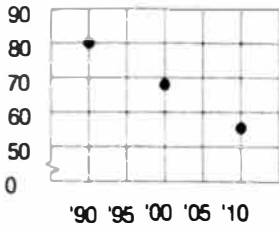


1. How would you describe the relationship between the x- and y-values in the scatter plot?



As the x-values increase,
the y-values decrease.

2. Based on the data in the scatter plot in #1, what would you expect the y-value to be for $x = 2020$? (The x-axis is years, and '90 = 1990.)

About 42-45

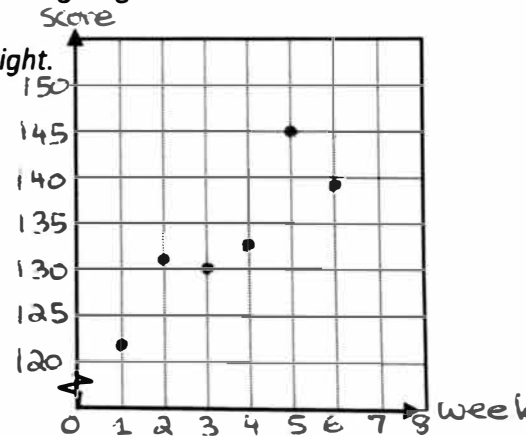
3. Which correlation coefficient corresponds to the best-fit line that most closely models the set of data in #1? How do you know?

- a) 0.84 b) 0.13 **c) -0.87** e) -0.15

4. The table below shows Kyle's bowling score each week he participated in a bowling league.

Week	1	2	3	4	5	6
Score	122	131	130	133	145	139

a. Make a scatterplot to the right.



b. Which of the following is the best equation for the line of best fit. EXPLAIN your choice.

- I) $y = 120.3x + 3.7$ II) $y = -120.3x + 3.7$
III) $y = 3.7x + 120.3$ IV) $y = -3.7x + 120.3$

Positive correlation
 Positive slope and y-intercept of 120.3

c. Estimate Kyle's score for week 9, round to the nearest whole number. Explain HOW you found this estimate.

$$y = 3.7(9) + 120.3$$

$$= 153.6 \approx 154$$

d. Find the equation of the line if you used the data points from week 1 and 3.

(1, 122) (3, 130) $m = \frac{130 - 122}{3 - 1} = \frac{8}{2} = 4$

$$y - 122 = 4(x - 1)$$

$$y - 122 = 4x - 4$$

$$y = 4x + 118$$

e. Finish the statement about the data:

As the number of weeks increases, Kyle's bowling score tends to increase.

5. Describe what it means for a scatter plot to present a negative correlation. Give an example of a situation that would create this type of graph.

Negative correlation \rightarrow As x increases, y tends to decrease.

i.e. - Time spent on playing video game and test scores.

6. What is the difference between correlation and causation? \swarrow cause-and-effect.

Correlation does not imply causation (cause and effect).

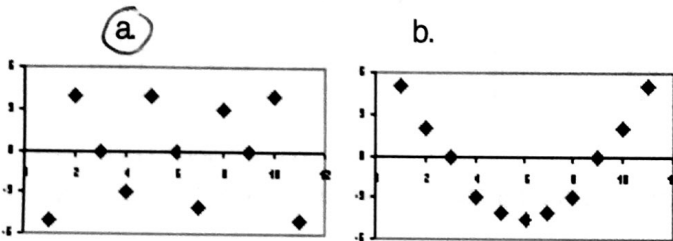
Correlation can be measured its strength by a number between

1 and -1. A strong correlation means that the two variables

tend to vary together in a predictable way, which might be due to

7. Which of the residual plots below would indicate that a linear model is appropriate? Why? other than

a cause-and-effect.



The points are randomly scattered for (a).

8. A line of best fit might be defined as

a. a line that connects all the data points.

(b) a line that might best estimate the data and be used for predicting values.

