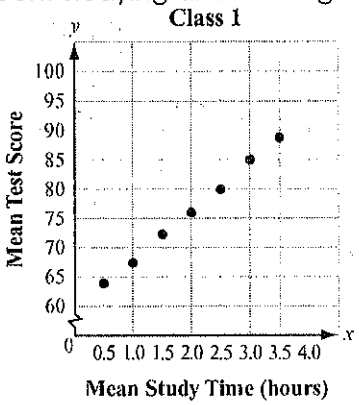
Name: _____ Date: _____

Correlation

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.9 Distinguish between correlation and causation.

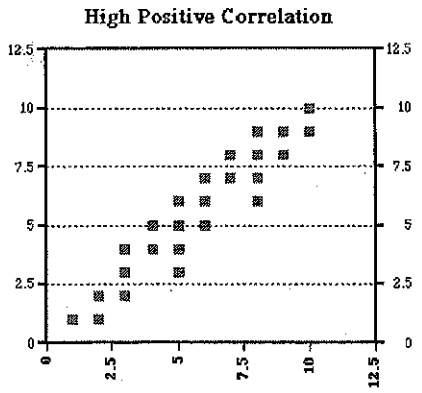
A **scatter plot** is often used to present bivariate **quantitative** data. Each variable is represented on an axis and the axes are labeled accordingly.

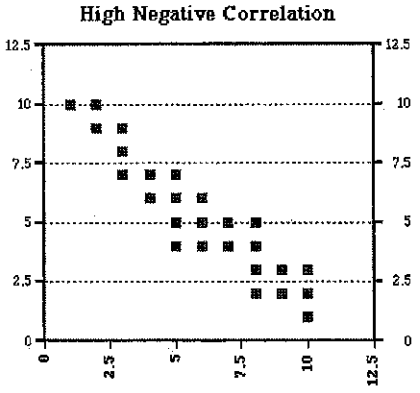
A scatter plot displays data as points on a grid using the associated numbers as coordinates or ordered pairs (x, y). The way the points are arranged by themselves in a scatter plot may or may not suggest a relationship between the two variables. For instance, by reading the graph below, do you think there is a relationship between the hours spent studying and exam grades?

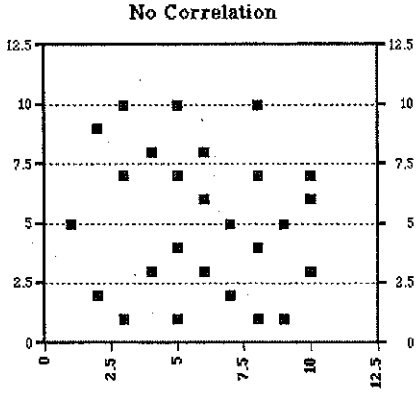


If y tends to increase as x increases, then the data have **positive** correlation.

If y tends to decrease as x increases, then the data have **negative** correlation.





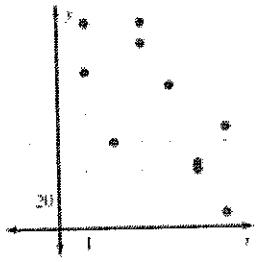


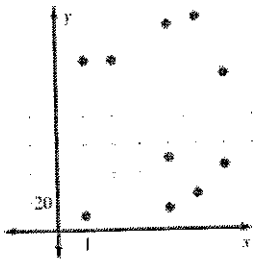
A correlation coefficient, denoted by r, is a number from -1 to 1 that measures how well a line fits a set of data pairs (x, y). If r is near 1, the points lie close to a line with a positive slope. If r is near -1, the points lie close to a line with a negative slope. If r is near 0, the points to not lie close to any line.

Give an example of negative correlation: r = -.3

Practice Problems:

For each scatter plot, tell whether the data have a positive correlation, a negative correlation, or no correlation. Then, tell whether the correlation is closest to -1, -0.5, 0, 0.5, or 1.

1. 

2. 

