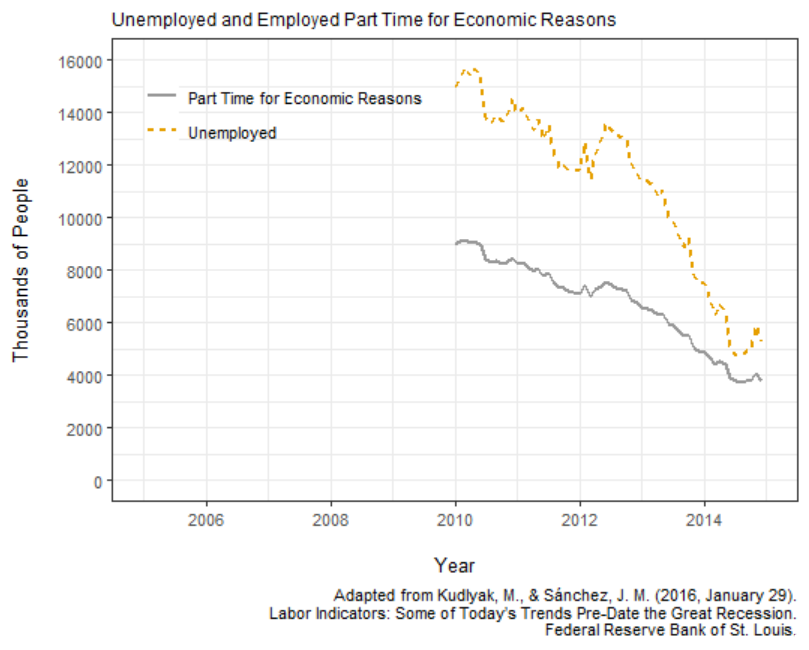
1. A team of researchers found that countries that mandated Bacillus Calmette-Guérin (BCG) vaccination had lower COVID-19 death rates compared to countries that did not mandate BCG vaccination. Based on this association alone, the researchers concluded that mandating BCG vaccinations causes lower COVID-19 death rates. Which one of the following fallacies have they committed?
2. Atomistic Fallacy
3. Ecological Fallacy
4. Confusing Correlation and Causation
5. None of the other options

*Explanation: See Chapter 2: Slide 57. The researchers found an association, not necessarily a causation. Healthcare infrastructure could be a confounder: countries with mandated BCG vaccination tend to have better healthcare, which leads to lower COVID-19 death rates.*

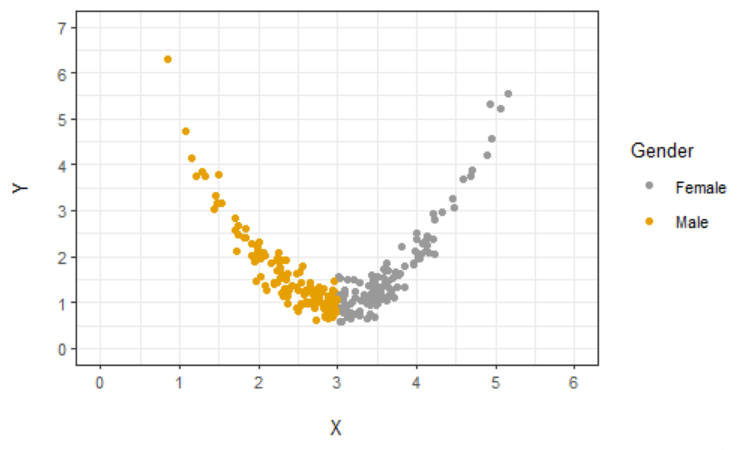
1. The graph shows the trend of the number of unemployed people (y) and the number of people who worked part time for economic reasons (x) during the period 2010-2015. You may assume that both variables were measured monthly. An economist plots y against x in a scatter diagram, so it has a total of 60 points. The correlation coefficient between x and y is



1. Negative
2. Zero
3. Positive

*Explanation: From 2010 to 2015, both numbers tend to decrease in time. But when x is large, y also tends to be large, and when x is small, y also tends to be small. In fact, the correlation coefficient is around 0.99.*

