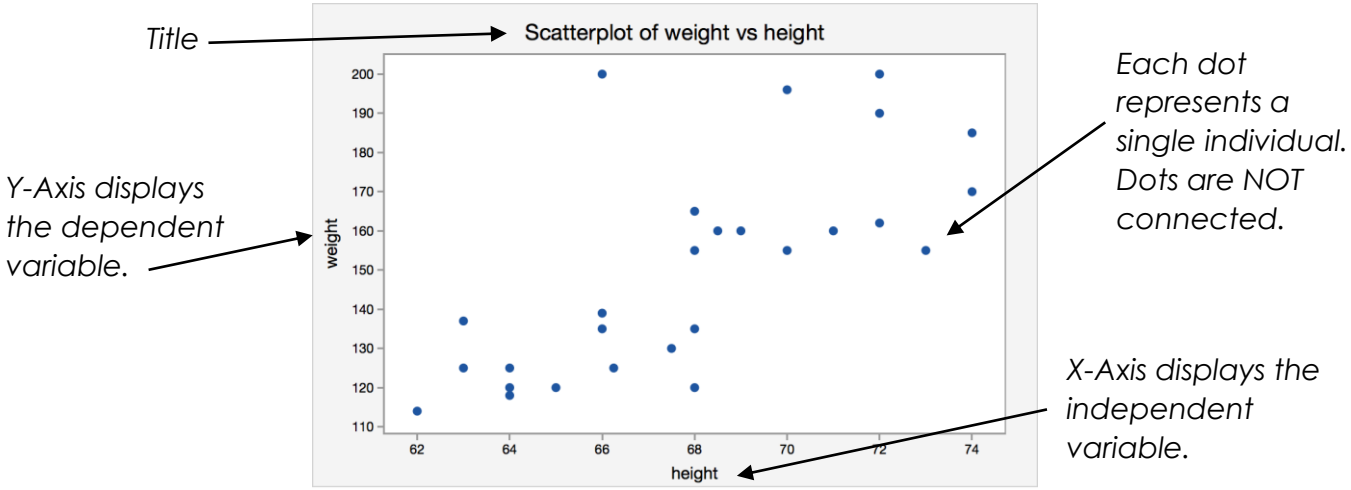
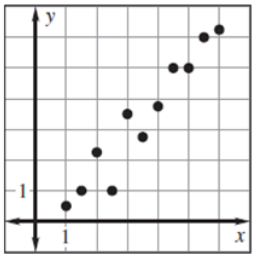
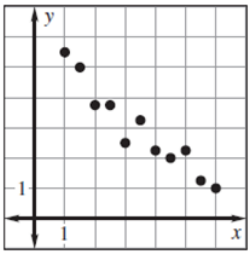
Scatterplots and Correlation

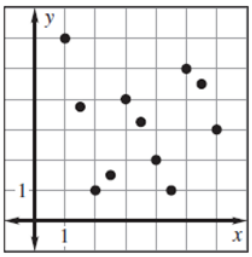
A **scatterplot** is a graph of data pairs (x, y). An example of a scatterplot is below.



Scatterplots are typically used to describe relationships, called **correlations,** between two variables (bi-variate). The **correlation coefficient** describes how well a line fits the data. A **trend line** can be drawn to help determine correlation.





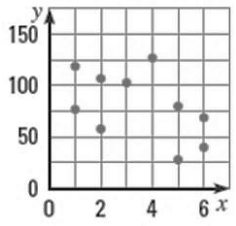


Positive Correlation	Negative Correlation	No Correlation
▫ as x increases, y increases ▫ positive slope ▫ Correlation Coefficient is close to 1	▫ as x increases, y decreases ▫ negative slope ▫ Correlation Coefficient is close to -1	▫ no relationship between x and y ▫ Correlation Coefficient is close to 0

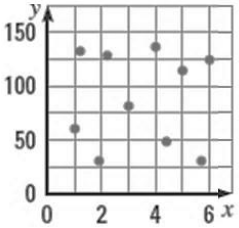
Correlation Coefficient Guidelines			
0.70 to 1.00	Strong Positive	-0.70 to -1.00	Strong Negative
0.30 to 0.69	Moderate Positive	-0.30 to -0.69	Moderate Negative
0.00 to 0.29	None to Weak Positive	0.00 to -0.29	None to Weak Negative

Example: Describe the correlation of each scatterplot below. Be sure to include direction and strength in your description.