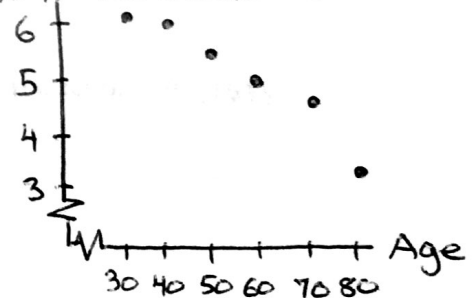
c. a vertical line halfway through the data.

d. a line that has a slope greater than 1.

9. a. Make a scatter plot relating the age to the % of the person's budget spent on entertainment. Label axes.

Age	30	40	50	60	70	80
% Spent on Entertainment	6.1	6.0	5.4	5.0	4.7	3.4

% spent on entertainment.



- b. Which equation below should be used to represent a line of best fit for the data?

Justify your answer.

(a) $y = -0.05x + 7.5$

b. $y = -0.05x - 7.5$

c. $y = 0.05x + 7.5$

d. $y = 0.05x - 7.5$

- c. Predict the % of a 65-year-old person's budget that would be spent on entertainment, round to the nearest tenth.

$$y = -0.05(65) + 7.5$$

$$= 4.25 \approx 4.3\%$$

- d. Is it reasonable to use the equation to estimate the entertainment spending for all ages? Explain your reasoning.

No, (0-5) too young.

- e. Make a statement about the data: As age increases, the percent of budget spent on entertainment tends to decrease.

10. Use the table and scatterplot below, which show the number of hours different players practice basketball each week and the number of baskets each player scored during a game.

Player	Bill	Ryan	Tanja	Cami	Sonia	Ingrid	Esther	Danae
Hours	5	10	7	0	12	3	9	6
Baskets	6	11	8	2	12	7	8	10

- a. Use the graph of the data to sketch the line of best fit.

- b. Identify and interpret the slope in the context of the situation.

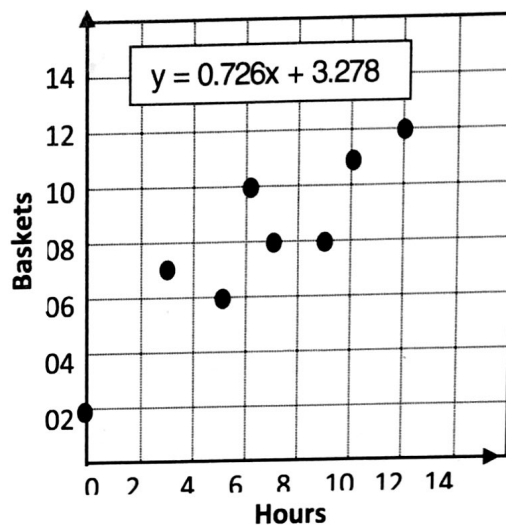
Slope = 0.726

Meaning: For every hour practiced, they typically make an additional 0.726 basket.

- c. Identify and interpret the y-intercept in the context of the situation, or explain why it is unreasonable.

y-Int = 3.278

Meaning: with no practice, the predicted number of baskets they will make is 3.278



- d. Which of the following would be a good estimate for the correlation coefficient (r)? Explain why.

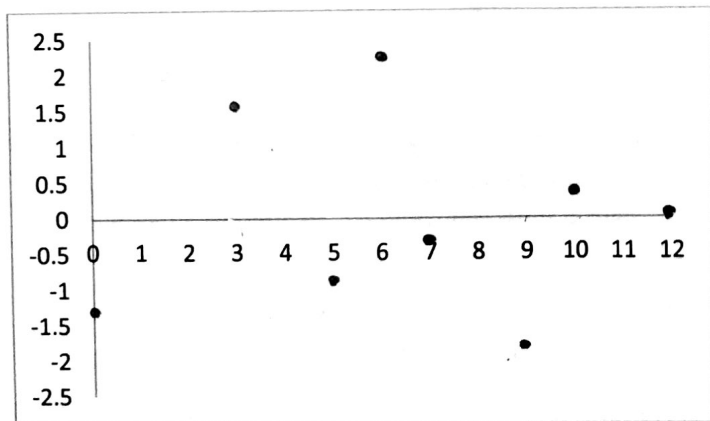
- a) -0.89 b) -0.50 c) 0.01 d) 0.50 e) 0.89

