#107

Name________________ Date____________

8A; Algebra 1 Period______

## 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.

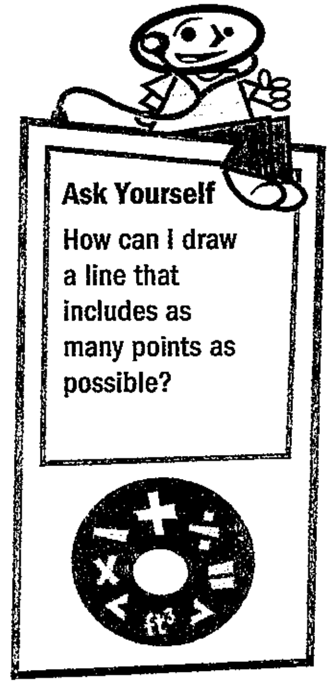
> **Predicting:**
> - If you are looking for values that fall within the plotted values when using the line of best fit, you are ***interpolating.***
> - If you are looking for values that fall outside the plotted values when using the line of best fit, you are ***extrapolating.*** **Be careful** when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

Another term for line of best fit is

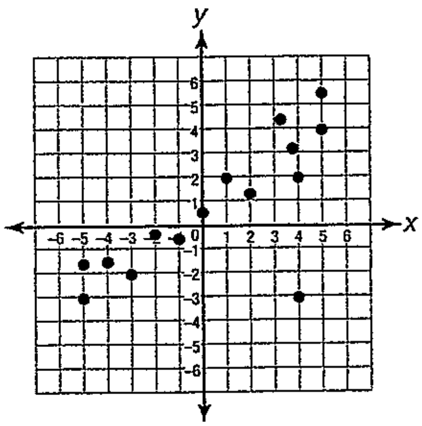
## Choose:

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

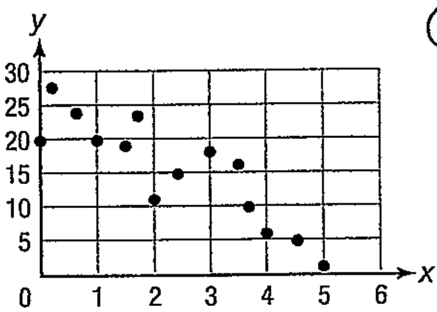
**Sketch a trend line for each scatter plot.**



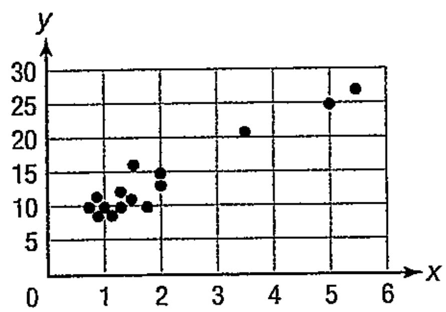

(1)




(2)

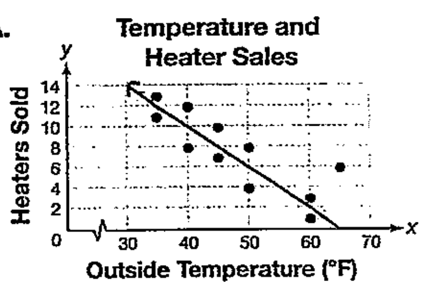



(3)

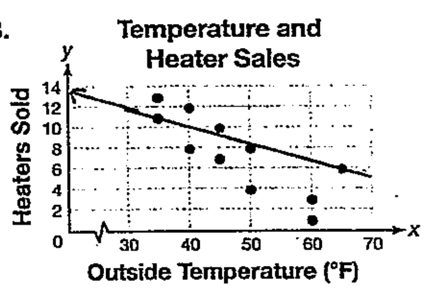



**Consider each pair of identical scatter plots. Circle the letter of the plot that shows the better trend line. Explain your choice.**