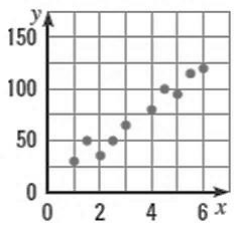
a)



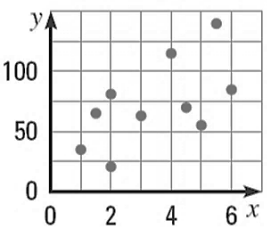
b)



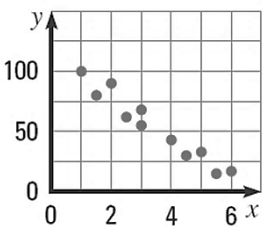
c)



d)



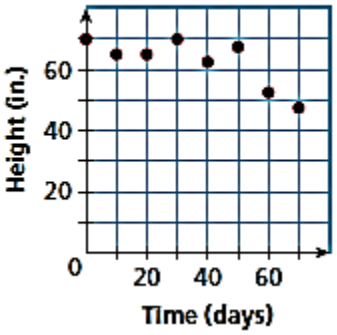
e)



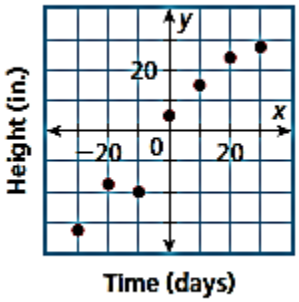
Example 1: Describe the scatterplot that best describes the scenario below and explain why you chose that graph.

Scenario: *The relationship between the number of days since a sunflower seed was planted and the height of the plant.*

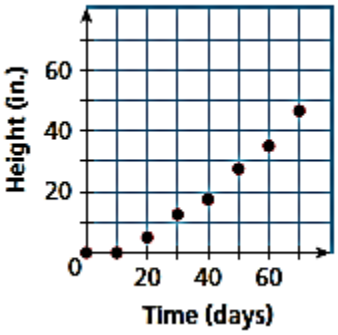
Graph A



Graph B



Graph C



I think graph ____ best represents this scenario because:

_____

_____

Example 2: Describe the correlation you would expect to see between each pair of data sets.

Positive Correlation: When the variables go in the _____ direction. (↑,↑ or ↓,↓)

Negative Correlation: When the variables go in _____ directions. (↓,↑ or ↓,↓)

a. The number of hours you work vs the amount of money in your bank account:

positive correlation *negative correlation* *no correlation*

b. The number of hours workers receive safety training vs the number of accidents on the job:

positive correlation *negative correlation* *no correlation*

c. The number of students at Harrison vs the number of dogs in Kennesaw:

positive correlation *negative correlation* *no correlation*

d. The number of meals eaten vs the number of cars on I-75 throughout the day:

positive correlation *negative correlation* *no correlation*

e. The number of calories burned/lost vs the amount of hours you worked out:

positive correlation *negative correlation* *no correlation*

Linear Regression

When given a scatterplot of data, we can use our calculators to run a linear regression model. This linear regression model (also called the least-squares regression), would give us the BEST line of best for the data.

Calculating Linear Regression (TI-83/TI-84)

1) Press 2nd, 0, x^{-1}, and scroll down until you see Diagnostic**On**. Press ENTER twice – your screen should then say "Done"

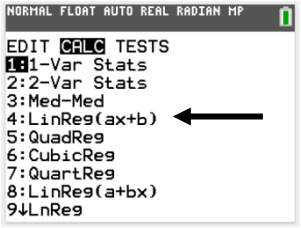
2) Press STAT, then ENTER.

3) To clear L_1, press ↑ to the go to the very top, CLEAR, ↓. Then press → to go to L_2. To clear L_2, press press ↑ to the go to the very top, CLEAR, ↓. ★**NEVER PRESS DELETE**★

4) Type your data for the x-variable into L_1. Be sure to hit ENTER after each number. You do NOT need to order your data first.

5) Press → to go to L_2. Type your data for the y-variable into L_2. Be sure to hit ENTER after each number. You do NOT need to order your data first.

4) Once all of your data is entered, press STAT, then → and select option 4:LinReg(ax+b). You can do this by either scrolling down to 4 and pressing ENTER or just pressing the 4 on the keyboard.

