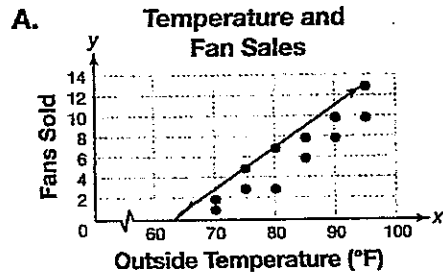


Consider each pair of identical scatter plots. Circle the letter of the plot that shows the better trend line. Explain your choice.

①



②




---



---



---



---

## More Examples

7786 -

- 1) Given the data in the chart below.

x	4	5	6	7
y	8	10	12	14

Determine a line of best fit.

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

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

Determine a line of best fit for this data.

- 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                      C) 44.25  
B) 39.75                      D) 42

Calc: 1) **STAT** **1:Edit** put #'s in  $y_1$  &  $y_2$   
2) **ZOOM** **9:STATS** to see the graph  
(make sure stat plot is on)  
3) **STAT** **→** **CALC** **4:LinReg (ax+b)**  
**↓** Calculate **enter** to get the equation

# Line of best fit

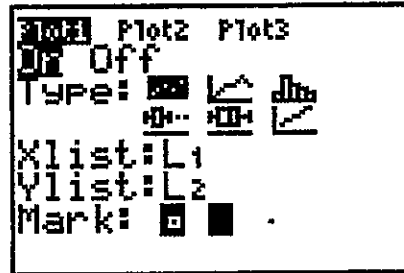
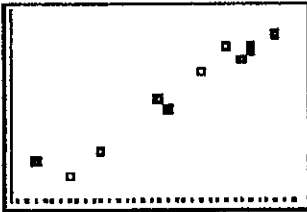
Can we predict the number of total calories based upon the total fat grams?

1. Enter the data in the calculator lists. Place the data in L<sub>1</sub> and L<sub>2</sub>.  
STAT, #1Edit, type values into the lists

L1	L2	L3	3
8	260		
13	320		
21	420		
30	530		
41	660		
51	800		
64	990		

L3(1)=

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



3. Have the calculator determine the line of best-fit.

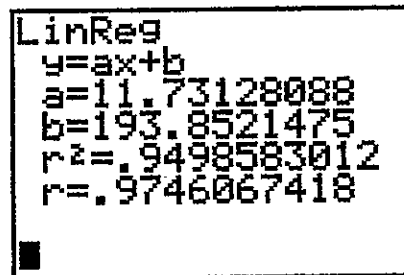
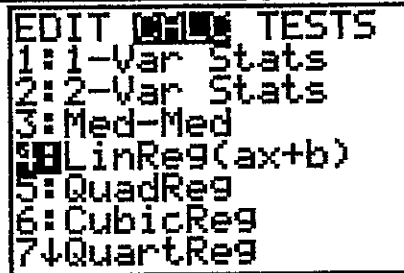
STAT → CALC #4 LinReg(ax+b)

Include the parameters L<sub>1</sub>, L<sub>2</sub>, Y<sub>1</sub>.  
(Y<sub>1</sub> comes from VARS → YVARS, #Function, Y<sub>1</sub>)

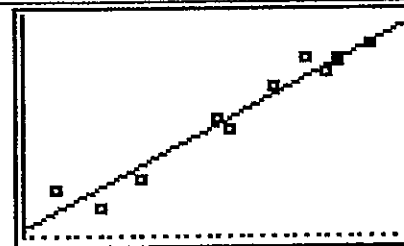
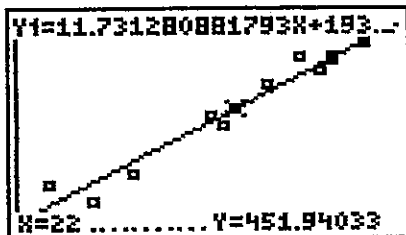
LinReg(ax+b) L1,  
L2, Y1

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

make sure you make off y =

Y = → VARS (STATS) → EQ [1: Reg EQ] (ZOOM) (9: ZOOMSTAT) → type equation on the graph

1. A factory is producing and stockpiling metal sheets to be shipped to an automobile manufacturing plant. The factory ships only when there is a minimum of 2,050 sheets in stock. The accompanying table shows the day,  $x$ , and the number of sheets in stock,  $f(x)$

- (a) Using the calculator, determine the line of best fit for this set of data, rounding the coefficients to *four decimal places*.

Day ( $x$ )	Sheets in Stock ( $f(x)$ )
1	860
2	930
3	1000
4	1150
5	1200
6	1360

- (b) Use this equation to determine the day the sheets will be shipped.