- e. Use the line of best fit to calculate the expected values, then find the residuals. Round to the nearest hundredth.

$$y = 0.726x + 3.278$$

Player	Hours x	Actual Basket y	Predicted Baskets	Residual
Bill	5	6	6.91	-0.91
Ryan	10	11	10.54	0.46
Tanja	7	8	8.36	-0.36
Cami	0	2	3.28	-1.28
Sonia	12	12	11.99	0.01
Ingrid	3	7	5.46	1.54
Esther	9	8	9.81	-1.81
Danae	6	10	7.63	2.37

- f. Create a residual plot below; determine if a linear model is appropriate for this data. Explain why or why not.



11. Mrs. Burhans' class took a Unit 2 Quiz, and then a Unit 2 Exam. The scatterplot of the data is shown below.

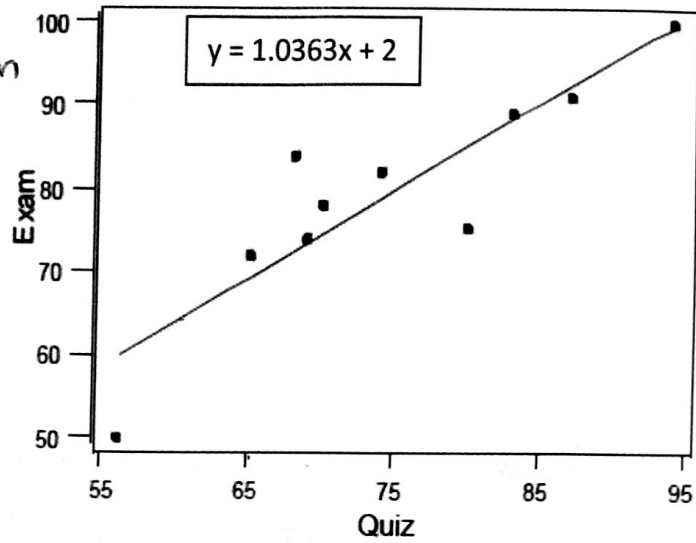
a. The graph shows what type of correlation?

a strong positive linear correlation

b. If a student scored a 60% on the quiz, what is their predicted test score? Round to the hundredth.

$$y = 1.0363(60) + 2$$

$$y = 64.18\%$$



c. Which of the following is true about this data?

- I) Someone scored a 68% on the quiz, but got over 80% on the exam.
- II) The students that scored between 65-75% on the quiz, earned strictly between 70-80% on the exam.
- III) A high quiz score DEFINITELY means you will get a high grade on the exam.
- IV) A slope of 1.063 means that in general the students did better on the exam than on the quiz.
- V) A y-intercept of 2 means that if you got a 0% on the quiz, you are predicted to get a 2% on the exam.
- VI) The correlation coefficient is closer to 1 than to -1.