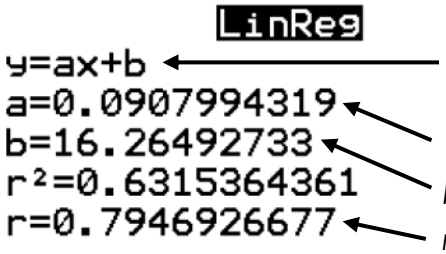


5) Your screen will then look like one of the following:

Make sure it says Xlist: L_1 and Ylist:L_2. FreqList and Store RegEQ are blank. Press ENTER until screen changes.

Press ENTER. Your calculator automatically assigns L_1 to Xlist and L_2 to Ylist.

6) Your screen will then change to this:



This is the form that you will write your equation. (It is slope intercept form just with different letters)

a represents the slope

b represents the y-intercept

r represents the correlation coefficient

★ *Be sure to check the signs of a, b, and r. Round to the nearest hundredth for a and b and the nearest thousandth for r unless told otherwise.* ★

Linear Regression Practice

1) Using your calculator, find the linear regression model for the data below.

Taylor had guests for dinner at her house eight times and has recorded the number of guests and the total costs for each meal in the table below:

Guests	3	4	4	6	6	7	8	8
Cost ($)	30	65	88	90	115	160	150	162

a = _____ b = _____ r = _____

Linear Regression Model: _____

Correlation: _____

Now use your linear regression model and information above to answer the following questions.

a) Predict the cost for 2 guests to have a meal.

b) If a meal costs $89.78, how many guests were there?

c) Interpret what the slope means in context.

When interpreting the slope, follow this general guide:
As (x context) increases by 1 (x-unit), the number of (y context) increases/decreases by (slope # with y-units).

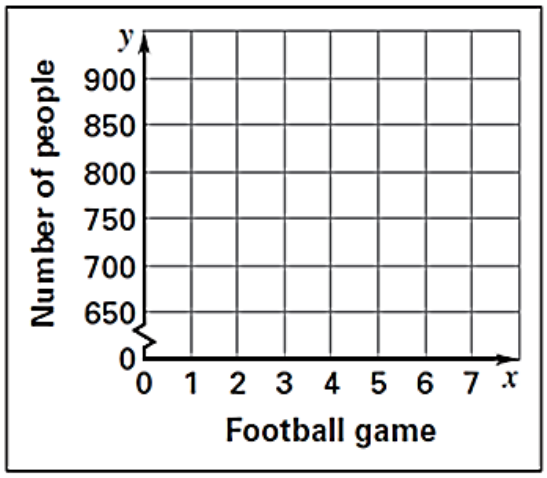
d) Interpret what the y-intercept means in context.

When interpreting the y-intercept, follow this general guide:
For 0 (x context), the number of (y context) is (y-intercept #).

2) The table below gives the number of people, y, who attended each of the first seven football games, x, of the season.

x	1	2	3	4	5	6	7
y	722	763	772	826	815	857	897

a) Create a scatterplot of the data.



b) Calculate the linear regression equation. Round your decimals to the nearest tenth.

a = _____ b = _____

Equation: _____

c) Identify the correlation coefficient. What does that tell you about the relationship between the two variables?

d) Interpret the slope. Does it make sense?

e) Interpret the y-intercept. Does it make sense?

f) Using your regression equation, predict the number of people at the last game if there are 9 home games this season.

3) Here is data from a group of students who measured the size of their hand in inches and counted the number of starbursts they could grab at one time.

Size of hand	6	6	7	7	7.5	8	8	9	9	9.5
# of Starbursts	30	26	31	30	31	39	29	40	43	50

a) Calculate the linear regression equation. Round all decimals to the nearest tenth.

b) Interpret the slope. Does it make sense?

c) Interpret the y-intercept. Does it make sense?

4) The table below gives the amount of time students in a class studied for a test and their test scores. Create a scatterplot. Then calculate the linear regression model.

Hours Studied	1	0	3	1.5	2.75	1	0.5	2
Test Score	78	75	90	89	97	85	81	80

a) Linear Regression Equation:

