Quiz 3:

1. A student has a set of 20 bivariate data, and calculated the correlation coefficient from it to be 0.5. After that, the student discovered that he had missed out an extra line of data. He has no time to re-calculate the correlation for the new set of 21 bivariate data, but he knows that the extra line of data has an x value which is larger than all other x values.

	X	Y
1		
2		
3		
...		
20		

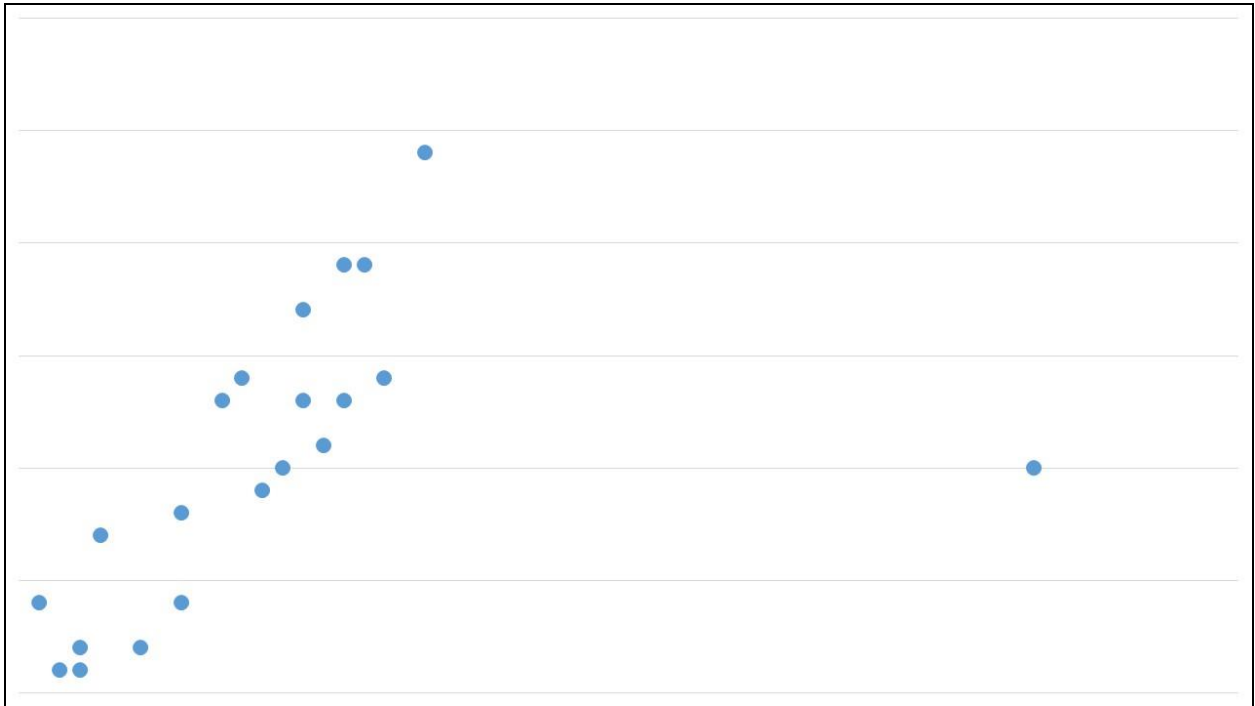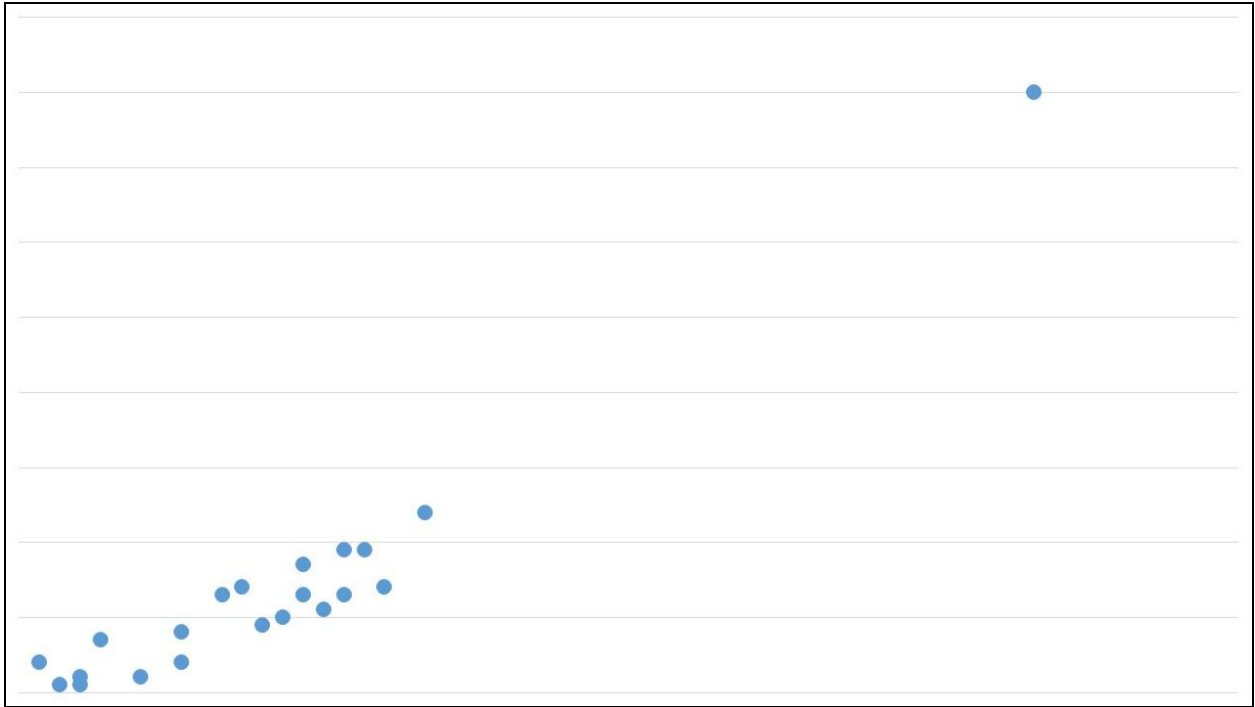| 21		