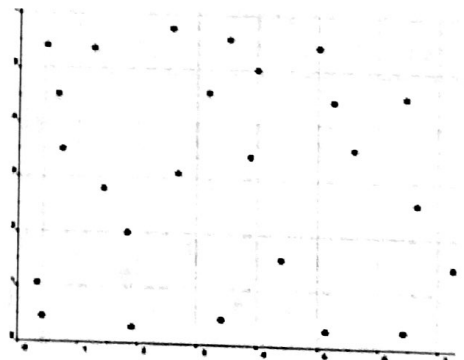
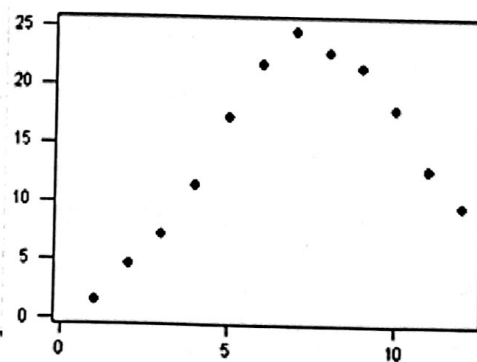
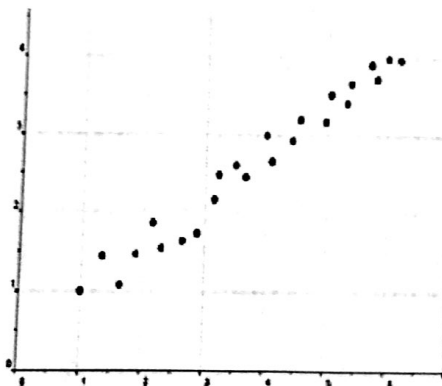
exam
quiz

12. Given the three scatter plots below, answer the following questions.

#1

#2

#3



a. Which scatterplot is best described by each of the following:

Quadratic #2 Linear #1 No Relationship #3

b. Scatterplot # 3 could be about shoe size and their score on the ACT test.

c. Scatterplot # 1 could be about how as height increases, weight also tends to increase.

d. Scatterplot # 2 could be about how over time iPod sales increased, and then decreased (as more people bought iPhones instead).

13. Sophie drew a line of best fit through the two points (2,3) and (11,75). Write the equation for her line of best fit in slope-intercept form.

$$m = \frac{75-3}{11-2} = \frac{72}{9} = 8$$

$$y - y_1 = m(x - x_1)$$

$$y - 3 = 8(x - 2)$$

$$y - 3 = 8x - 16$$

$$\boxed{y = 8x - 13}$$

14. Lara drew a line of best fit through the two points (15,7) and (43,22). Write the equation for her line of best fit in slope-intercept form.

$$m = \frac{22-7}{43-15} = \frac{15}{28} = \frac{1}{2}$$

$$y - 7 = \frac{1}{2}(x - 15)$$

$$y - 7 = \frac{1}{2}x - \frac{15}{2} + 7$$

$$\boxed{y = \frac{1}{2}x - \frac{1}{2}}$$

$$\begin{aligned} &-\frac{15}{2} + \frac{7}{1} \\ &-\frac{15}{2} + \frac{14}{2} = -\frac{1}{2} \end{aligned}$$

15. Use **desmos.com** to complete the following problem. Round all answers to the nearest hundredth.

- a. Least squares line:

$$m = 11.73$$

$$b = 193.85$$

$$y = 11.73x + 193.85$$

Sandwich	Total Fat (g)	Total Calories
Hamburger	9	260
Cheeseburger	13	320
Quarter Pounder	21	420
Quarter Pounder with Cheese	30	530
Big Mac	31	560
Arch Sandwich Special	31	550
Arch Special with Bacon	34	590
Crispy Chicken	25	500
Fish Fillet	28	560
Grilled Chicken	20	440
Grilled Chicken Light	5	300

- b. Correlation coefficient: $r = \underline{0.98}$

- c. Interpretation of the correlation coefficient:

A strong positive linear relationship.

- d. If a sandwich has 410 calories, what is its predicted grams of fat?

$$y = 11.73x + 193.85$$

$$410 = 11.73x + 193.85$$

$$x \approx 18.43$$

$$216.15 = 11.73x$$

$$18.43 \approx x$$

- e. If a sandwich has 27 grams of fat, what is its predicted number of calories?

$$y = 11.73(27) + 193.85$$

$$y = 510.56 \text{ calories}$$