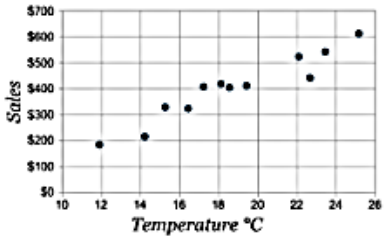
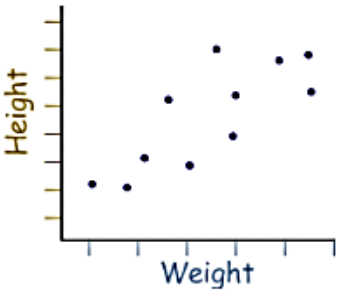
Scatterplots and Correlation

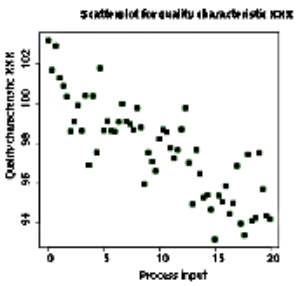
Name_____

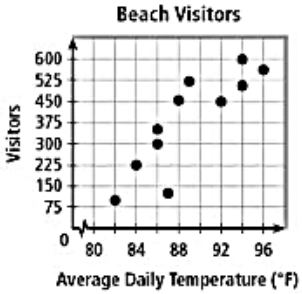
The following are all examples of scatterplots.

Period:_____





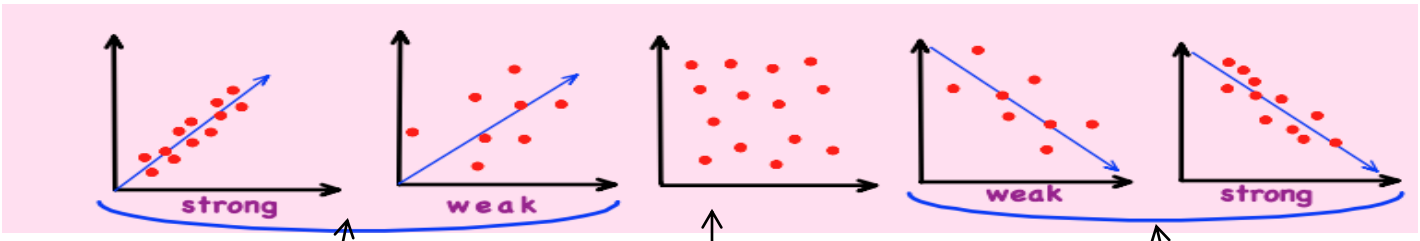




1) How do you think you could explain to someone what a scatterplot is after looking at these examples?

2) Do you notice any 'trends' in the data that is displayed? Explain.

The formal word for these trends is **correlation.** Here are examples of all possible linear correlations.



_____ correlation: As one set of data _____, the other set of data _____.

_____ correlation: There is no visible relationship between the values.

_____ correlation: As one set of data _____, the other set of data _____.

Do the following data sets have a positive, a negative, or no correlation?

3. The age and weight of a baby.

   As a baby gets older, his weight _____. *Both sets of data* _____, *so the data has a* _____ *correlation.*

4. The amount of free time you have and the number of sports that you play.

   Your free time _____ *as the number of sports you play* _____; *therefore, the data has a* _____ *correlation.*

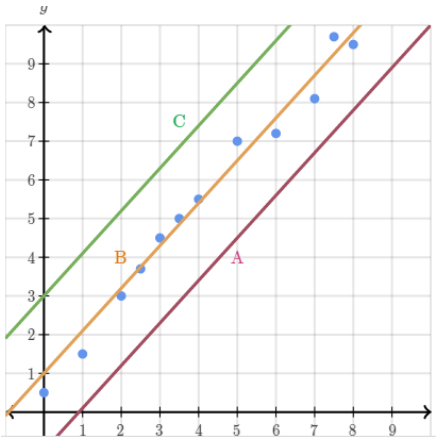
5. The price of a shirt and the color of its buttons.

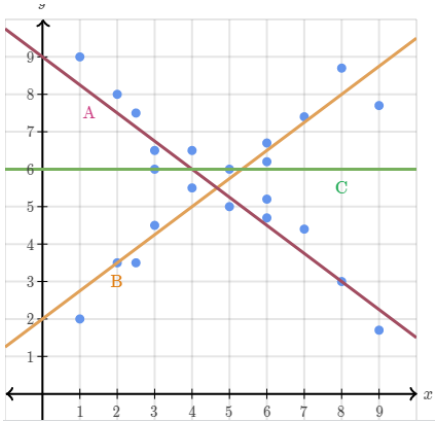
   The price is not affected by the color of its buttons, so the data has _____ *correlation.*

When we see a relationship in a scatterplot, we can use a line to summarize the relationship in the data. We can also use that line to make predictions in the data. This process is call **linear regression**.

There are more advanced ways to fit a line to data, but in general, we want the line to go through the middle of the data points so that half of the points are above the line while the other half is below. The line should also go with the flow of the data.