The correlation for the new set of 21 bivariate data...

A) is less than 0.5

B) is more than 0.5

C) is exactly 0.5

D) is unable to be estimated from the above information

Explanation:

D is the answer. There are many possible positions for the outlier. For example:

2. A large number of individuals take two IQ tests of similar difficulty. For both attempts, the average is 100, and the standard deviation is 15. The two scores are positively correlated, but not perfectly, so that the two scores for a typical individual are different. Consider the individuals who score 130 in the first test. Which of the following is a consequence of the regression effect? Their average score in the second test is around ___

    (A) 140
    (B) 120
    (C) 100
    (D) 80

Explanation:
B is the answer because regression effect states if the group of individuals has an extreme value on their first test, the second test will have a value closer to the average (Refer to Chapter 2 Unit 9).

3. Mr G attempted to find the correlation between the amount of time Singapore students spent at Co-Curricular Activities (CCA) and their marks. In 2015, he conducted the study on 200 schools in the Singapore. He recorded the "average number of CCA hours" and "average mark" from each school, and used it to calculate a correlation coefficient of r=0.6.

In his report, Mr G wrote "Based on the data I collected, I conclude that the correlation between the amount of CCA hours a student spends and his/her marks is also going to be around 0.6" in Singapore.

Which of the following may be present in this reasoning?

A) Atomistic fallacy

B) Ecological fallacy

C) Regression fallacy

Explanation:

B is the answer. The correlation coefficient observed for CCA hours and grades are both **for schools**. Mr G stated that based on the data, there is a moderate positive correlation between CCA hours and grades **for individual students**. This is not always true.

Refer to Chapter 2 unit 8.

4. The following table lists the number of hours to produce, cost and the type of 12 cakes from a bakery. Type S means "Shortened cakes" while U means "Un-Shortened cakes".

Type	U	U	U	S	S	U	S	S	S	U	S	U
Hours	2.5	2.5	2	5	2	3	3	4.5	2	1.5	3.5	2
Cost	30	35	30	25	20	25	30	35	25	35	20	30

Among the Un-Shortened cakes, the correlation coefficient between the hours and cost is closest to:

A) 0.6

B) 0.4

C) 0.2

D) -0.4

E) -0.6

Explanation:

D is the answer. Using Microsoft Excel to calculate r, r = -0.6333.

5. The scatter diagram comes from a study on 200 adults. The correlation between height and weight is 0.6. A student commented:

"Among those adults who are between 165 cm and 175 cm tall, the correlation between weight and height is 0.6."

What is wrong with the comment?