3. Positive, negative, or no correlation?

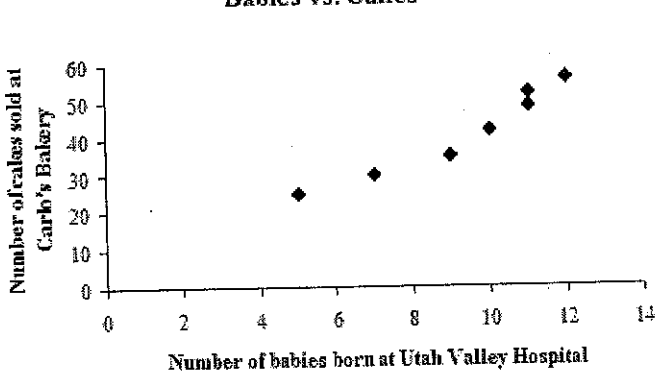
a. Amount of exercise and percent of body fat negative
b. A person's age and the number of medical conditions they have positive
c. Temperature and number of ice cream cones sold positive
d. The number of students at Hillgrove and the number of dogs in Atlanta NO
e. Age of a tadpole and the length of its tail positive as age

Correlation vs. Causation

When a scatter plot shows a correlation between two variables, even if it's a strong one, there is not *necessarily a cause-and-effect relationship*. Both variables could be related to some third variable that actually causes the apparent correlation. Also, an apparent correlation simply could be the result of chance.

Example 1: During the month of June the number of new babies born at the Utah Valley Hospital was recorded for a week. Over the same time period, the number of cakes sold at Carlo's Bakery in Hoboken, New Jersey was also recorded. What can be said about the correlation? Is there causation? Why or why not?

Number of babies born	Number of cakes sold
5	25
7	30
9	35
10	42
11	48
11	52
12	56



Example 2: Mr. Jones gave a math test to all the students in his school. He made the startling discovery that the taller students did better than the short ones. His Causation Statement: *As your height increases, so does your math ability.*
What can be said about the correlation? Is there causation? Why or why not?
correlation ⇒ positive NO causation ⇒ your height does not make you smarter

Example 3: In this present economy families are trying to find ways to save money Families might be thinking about not eating out to spend less money. Causation Statement: *The more you eat out, the more money you spend at restaurants.*
What can be said about the correlation? Is there causation? Why or why not?
positive correlation There is causation if you eat at you spend $

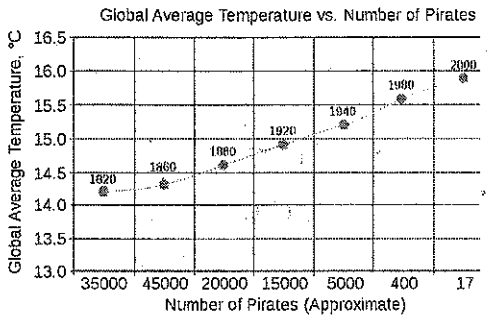
Name: _____ Date: _____

Correlation and Causation Homework

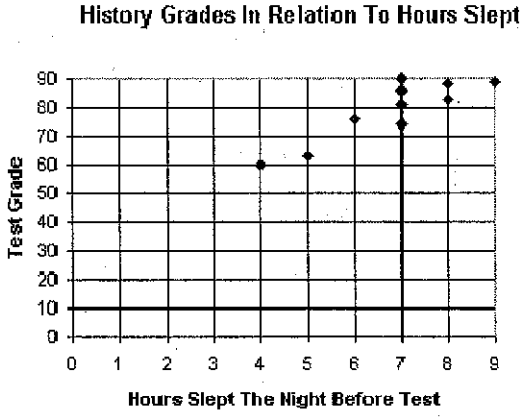
1. From the information given,
   a. Determine if the correlation is positive, negative or none. positive

   b. Estimate the correlation coefficient. .89

   c. Is there causation? Why or why not? There may be causation → pirates may only want to be in warmer parts of the world



2. A history teacher asked her students how many hours of sleep they had the night before a test. The data above shows the number of hours the student slept and their score on the exam. The graph is a scatter plot from the given data.



   a. Determine if the correlation is positive, negative, or none. positive

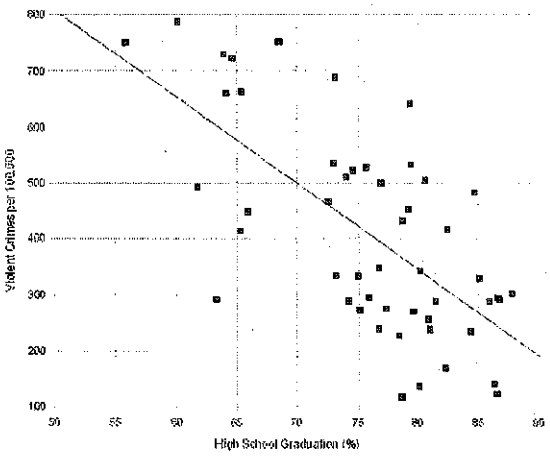
   b. Estimate the correlation coefficient. .85

   c. Is there causation? Would this information affect your behavior the night before a test? yes there is causation → more sleep → clearer head for test !!

