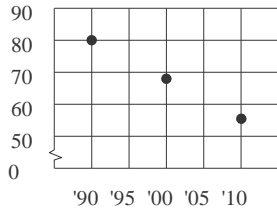


Unit 6 Review

Name _____ Per _____

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



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.)

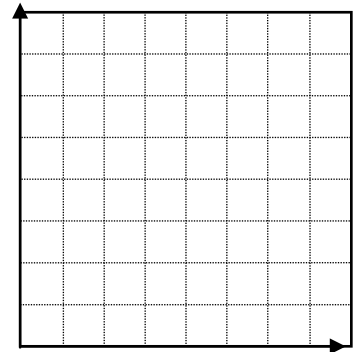
3. Which correlation coefficient corresponds to the best-fit line that most closely models the set of data in #1? How do you know?

- 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) $y = -3.7x + 120.3$

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

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

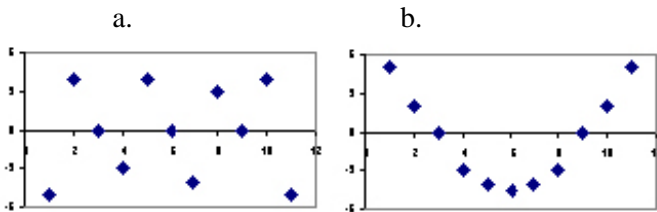
e. Make a statement about the data:

As the number of weeks increases, _____.

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.

6. What is the difference between correlation and causation?

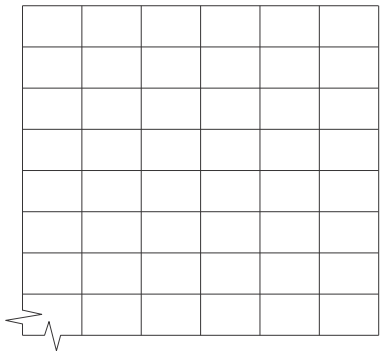
7. Which of the residual plots below would indicate that a linear model is appropriate? Why?



8. A line of best fit might be defined as
- a line that connects all the data points.
 - a line that might best estimate the data and be used for predicting values.
 - a vertical line halfway through the data.
 - 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



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

- $y = -0.05x + 7.5$
- $y = -0.05x - 7.5$
- $y = 0.05x + 7.5$
- $y = 0.05x - 7.5$

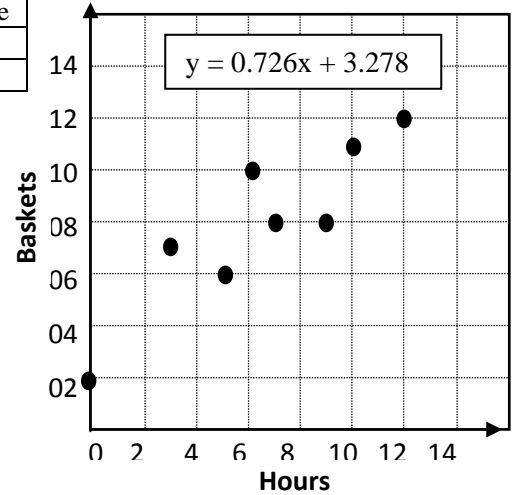
c. Predict the % of a 65-year-old person’s budget that would be spent on entertainment, round to the nearest tenth.

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

e. Make a statement about the data: As age increases, _____.

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	Ayo	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 a line of best fit.
- b. Identify and interpret the slope in the context of the situation.

Slope = _____

Meaning:

- c. Identify and interpret the y-intercept in the context of the situation.

y-Int = _____

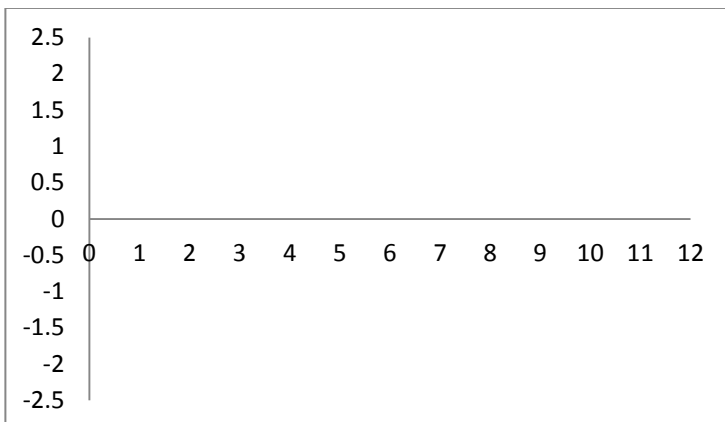
Meaning:

- 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.