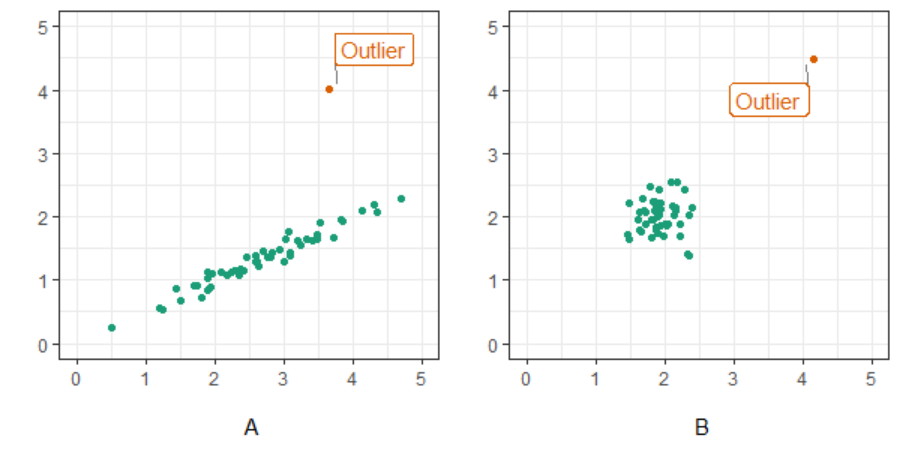
1. A researcher examined the relationship between Variables X and Y among 250 male and female subjects. He graphed the relationship in the scatter diagram below. Let r1 be the correlation coefficient among all 250 subjects, and r2 be the correlation coefficient among male subjects only. Which of the following statements is true?



1. r1 > r2
2. r1 = r2
3. r1 < r2

*Explanation: See Chapter 2, Slide 64. The correlation for all subjects is approximately zero. The correlation for males only is negative.*

1. What will the new correlation coefficients for Plot A and Plot B be respectively, after removing the outliers?



1. Close to 1; Close to 1
2. Close to 1; Close to 0
3. Close to 0; Close to 1
4. Close to 0; Close to 0

*Explanation: See Chapter 2, Slides 61-62. For Plot A, removing the outlier brings the correlation close to 1. For Plot B, removing the outlier brings the correlation close to 0.*

1. A scatter diagram contains several hundred points. The horizonal and vertical axes represent variables X and Y respectively. Which of the following statements is/are correct?

(I) If the correlation coefficient is positive, the gradient of the regression line will be positive.

(II) If the correlation coefficient is 0.9, the gradient of the regression line will be 0.9.

(A) Both (I) and (II)