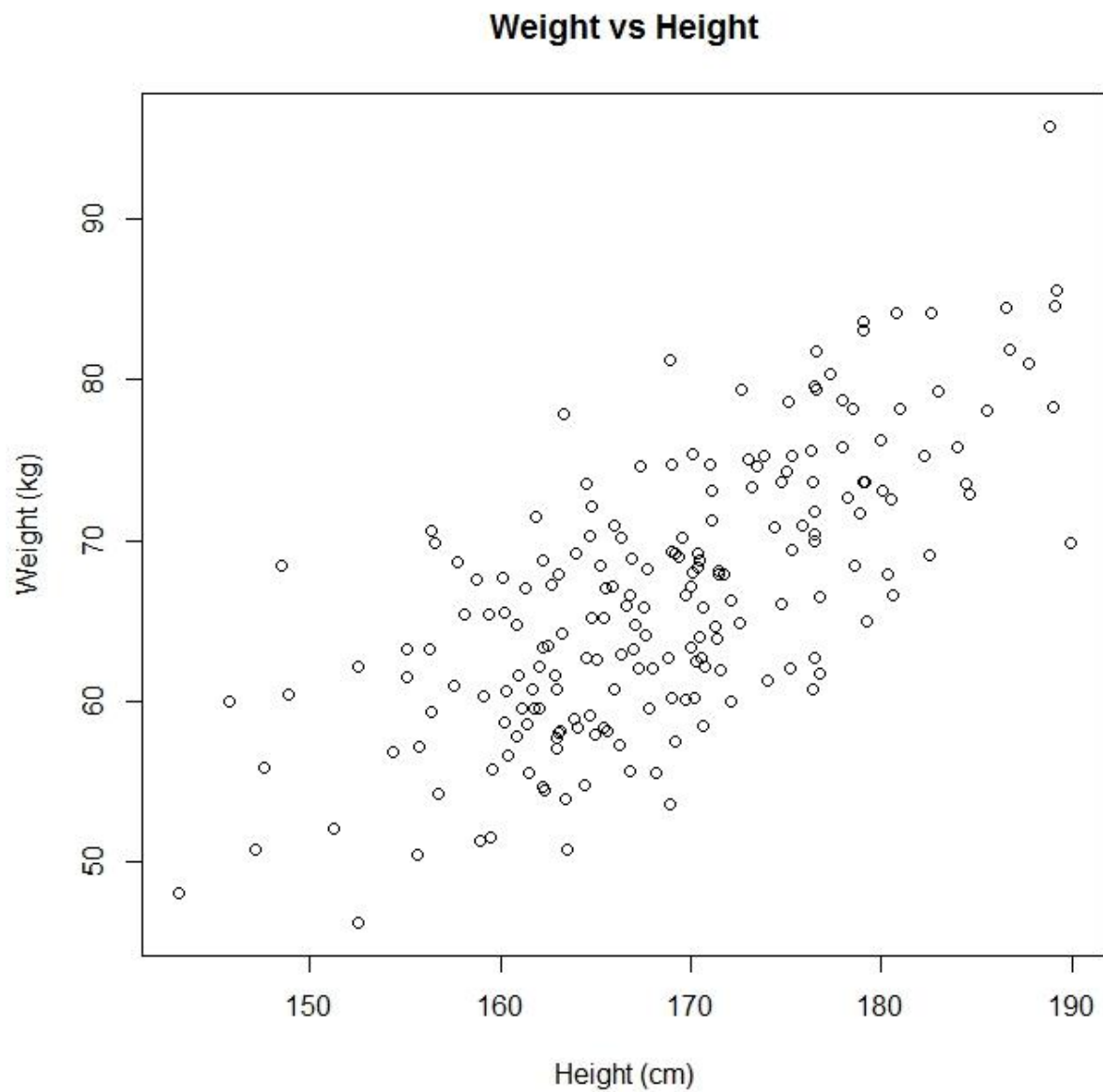I. The correlation should be between height and weight, not weight and height.

II. The correlation is less than 0.6.

(A) None of the above.
(B) Only (I).
(C) Only (II).
(D) Both (I) and (II).

Weight vs Height

Explanation:

C is the answer. Correlation is unchanged when swapping the variables (unit 6 slide 15). Restricting the range is likely to weaken the association, so the correlation among adults of height 165cm and 175cm should be less than 0.6 (Chapter 2 unit 9 slide 4).