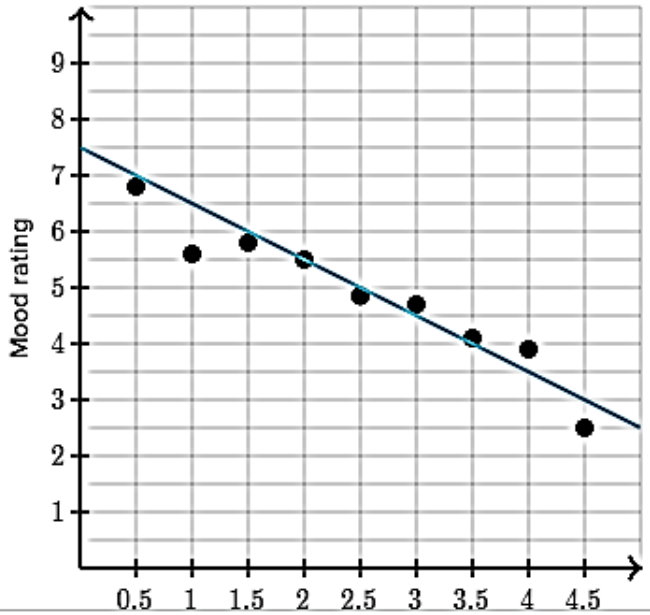
6) Which line best fits the data in each graph? Why?

a) 

b) 

We can also estimate equations of lines and use them to make predictions about the data.

7) Jacob distributed a survey to his fellow students asking them how many hours they spent on the Internet in the past day. He also asked them to rate their mood on a scale from 0 to 10, with 10 being the happiest. A line was fit to the data to model the relationship.



Which of these linear equations best describes the given model?

Choose 1 answer:

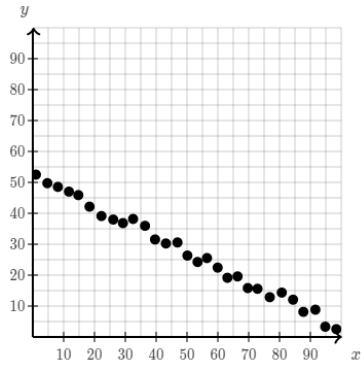
(A) $\hat{y} = x + 7.5$

(B) $\hat{y} = -x + 7.5$

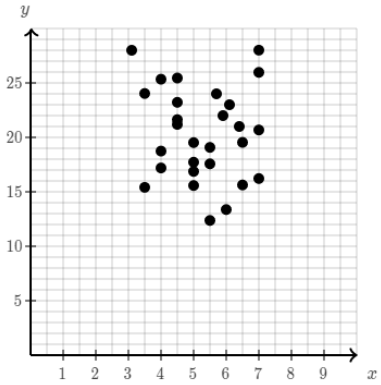
(C) $\hat{y} = -\frac{1}{2}x + 7.5$

You try ☺ Label the following as positive, negative, or no correlation. Also, tell whether it's strong or weak.

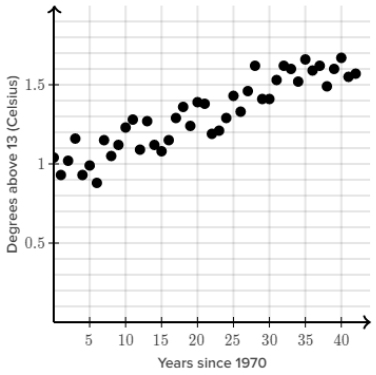
1)



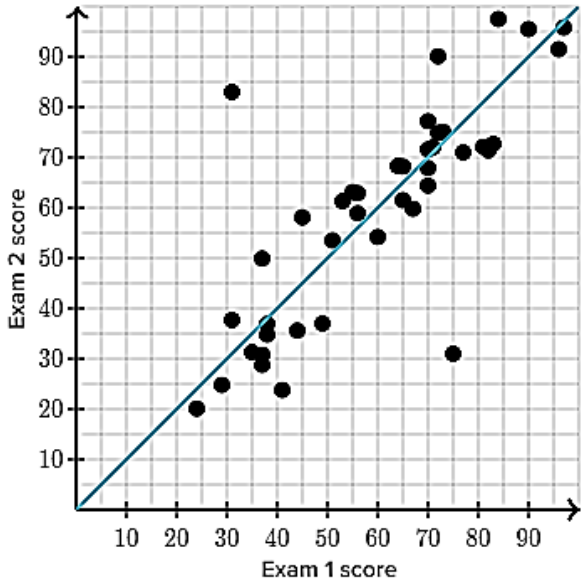
2)



3)



4) The plot shown below describes the relationship between students' scores on the first exam in a class and their corresponding scores on the second exam in the class. A line was fit to the data to model the relationship.



Which of these linear equations best describes the given model?