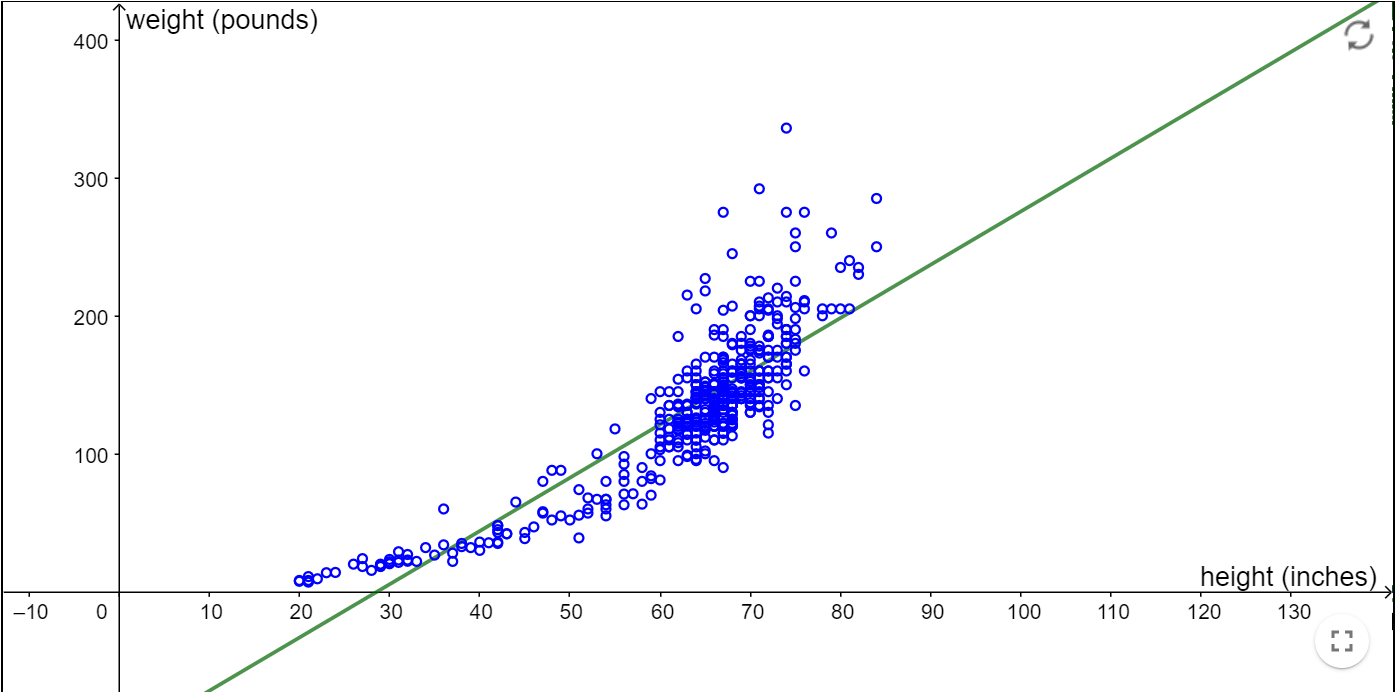
(B) Only (I)

(C) Only (II)

(D) Neither (I) nor (II)

*Explanation: See Chapter 2 Slides 25 and 90. The correlation coefficient only tells the direction and strength of the relationship. The gradient generally does not have the same value as the correlation coefficient.*