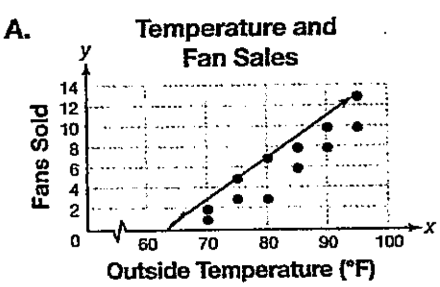
(1) **A.**



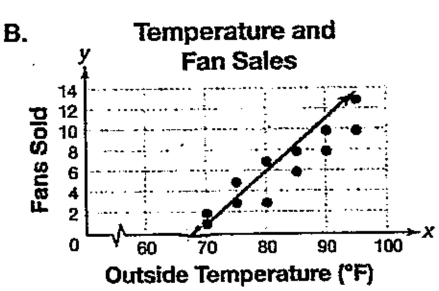

**B.**




(2) **A.**




**B.**




______________________________ ______________________________

______________________________ ______________________________

## 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?

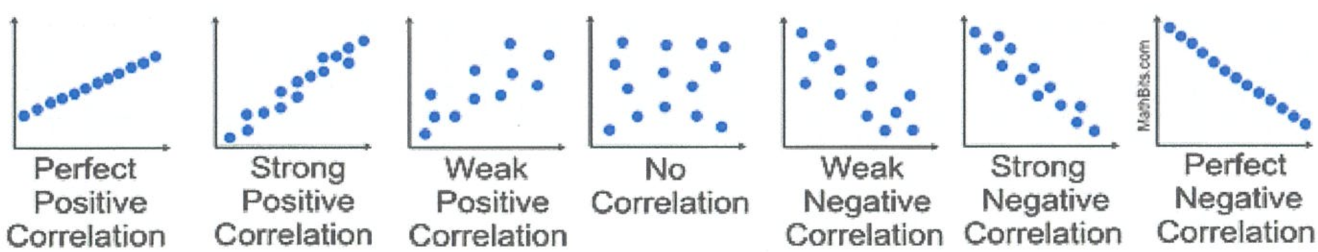
Yes! There is a way of measuring the "goodness of fit" of the best fit line (least squares line), called the correlation coefficient. It is a number between -1 and 1, inclusive, which indicates the measure of linear association between the two variables, and also shows whether the correlation is positive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

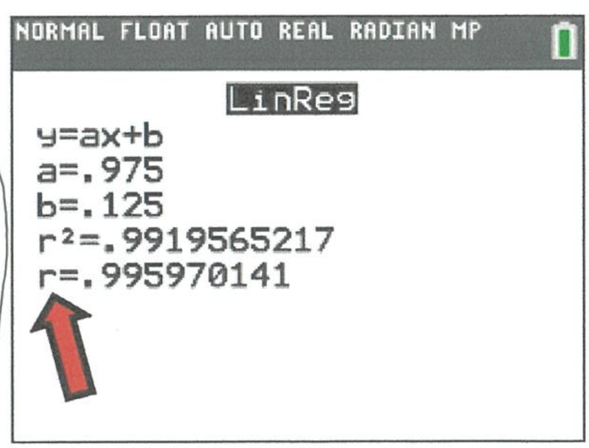
Be sure the TI-84+'s "Diagnostics" are turned on.
If not, you will not see the $r$-value.

Calc: [2nd] [0] [↓] Diagnostic on [Enter] [Enter] 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 [Mode] + scroll down until you see STAT Diagnostics + [→] [On] + press [Enter] then exit

Name: ______________________________ Date: ______________

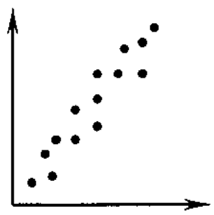
## QUANTIFYING PREDICTABILITY
## COMMON CORE ALGEBRA I HOMEWORK

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).

(a) 

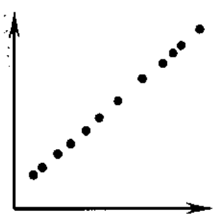
$r = 1.0$

Weak Negative

(b) 

$r = 0.35$

Perfect Positive

(c) 

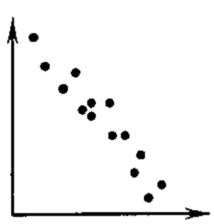
$r = -0.82$

Strong Positive

(d) 

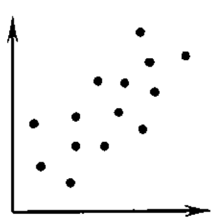
$r = 0$

Weak Positive

(e) 

$r = -0.56$

Moderate Negative

(f) 

$r = 0.93$

No Correlation

# More Examples



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 $x = 6$.

