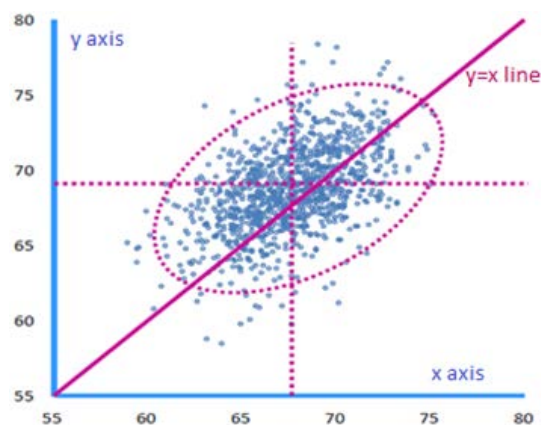1. In the scatter diagram shown below, the dotted straight lines mark the average values of X and Y. From the diagram, we can say that

   (A) The Y=X line cuts through the data points in half, with 50% of the data points on either side of the line.

   (B) The average of X is larger than the average of Y.

   (C) The average of Y is larger than the average of X.

   Explanation:

   Answer is (C). Since the dotted straight lines intersect above the Y=X line, the average of Y is larger than the average of X. The Y=X line does not cut through the data points in half.

2. For the scatter plots below, choose the most appropriate option corresponding to the correlation coefficients r1 and r2 in figures (1) and (2) respectively.
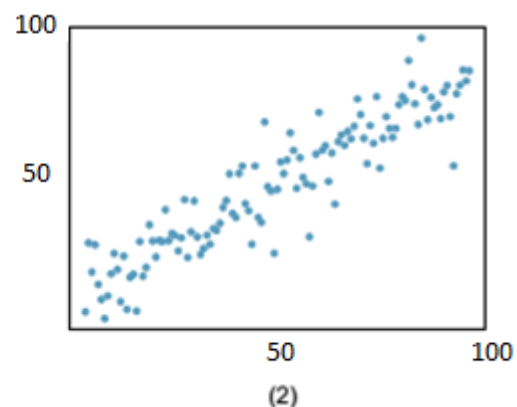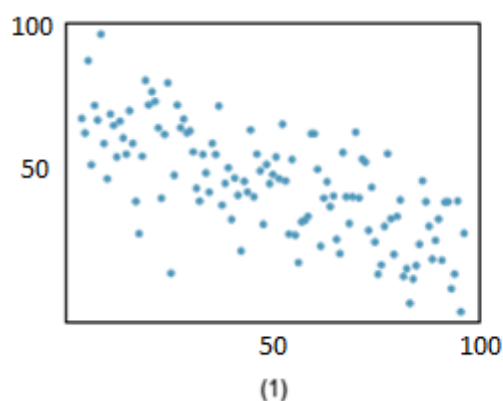
   (A) r1 = 0.45, r2 = 0.7
   (B) r1 = -0.45, r2 = -0.7
   (C) r1 = 0.45, r2 = -0.7
   (D) r1 = -0.45, r2 = 0.7

Explanation:

Answer is (D). For (1), there is a negative linear association. (2) shows a positive linear correlation.

3. Company ABC follows a 'buddy-system' where each new employee is paired with a mentor, who is older in age. Imagine a scatter diagram with Y axis representing mentors' age and X axis representing mentees' age, where each point represents a mentor-mentee pair. Which of the following must be true about the employees?

(I) The correlation between mentors' age and mentees' age is positive.

(II) The relationship between mentors' age and mentees' age may best be described by a line.

(A) (I) only

(B) (II) only

(C) (I) and (II)

(D) Neither (I) nor (II)

Explanation:

Answer is (D). Since mentors are strictly older than their respective mentees, all points will lie above the Y=X line. With the single condition that all points lie above Y=X line, we are unable to conclude the direction of correlation between X and Y. Moreover, we also cannot determine whether a line or a curve is more appropriate in describing the relationship between X and Y.

4. As a part of an assignment, the students in a physics class are asked to read a 10-page document. At the end of the class, the number of pages read varies across students. Each student is represented in a scatter diagram, where the x-axis is the number of pages read, and the y-axis is the number of pages left to read. The correlation coefficient is

(A) -1

(B) -0.5

(C) 0

(D) +0.5

(E) 1

Explanation:

Answer is (A). Let X and Y be the number of read and unread pages respectively. Since there are 10 pages, X+Y=10 or Y=10-X, which means all points fall on a perfect line. X and Y are negatively correlated: When X increases, Y decreases and when X decreases, Y increases. Thus, r = -1.

5. Which of the following statements about correlation coefficient is most appropriate?

(A) If correlation coefficient is 0, there is no relationship between two variables.

(B) When there is a weak linear association, the correlation coefficient is close to 0.

(C) If the correlation coefficient is 0.7, then 70% of the data points are tightly clustered along a line.

(D) The correlation coefficient is changed if we add/multiply 3 to all the values of one variable.

Explanation:

Answer is (B). Correlation coefficient r measures linear association between two variables. The lower the value, the weaker the linear association. A zero r value indicates no LINEAR association, but there can still be a nonlinear relationship between two variables. We cannot interpret r as the percentage of data points tightly clustered along a line. The r value is not affected by adding any number or multiplying any positive number to all values of any of the 2 variables.

6. Alicia was testing a new weighing equipment. To see if her readings are consistent, she weighs a chocolate bar 10 times. Every time she weighed the bar, she got the same reading of 20 grams. Which of the following statements is NOT true about the dataset Alicia has?

(A) The average is equal to 20 g.

(B) Standard deviation describes the spread of data around the average.

(C) The standard deviation is equal to 0 g.

(D) The standard deviation is equal to 20 g.

Explanation:

Answer is (D). All the other statements are true about the dataset. Since all the readings are equal to 20 g, the average is 20 g. By definition, standard deviation describes the spread of data around the average. In this dataset, all the readings are equal to the average of 20 g. Hence the standard deviation is equal to 0 g.