1. A researcher examined the relationship between weight (y axis) and height (x axis) among 475 male subjects. He graphed the relationship in the scatter diagram below. Weight is measured in pounds, and height in inches.



The equation of the regression line is y = 3.86\*x – 110.42.

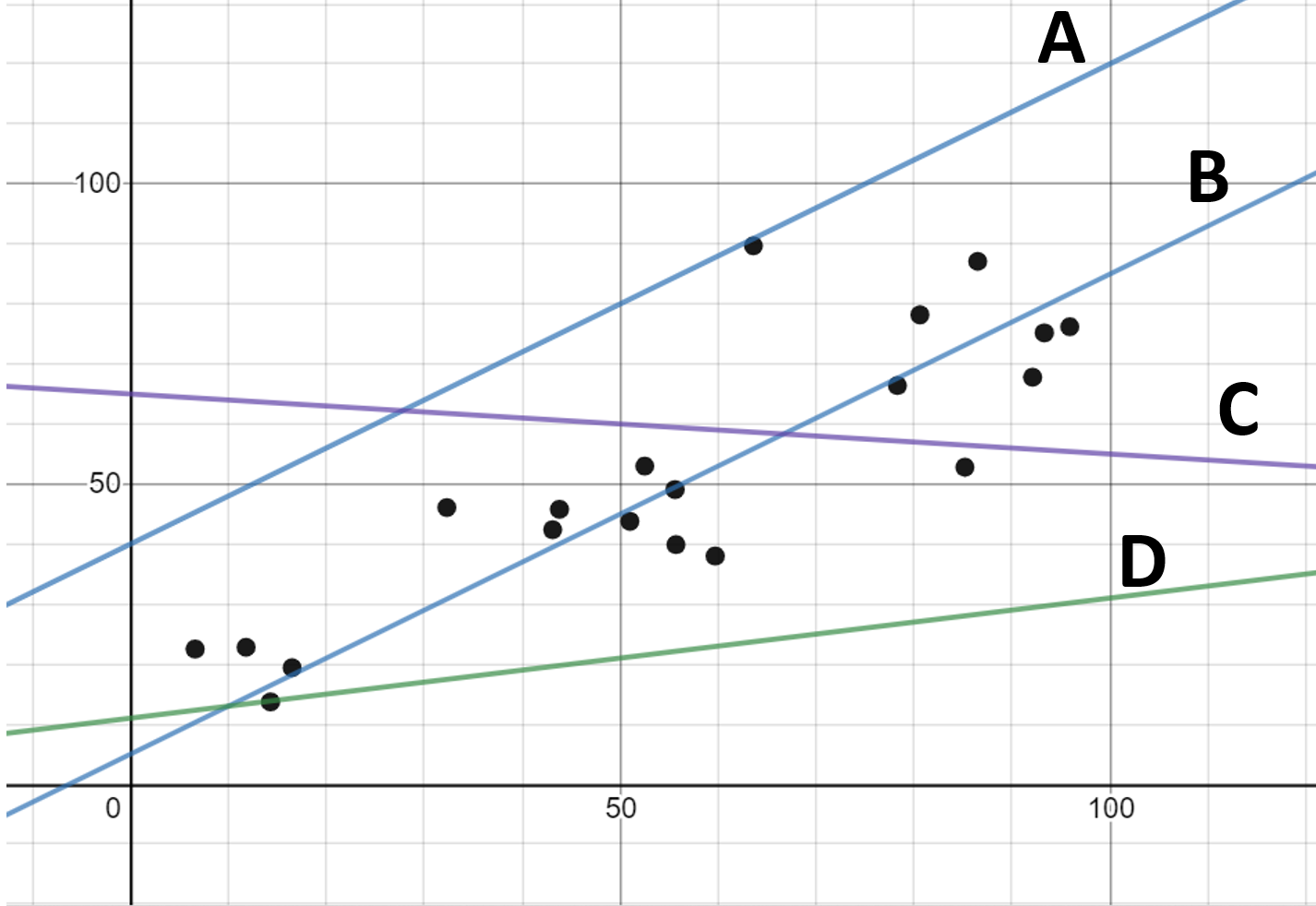
Is it reasonable to presume that if a male is 107 inches tall, his weight will be 302.6 pounds?