Player	Hours	Actual Basket	Predicted Baskets	Residual
Bill	5	6		
Ryan	10	11		
Tanja	7	8		
Cami	0	2		
Sonia	12	12		
Ingrid	3	7		
Ayo	9	8		
Danae	6	10		

- 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.

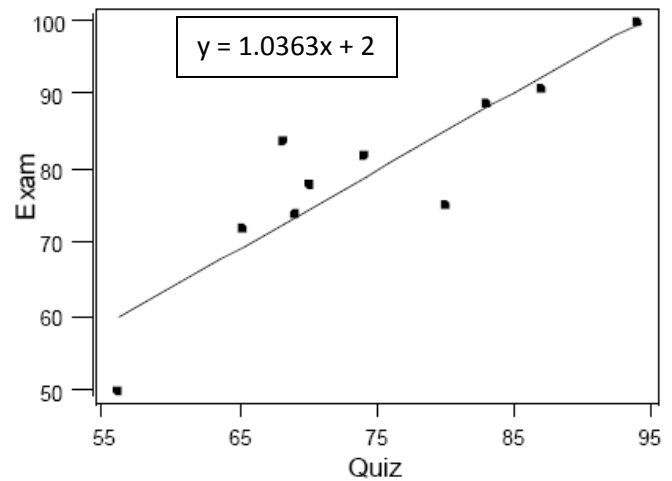
a. The graph shows what type of correlation?

b. If a student scored a 60% on the quiz, which of the following is the best choice for their predicted test score?

- I) 70% II) 65% III) 60%

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.

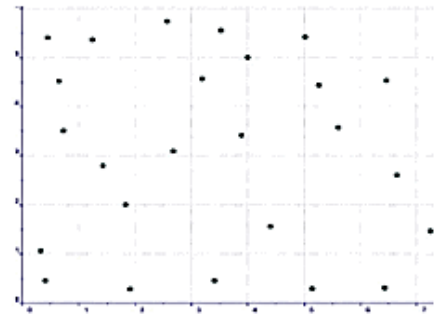
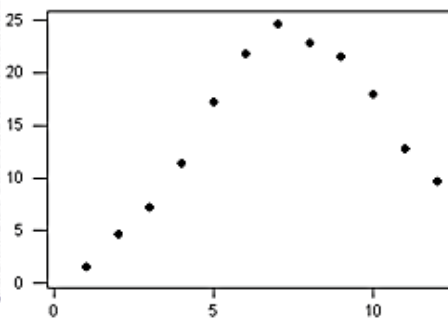
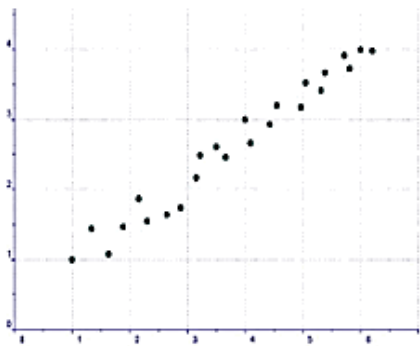


12. Given the three scatter plots below, answer the following questions.

#1

#2

#3



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

Quadratic _____ Linear _____ No Relationship _____

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

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

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

Answer Key

1. As the x-values increase, the y-values decrease.	2. About 42	3. c - There is a strong negative correlation between the points.	4. a. Scatter plot
4. b. III is the best equation since the data has a positive correlation, and III has a positive slope with a y-intercept of 120.3.	4. c. 154 points I substituted 9 into the x-value of the line of best fit and got an answer of 153.6, it rounds to 154.	4. d. $y = 4x + 118$ 4. e. Kyle's bowling score tends to increase.	5. A negative correlation means that as x increases, the y's tend to decrease. (Examples will vary.)
6. Responses will vary.	7. Residual plot a indicates that a linear model is appropriate because the points are scattered randomly above and below the line.	8. b.	9. a. Scatter plot
9. b. Line a is correct since there is a negative correlation, and the y-intercept appears to be about 7.5.	9. c. 4.3%	9. d. No, it is not reasonable since [explanations will vary]. 9. e. the percent of budget spent on entertainment tends to decrease.	10. a. Scatter plot
10. b. The slope is 0.726 which means that for every hour practiced, they typically make an additional 0.726 baskets in a game.	10. c. The y-intercept is 3.278 which means that if someone practices for 0 hours, the predicted number of baskets they will make in a game is 3.278.	10. d. (e) is the best choice, since the data has a strong positive correlation.	10. e. [table below]

10. e.

Player	Hours	Actual Basket	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
Ayo	9	8	9.81	-1.81
Danae	6	10	7.63	2.37

10. f. Residual plot Yes, a linear model is appropriate since the points are scattered randomly above and below the residual line.	11. a. Strong positive linear correlation	11. b. II	11. c. I, IV, V, VI
12. a. Quadratic #2 Linear #1 No Relationship #3	12. b. #3	12. c. #1	12. d. #2