Choose 1 answer:

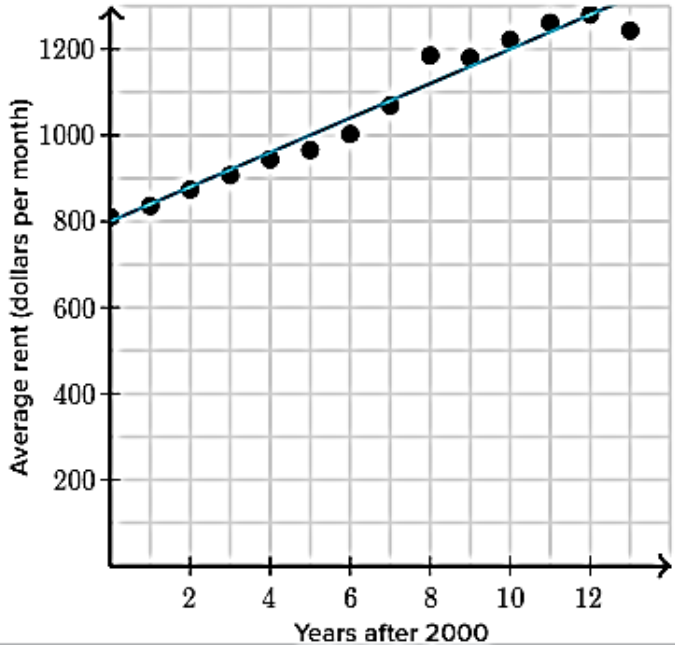
(A) $\hat{y} = 2x + 10$

(B) $\hat{y} = x + 10$

(C) $\hat{y} = 2x$

(D) $\hat{y} = x$

5) The scatter plot below shows the average rent (in dollars per month) for a 1-bedroom apartment in New York City each year between 2000 and 2013. A line was fit to the data to model the relationship.



Which of these linear equations best describes the given model?

Choose 1 answer:

(A) $\hat{y} = \frac{2}{5}x + 800$

(B) $\hat{y} = 40x + 800$

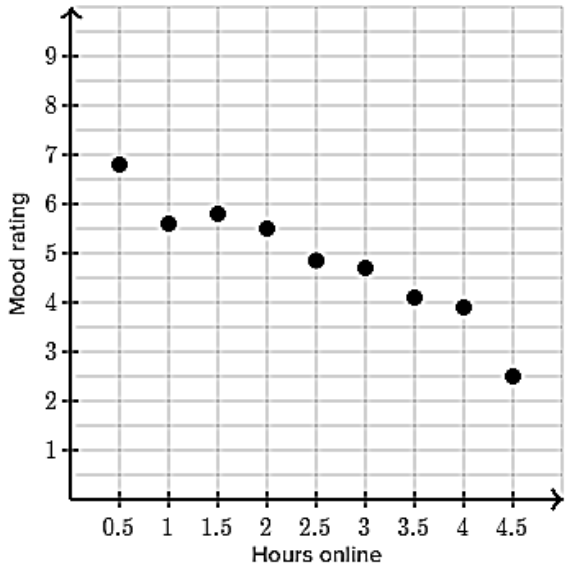
(C) $\hat{y} = 800x + \frac{2}{5}$

(D) $\hat{y} = 800x + 40$

6) Jacob distributed a survey to his fellow students asking them how many hours they spent on the Internet in the past day. He also asked them to rate their mood on a scale from 0 to 10, with 10 being the happiest.

The scatter plot below shows the mood rating for each amount of time online.

Which of the following is the best estimate of the average change in mood rating associated with a 1 hour increase in hours online?



Choose 1 answer:

(A) -4 points

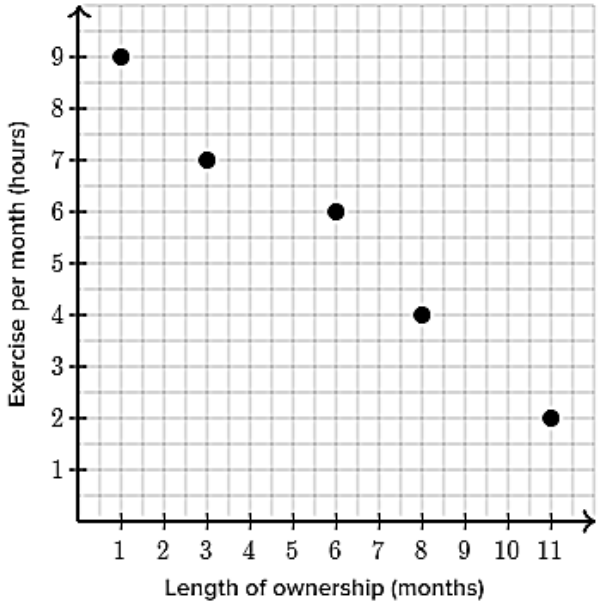
(B) -2 points

(C) -1 point

(D) $-\frac{1}{2}$ point

7) A company that makes an exercise machine asked some of its recent customers how long they have owned their machines, and how many hours per month they currently use their machine.

Which of the following is the best estimate of the average change in hours of exercise per month that was associated with a 1 month increase in ownership?



Choose 1 answer:

(A) -5 hours

(B) -3 hours

(C) $-\frac{3}{4}$ hours

(D) $-\frac{1}{3}$ hours