3. The following chart shows violent crime rates compared to high school graduation for all fifty states.
   a. Determine if the correlation is positive, negative, or none. negative

   b. Estimate the correlation coefficient. .6

   c. Is this an illustration of cause and effect, or are these two variables simply correlated? Yes, there is causation, this does illustrate cause & effect. The more people in school, the less crimes bc kids are in the school building & not out making crime!



Adapted from: Mathematics Vision Project

For the given situations below,

a. **Is the association positive, negative or none?**
b. **Is the causation statement is true or false?**

4. When you are on a diet, the less calories you eat daily vs. the more weight you lose. Causation statement: *Therefore, eating less calories makes you lose weight.*

weight / cal

a) positive

b) true

5. The more ice cream consumed on a beach vs. the increased number of people who go in the water. Causation statement: *Therefore, eating more ice cream on the beach makes people go in the water.*

pos / isec c

a) positive

b) false

6. The more people in a family vs. the increased number of cars the family owns. Causation Statement: Therefore, the more people there are in a family determines how many cars a family owns.

a) positive

b) true

7. The average speed cars travel from Philadelphia to New York on the turnpike vs. the average amount of times it takes. Causation Statement: *Therefore, the speed cars travel from Philadelphia to New York determines the time it takes to go between them.*

a) negative

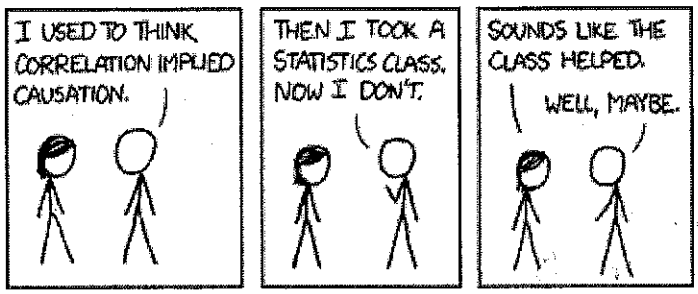
b) true

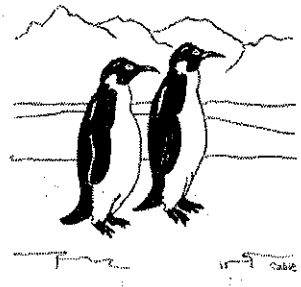
speed

8. How much you pay for a house vs. how much you pay for a car. Causation statement: *Therefore the more you pay for a house makes you spend more for a car.*

a) Negative

b) false





"Do you think all these film crews brought on global warming or did global warming bring on all these film crews?"

Adapted from: Mathematics Vision Project

Name: _____ Date: _____

Scatter Plots and Line of Best Fit – TV Task

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

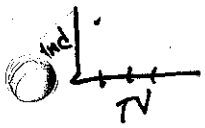
1. Students in Ms. Garth's Algebra II class wanted to see if there are correlations between test scores and height and between test scores and time spent watching television. Before the students began collecting data, Ms. Garth asked them to predict what the data would reveal. Answer the following questions that Ms. Garth asked her class.

   a. Do you think students' heights will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation?

   • I do not think students' height correlate to their test grades.

   b. Do you think the average number of hours students watch television per week will be correlated to their test grades? If you think a correlation will be found, will it be a positive or negative correlation? Will it be a strong or weak correlation? Do watching TV and low test grades have a cause and effect relationship?

   • Yes, I do think the more students watch TV, the lower their test grades. It will be a negative correlation. I do believe there is a cause & effect relationship.