(A) Yes

(B) No

*Explanation: See Chapter 2 Slide 100. Any prediction of the weight given a height that is beyond the range of data is unreliable. If a larger range of data is collected, the best-fit regression line may actually change, and maybe a line may not even be the best way to describe the data. A male who is 107 inches tall is way beyond the usual human height and it would not be wise to predict the weight of such a male with the regression line alone. – this is actually one of the world’s tallest male heights recorded.*

1. A researcher examined the relationship between Variables X and Y among 20 male subjects, and he graphed a scatter plot as seen below. One of the lines (A, B, C, or D) is the regression line for predicting Y from X. Which one is it?



(A) A

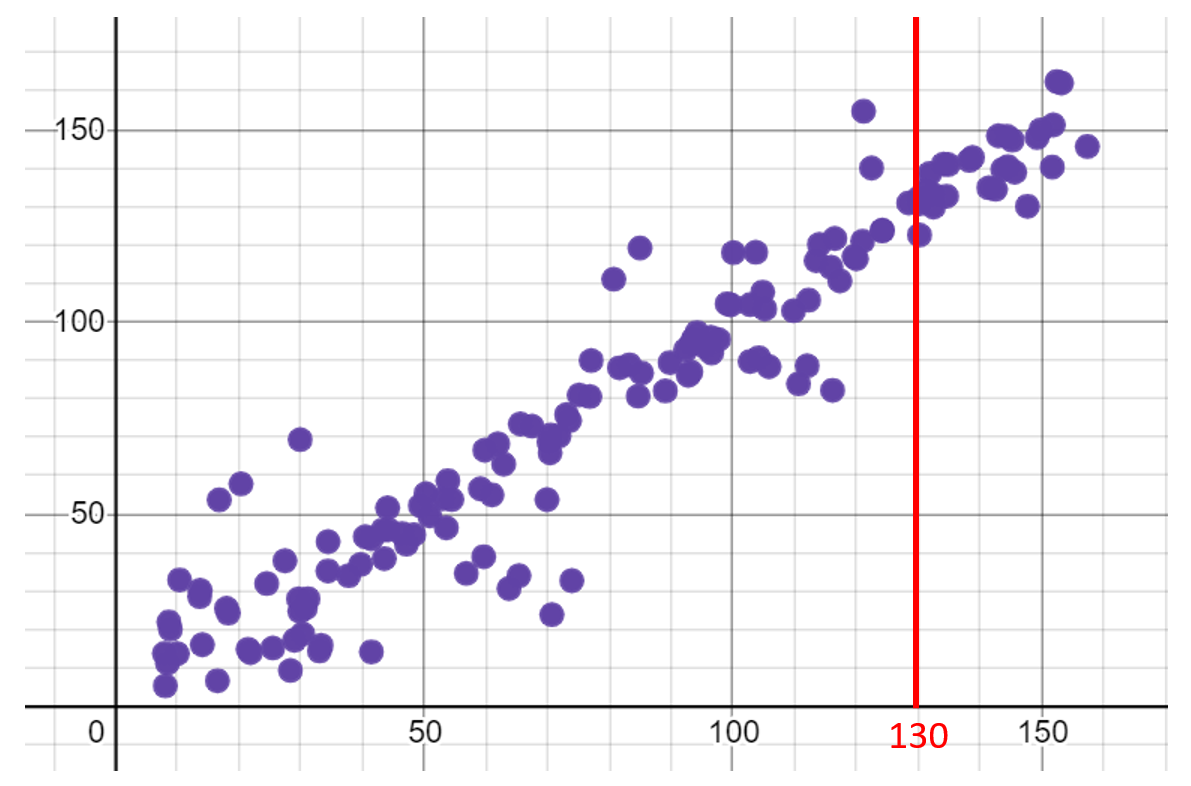
(B) B

(C) C

(D) D

*Explanation: See Chapter 2, Slides 25 and 98. The regression line is the line that is closest to all the points in the y direction on average.*

1. A researcher examined the relationship between Variables X and Y among 150 male subjects, and he graphed a scatter plot as seen below. The correlation coefficient for all the 150 data points is about 0.5. Let K be the correlation coefficient for the data points with X values lying between 130 to 150. Which of the following statements is correct?



(A) K is more than 0.50.

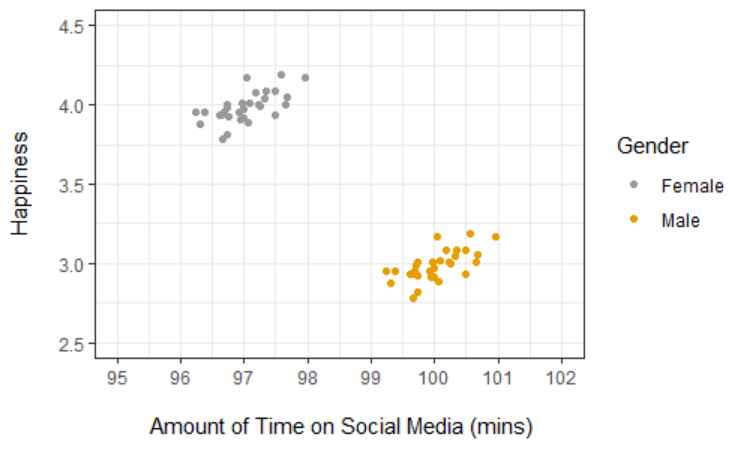
(B) K is less than 0.50.

(C) None of the other options.

*Explanation: See Chapter 2 Slide 78. Attenuation effect has occurred.*

1. A researcher is interested in the correlation between the amount of time an individual spends on social media and the individual’s level of happiness. Suppose that she observed that the correlation coefficient (r) for males only is 0.8, and that r for females only is also 0.8. Which of the following statements must be true for r for males and females combined?
2. 0 ≤ r < 0.8
3. r = 0.8
4. 0.8 < r ≤ 1
5. Not possible to tell

*Explanation. It is possible that the correlation coefficient in the combined dataset is negative, so that none of the other three options is correct.*



1. Which of the following statements is/are correct?

(I) Removal of outliers can increase the correlation coefficient.

(II) Removal of outliers can decrease the correlation coefficient.

(A) Both (I) and (II)

(B) Only (I)

(C) Only (II)

(D) Neither (I) nor (II)

*Explanation: See Chapter 2 Slides 61-63.*