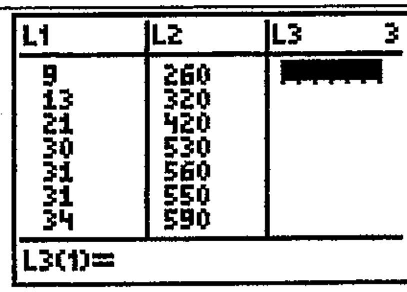
A) 15.25　　C) 44.25
B) 39.75　　D) 42

Calc: 1) [STAT] [1: Edit] put #'s in L1 + L2
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

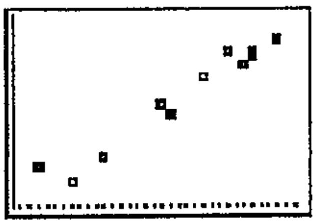
# Line of best fit

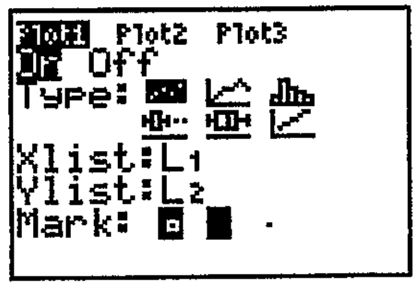
**Can we predict the number of total calories based upon the total fat grams?**

1. Enter the data in the calculator lists. Place the data in $L_1$ 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.

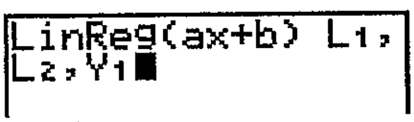


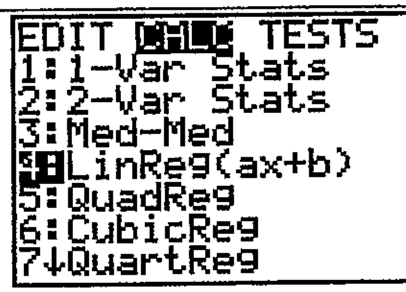


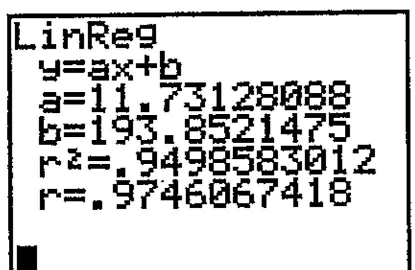

3. Have the calculator determine the line of best-fit.
STAT → CALC #4 LinReg(ax+b)

calc stat

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

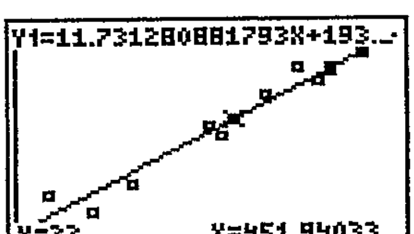


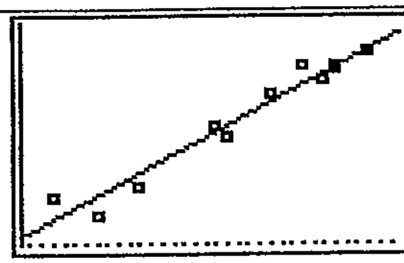




You now have the values of a and b needed to write the equation of the line of best fit. See values at the right.
$y = 11.73128088x + 193.8521475$

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.




make sure you turn off y=



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

y= → VARS STATS → EQ 1:RegEQ ZOOM 9: ZOOMSTAT → to put equation on the graph

**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.

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

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*.

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

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

| **Mathematics** | 55 | 93 | 89 | 60 | 90 | 45 | 64 | 76 | 89 |
|---|---|---|---|---|---|---|---|---|---|
| **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.

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

**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.

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

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

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

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

**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.

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

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

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.