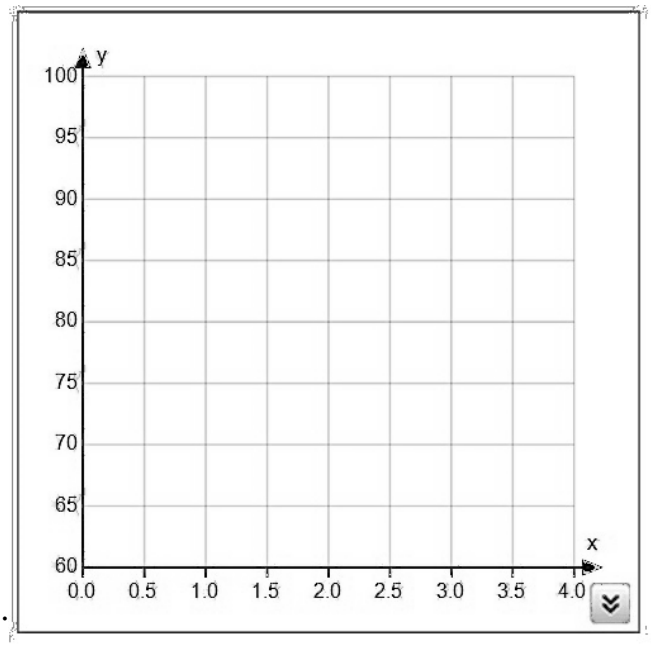
_____

b) Correlation Coefficient (r): _____

c) Type of Correlation: _____

d) Using the linear regression equation predict a students test score if they studied for 4 hours.

e) Explain what the y-intercept means in context.



f) Explain what the slope means in context.

5) The table below gives the estimated world population (in billions) for various years.

Year	1980	1990	1997	2000	2005	2011
Population	4400	5100	5852	6080	6450	7000

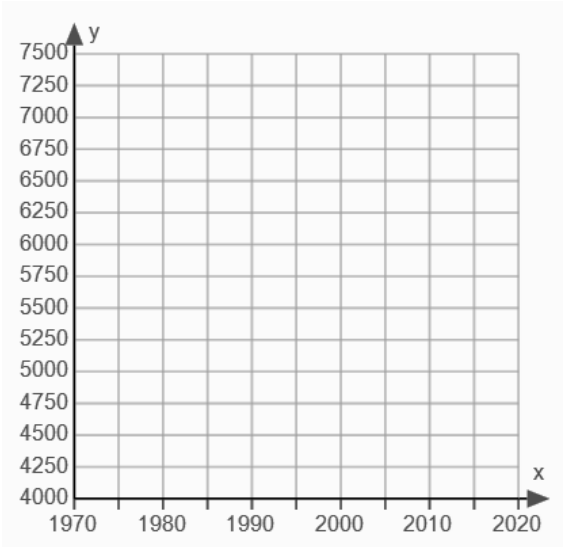
a) Linear Regression Equation:

_____

b) Correlation Coefficient (r): _____

c) Type of Correlation: _____

d) Using the linear regression equation predict the world population in the year 2030.



Extra Practice

Decide whether each scenario has a positive, negative, or no correlation. Remember to think **generally**!

1. Hours studying vs. Grades
   positive correlation *negative correlation* *no correlation*

2. Hours shopping vs. Amount of money spent
   positive correlation *negative correlation* *no correlation*

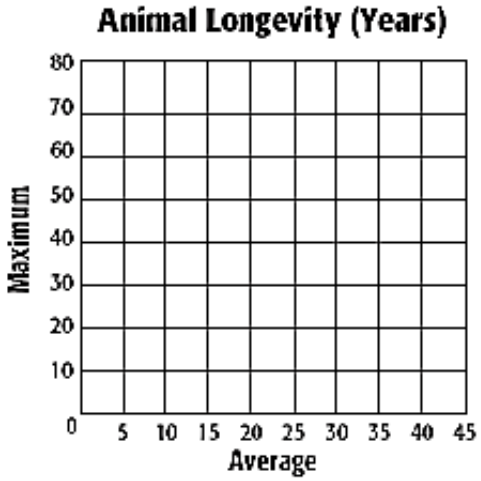
3. Person's height vs. Person's age
   positive correlation *negative correlation* *no correlation*

4. Person's height vs. How fast they drive
   positive correlation *negative correlation* *no correlation*

5. Temperature vs. Number of people wearing jackets
   positive correlation *negative correlation* *no correlation*

2. The table shows the average and maximum longevity of various animals in captivity.

Longevity (years)								
Avg.	12	25	15	8	35	40	41	20
Max.	47	50	40	20	70	77	61	54



a) Draw a scatterplot and determine, what relationship, if any, exists in the data.

b) Calculate the linear regression equation.

c) Use your equation to predict the maximum life span of an animal that has an average life span of 17 years. *Round to the nearest tenth if needed.*

d) If an animal has a maximum life span of 30 years, what is the average life span of the animal? *Round to the nearest tenth if needed.*