2. A real estate agent plans to compare the price of a cottage,  $y$ , in a town on the seashore to the number of blocks,  $x$ , the cottage is from the beach. The accompanying table shows a random sample of sales and location data.

- (a) Using the calculator, determine the line of best fit that relates the price of a cottage to its distance from the beach.

Number of Blocks from the Beach ( $x$ )	Price of a Cottage ( $y$ )
5	\$132,000
0	\$310,000
4	\$204,000
2	\$238,000
1	\$275,000
7	\$60,800

- (b) Use the equation to predict the price of a cottage, to the *nearest dollar*, located three blocks from the beach.

## 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.

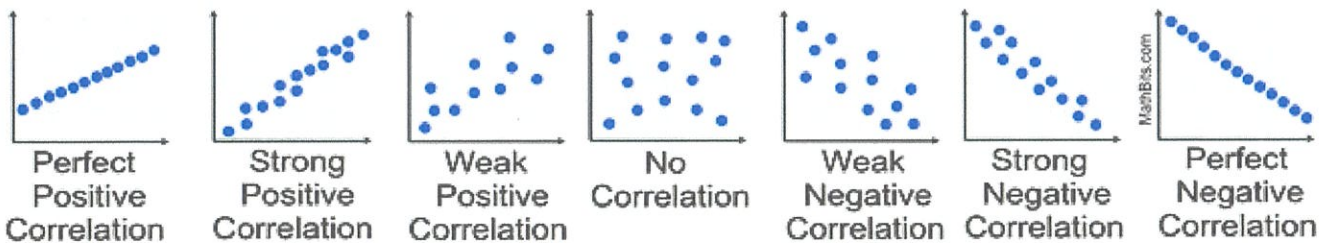
### Definition:

A **correlation coefficient**, designated by  $r$ , is a number in the range  $-1 \leq r \leq 1$ , 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.

Calc:  $2^{nd}$   $0$   $\downarrow$  Diagnostic on  $\left[ \text{Enter} \right]$   $\left[ \text{Enter} \right]$  OR  
 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.



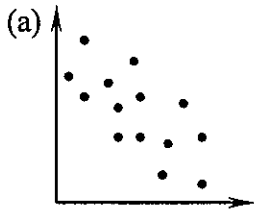
press  $\left[ \text{mode} \right]$  + scroll down until you see STAT Diagnostics +  $\left[ \rightarrow \right]$   $\left[ \text{on} \right]$  + press  $\left[ \text{enter} \right]$  then exi

Name: \_\_\_\_\_

Date: \_\_\_\_\_

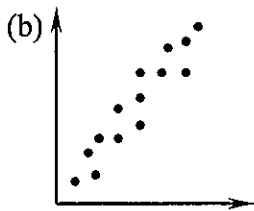
**QUANTIFYING PREDICTABILITY**  
**COMMON CORE ALGEBRA I HOMEWORK**

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).