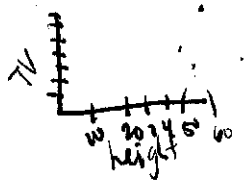


2. The students then created a table in which they recorded each student's height, average number of hours per week spent watching television (measured over a four-week period), and scores on two tests. Use the actual data collected by the students in Ms. Garth's class, as shown in the table below, to answer the following questions.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13
Height (in inches)	60	65	51	76	66	72	59	58	70	67	65	71	58
TV hrs/week (average)	30	12	30	20	10	20	15	12	15	11	16	20	19
Test 1	60	80	65	85	100	78	75	95	75	90	90	80	75
Test 2	70	85	75	85	100	88	85	90	90	90	95	85	85

   a. Which pairs of variables seem to have a positive correlation? Explain.

   Test 1 & Test 2



   b. Which pairs of variables seem to have a negative correlation? Explain.

   hours watching TV & test 1 & 2

   c. Which pairs of variables seem to have no correlation? Explain.

   height & hrs watch TV

3. Using the statistical functions of your graphing calculator, determine a line of good fit for each of the following categories.

   a. Score on Test 1 versus hours watching television:
   $y = -.46x + 55.5$ $r = -.8$

   b. Score on test 1 versus score on test 2:
   $y = .6x + 37.8$ $r = .89$

   c. Hours watching television versus score on test 2:
   $y = -1x + 104.13$ $r = -.85$

4. Use your answer to 3a to predict the test 1 score of someone who watches tv for 40 hours per week.

   $40 = -.46x + 55.5$
   $-55.5 \quad -55.5$
   $-15.5 = -.46x$
   $x = 33.7\%$

5. Use your answer to 3c to predict the test 2 score of someone who watches tv for 5 hours per week.

   $y = -1(5) + 104.13$
   $y = 99\%$ 11
   C

TI 83 or 84 of Best Fit Steps

1. DATA, then EDIT (type in data)
2. DATA, then CALC
3. 4: LinReg(ax+b)
4. a =

   b =

   r =

   ★ You can use the x variable button to find a, b, and r.
5. The equation of the line is y = **a**x + **b.**
6. Correlation Coefficient is r.
8. To predict use **a(predict #) + b**. *Estimated method*

Name: _____ Date: _____

Scatter Plots and Line of Best Fit

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.6c Fit a linear function for a scatter plot that suggests a linear association.

The **best fitting line or curve** is the line that lies as close as possible to all the data points.

Regression is a method used to find the equation of the best fitting line or curve.

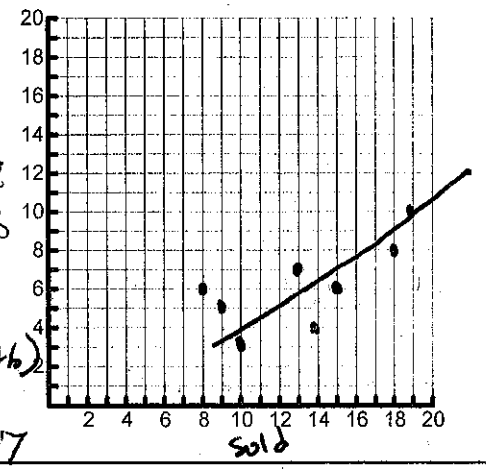
Extrapolation – the use of the regression curve to make predictions outside the domain of values of the independent variable.

Interpolation – Interpolation is used to make predictions within the domain of values of the independent variable.

Line of Best Fit by Hand:

1) The environment club is interested in the relationship between the number of canned beverages sold in the cafeteria and the number of cans that are recycled. The data they collected are listed in this chart.

Beverage Can Recycling								
Number of Canned Beverages Sold	18	15	19	8	10	13	9	14
Number of Cans Recycled	8	6	10	6	3	7	5	4



a) Plot the points to make a scatter plot.

b) Use a straightedge to approximate the line of best fit by hand.

c) Find an equation of the line of best fit for the data.

