Quiz 4:

1. A large number of students take a standardized test consisting of 10 multiple-choice questions. The number of correctly answered questions varies across students. Each student is presented in a scatter diagram, where the x-axis is the number of correct answers, and the y-axis is the number of wrong answers. The correlation coefficient is

(A) -1
(B) -0.5
(C) 0
(D) +0.5
(E) 1

Explanation:

Let X and Y be the number of correct and wrong answers respectively. Since there are 10 questions, 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.

2. The correlation between height and weight among men aged 18 to 74 in the US is about 0.4. Which of the following is true?

I. The correlation between weight and height among men aged 18 to 74 in the US is about 0.4.
II. If someone eats more and puts on 2 kg, he is quite likely to get somewhat taller.

(A) Neither I nor II.
(B) I only.
(C) II only.
(D) Both I and II.

Explanation:

Correlation is unchanged when swapping the variables (unit 6 slide 15), so [I] is true. [II] is false: correlation is not causation in general (unit 7). In this case, common sense tells us the height of an adult will hardly be affected by weight.

3. Last week, Sam ranked very highly on a Pokemon TCG Masters Division, having score 29 out of 30. This week, his score is 23, which is still above average, but his rank has dropped. Which of the following conclusions can reasonably be made by Sam?

I. Relative to other players in the Division, he has become worse.
II. The change in fortune can be explained by the regression effect: perhaps he was lucky last week to get such a high score, and this time a bit unlucky to fall behind.

(A) Neither I nor II.

(B) I only.
(C) II only.
(D) Both I and II.

Explanation:

One explanation of the decline in the score can always be becoming worse compared to others. However, this can be an example of regression effect also (unit 9 slide 14). If Sam thought only [I] must be true, he has committed the regression fallacy.

4. In 2012, the Organisation for Economic Cooperation and Development (OECD) conducted a study to test the impact of technology on students' grades, by measuring the **average ICT** (Information and Communication Technologies) **usage** in school **for each country** against the **average PISA** test scores amongst 15 year-olds of that country. Based on their findings, they observed that there is a strong negative correlation coefficient between average ICT usage and average PISA scores of countries, r = -0.7.

A newspaper wrote "The correlation between ICT usage and PISA score in schools is -0.7". Which of the following may be present in this reasoning?

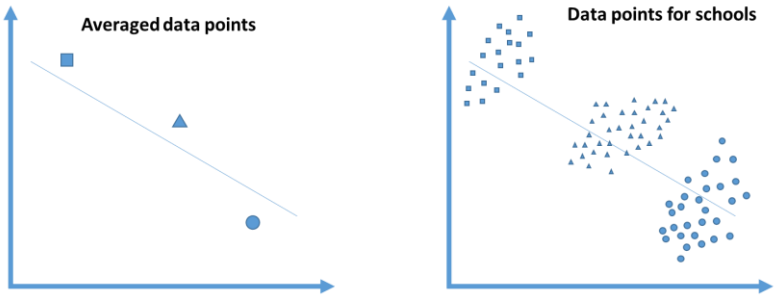
A) Atomistic fallacy

B) Ecological fallacy

C) Regression fallacy

Explanation:

The strong correlation coefficient observed is for the aggregated data as average ICT usage and PISA score are both averaged **for countries**. Newspaper article states that average ICT usage and average PISA scores **for schools** have a strong negative correlation. This is not always true. Figure below shows a hypothetical scatter plot of the same data when they are aggregated based on countries and when data of all schools are present:



From above illustration, one can see that same data when aggregated may yield to a stronger correlation coefficient. You may also refer to unit 8 of chapter 2.

5. The following table lists the number of pages, price and the type of 15 books. Type H means hardcover while S means softcover.

Pages	104	188	220	264	336	342	378	385	417	417	436	458	466	469	585
Price	32.95	24.95	49.95	79.95	4.5	49.95	4.95	5.99	4.95	39.75	5.95	60	49.95	5.99	5.95
Type	H	H	H	H	S	H	S	S	S	H	S	H	H	S	S

Among the Hard Cover books, the correlation coefficient between the price and number of pages is closest to:

A) 0.65

B) 0.35

C) 0.15

D) 0.85

Explanation:

Using Microsoft Excel to calculate r, r = 0.35.