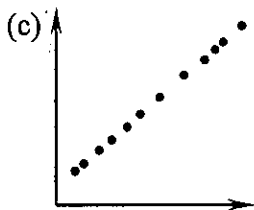
$r = 1.0$

Weak Negative



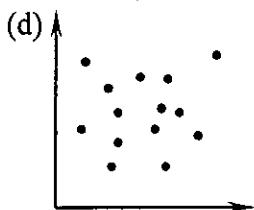
$r = 0.35$

Perfect Positive



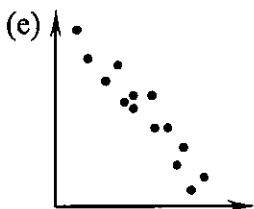
$r = -0.82$

Strong Positive



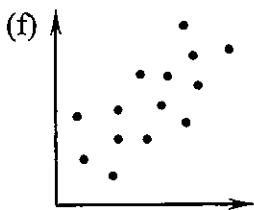
$r = 0$

Weak Positive



$r = -0.56$

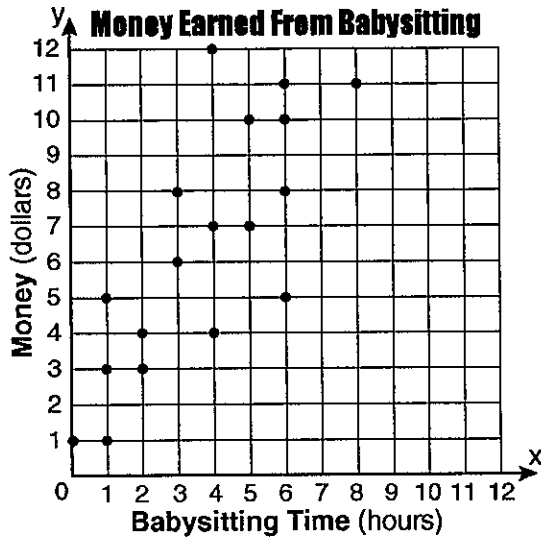
Moderate Negative



$r = 0.93$

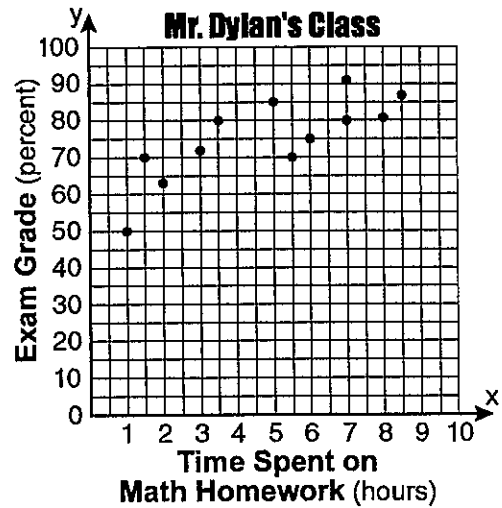
No Correlation

- 1) Which equation most closely represents the line of best fit for the scatter plot below?



- A)  $y = \frac{3}{2}x + 4$       C)  $y = \frac{3}{2}x + 1$   
 B)  $y = x$       D)  $y = \frac{2}{3}x + 1$

- 2) The number of hours spent on math homework each week and the final exam grades for twelve students in Mr. Dylan's algebra class are plotted below.

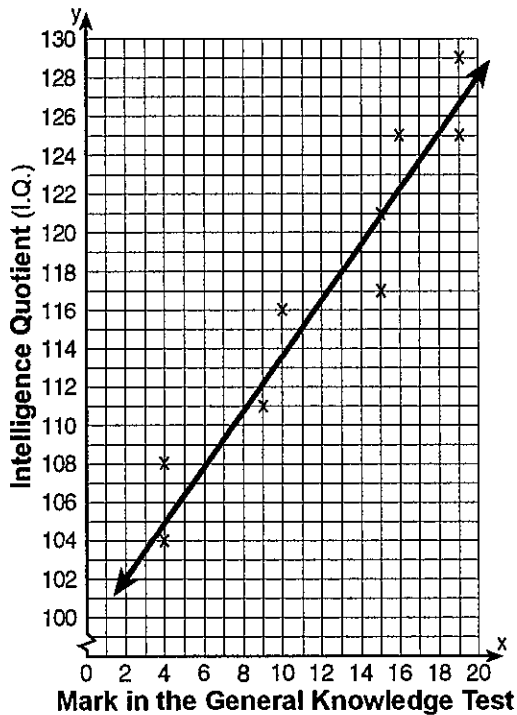


Based on a line of best fit, what exam grade would be the *best* prediction for a student who spends about 4 hours on math homework each week?

- A) 92      C) 72  
 B) 82      D) 62

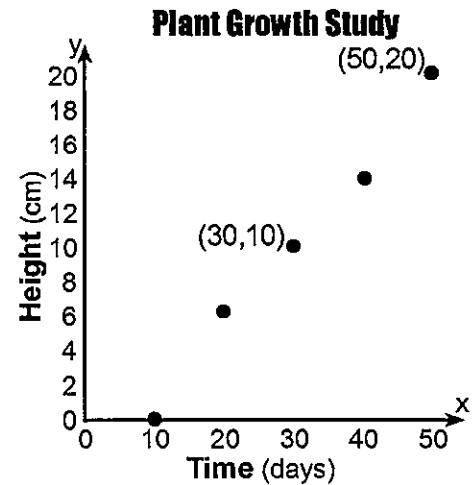


- 3) A published report shows how the United States student population scored on the general knowledge portion of the "Wechsler Adult Intelligence Scale" IQ test. In the scatter plot below, the line of best fit has been drawn.



A student has an IQ of 118. Use the scatter diagram shown to estimate this student's mark in the general knowledge test. Does 118 represent interpolating or extrapolating data?

- 4) The graph below shows the height of a plant over a period of days.



- (a) Using the points identified on the scatter plot above, determine a line of best fit.
- (b) Use your line of best fit from part (a) to predict the height of the plant after 70 days.
- (c) Use your line of best fit from part (a) to predict how many days it will take for the plant being observed to reach a height of 12 cm.