$y = \frac{2}{3}x - 2.7$ (10,4) (19,10)

① slope $m = \frac{10-4}{19-10} = \frac{6}{9} = \frac{2}{3}$

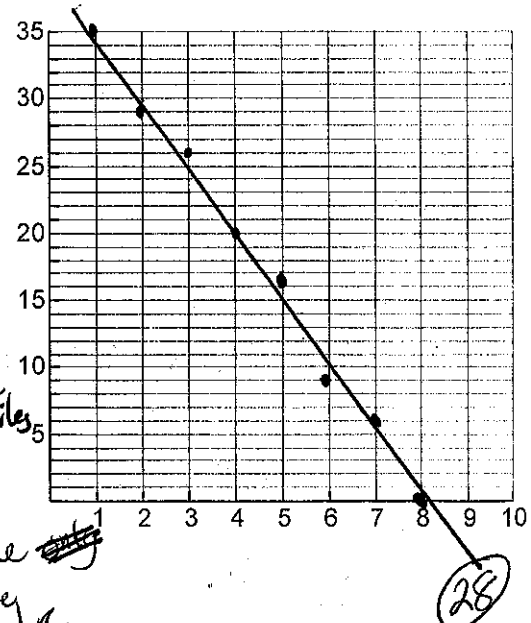
② pick a point & use slope to solve for b ($y = mx + b$)

$4 = \frac{2}{3} \cdot \frac{10}{1} + b$

$4 = 6\frac{2}{3} + b$ $b = -2.7$

2. Mike is riding his bike home from his grandmother's house. In the table below, x represents the number of hours Mike has been biking and y represents the number of miles Mike is away from home. Make a scatter plot for this data on the grid below.

Hours (x)	1	2	3	4	5	6	7	8
Miles (y)	35	29	26	20	16	9	6	0



a) Describe the association between the data points on the scatter plot.

b) Use a straightedge to approximate the line of best fit.

c) Find an equation of the line of best fit for the data.

(7,6) (2,29) $y = -4.5x + 37.5$ $6 = -4.5(7) + b$

d) What does the slope represent in the context of the problem? What does the y-intercept represent in the context of the problem?

Slope = how to get from 1 pt to next.... each hr he goes 4.5 miles

y-intercept → where he began trip at Mile marker 35

e) Could you use your equation to predict how far Mike would be after 10 hours? Use mathematics to justify your answer.

NO, bc he will be done after 10 hrs.

Line of Best Fit using the calculator

3) Use the table below to answer the questions about the population p (in millions) in Florida.

Year, t	2002	2003	2004	2005
Population (millions)	16.4	17.0	17.4	17.8

a) Find the best-fitting line for the data and the correlation coefficient.

$y = -.018x + 53.$ $r = -.4$

b) Using this model, what will be the population in 2020?

~~$y = 16.64$~~ $y = .46x - 904.46$ $r = .99$

4) Use the table below to answer the questions about the U.S. residential carbon dioxide emissions from 1993 to 2002. Emissions are measured in million metric tons.

Year, t	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Emissions	1027.6	1020.9	1026.5	1086.1	1077.5	1083.3	1107.1	1170.4	1163.3	1193.9

a) Find the best-fitting line for the data and the correlation coefficient.

$y = 19.9x - 38634$ $r = .96$

b) Using this model, how many residential tons were emitted in 1990? In 2010?

$y = 19.9(1990) - 38634$ $y = 967$

$y = 19.9(2010) - 38634$ $y = 1365$

5) Use the table below to answer the questions about the operating costs in thousands of a small business from 2000 to 2007.

Year, t	2000	2001	2002	2003	2004	2005	2006	2007
Operating Costs	2.3	2.6	3.1	3.3	4.0	5.2	5.9	7.0

a) Find the best-fitting line for the data and the correlation coefficient.

$y = -.04x + 89.7$ $r = -.99$

b) Using this model, what will be the operating costs in 2015?

$y = -.04(2015) + 89.7$

$y = 9.1$

Name: _____ Date: _____

Exponential Regression

MCC9-12.S.ID.6 Represent data on two quantitative variables on a scatter plot, and describe how the variables are related.
MCC9-12.S.ID.6a Fit a function to the data; use functions fitted to data to solve problems in the context of the data. Use the given functions or choose a function suggested by the context. Emphasize linear and exponential models.
MCC9-12.S.ID.8 Compute (using technology) and interpret the correlation coefficient of a linear fit.

Hot Coffee

The data at the right shows the cooling temperatures of a freshly brewed cup of coffee after it is poured from the brewing pot into a serving cup. The brewing pot temperature is approximately 180° F.

Time (mins)	Temp (F°)
0	179.5
5	168.7
8	158.1
11	149.2
15	141.7
18	134.6
22	125.4
25	123.5
30	116.3
34	113.2
38	109.1
42	105.7
45	102.2
50	100.5

a) Determine an exponential regression model equation to represent this data. have to use graphing calc. $y = 171.46(.99)^x$

*use on-line graphing calc. if don't have your own

b) Decide whether the new equation is a "good fit" to represent this data.

r - was not given???

c) Based upon the new equation, what was the initial temperature of the coffee? What is the decay rate?

initial → 171.46 rate – 1% or .01 ~~2% or .02~~

d) Interpolate data: When is the coffee at a temperature of 106 degrees?

$106 = 171.46(.99)^x$ around 41 min

e) Extrapolate data: What is the predicted temperature of the coffee after 1 hour?

$171.46(.99)^{60} \approx 93.8°$

f) In 1992, a woman sued McDonald's for serving coffee at a temperature of 180° that caused her to be severely burned when the coffee spilled. An expert witness at the trial testified that liquids at 180° will cause a full thickness burn to human skin in two to seven seconds. It was stated that had the coffee been served at 155°, the liquid would have cooled and avoided the serious burns. The woman was awarded over 2.7 million dollars. As a result of this famous case, many restaurants now serve coffee at a temperature around 155°. How long should restaurants wait (after pouring the coffee from the pot) before serving coffee, to ensure that the coffee is not hotter than 155°?

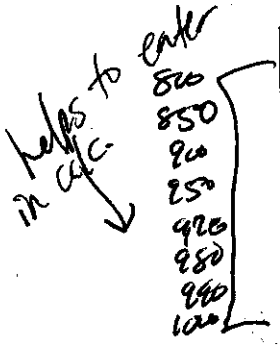
They should wait around 10 minutes before they serve

g) If the temperature in the room is 76° F, what will happen to the temperature of the coffee, after being poured from the pot, over an extended period of time?

The coffee will remain at 76°

Practice Problems:

1. Estimates for world population vary, but the data in the accompanying table are reasonable estimates of the world population from 1800 to 2000.



Year	Total Population (millions)
1800	980
1850	1260
1900	1650
1950	2520
1970	3700
1980	4440
1990	5270
2000	6080

a. Identify your independent and dependent variables.

year – independent population – dependent

b. Generate a best fit exponential function using your variables. Round to 3 decimals.

$y = .539(1.009)^x$

c. What does your model give for the growth rate? Describe this in the context of the problem.

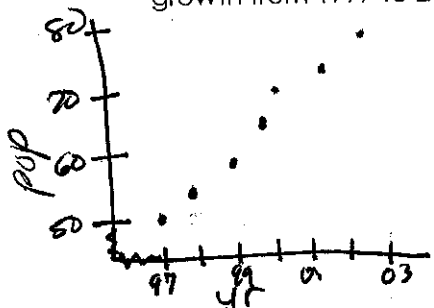
.009 → .09% The population grew at .09%.

d. Using the function, estimate the world population in 1750 and 2050 to 3 decimal places.

$y = .539(1.009)^{750}$
$= 446.6$ millions

$y = .539(1.009)^{2050}$
$= 6566$ million

2. **Town Planning**: The town planners designed their town for an optimal growth of 8% per year. The present school construction will serve a population of 200,000. Below is a table representing the growth from 1997 to 2003.



Year	Population
1997	50,000
1998	54,000
1999	58,000
2000	62,986
2001	68,024
2002	73,466
2003	79,344

a) Find and write the model of a linear regression. Use the model to determine what the population was in 1977. Round to 2 decimals.

$y = 4.89x - 425.6$ $r = .99$

b) Find and write the model of an exponential regression. Use the model to determine what the population was in 1977. Round to 2 decimals.

$y = 1.95(1.08)^x$ $y = 1.95(1.09)^{977}$
$= 7174$

c) Determine which model is better to use. Explain why you selected your model.

look at r – correlation coefficient (linear) → $r = .99$

d) Using the better model, predict what the population will be in the year 2017.

~~$y = 1.95$~~

e) In what year will the population double for the better model?

$100,000 = 4.89x - 425.6$ 2007 pop will double

~~$x = 20$~~