GER1000 2018 Sem 2
Quiz 3 and solutions

1) A researcher wishes to examine the association between the number of hours students engage in gaming and their academic performance. He conducts his study in 15 schools, and calculates for each school the average number of hours students spend gaming and the average academic score of students. He notices that there is a correlation of -0.9 between the two sets of averages. He concludes that

"The correlation between gaming hours and academic score for all students from the 15 schools is -0.9".

What fallacy is the researcher committing here?

A) Atomistic Fallacy
B) Ecological Fallacy

The researcher is trying to obtain the correlation at the individual level solely from the correlation at the aggregate level. This is an example of ecological fallacy. See Unit 6 Slide 6 (pg 72 of script).

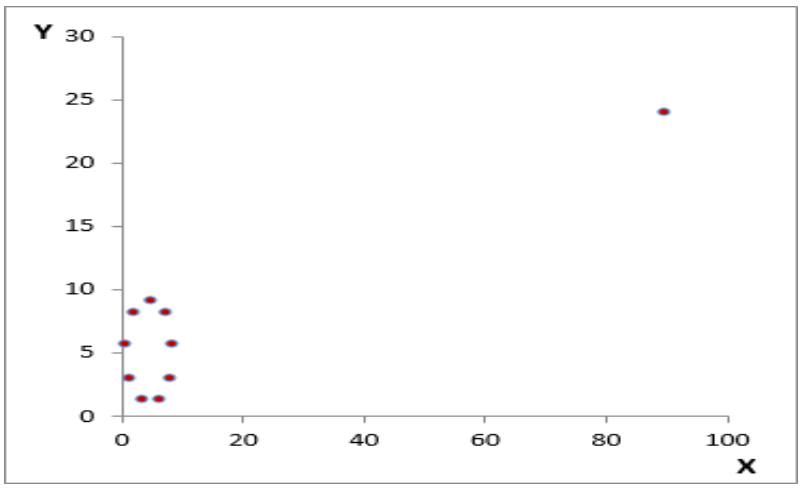
2) There is a weak positive association between numerical variables X and Y, where X ranges from 0 to 5 (inclusive). Based on the data from X and Y, the regression line is given by the equation $Y = 0.25X + 2$. Which of the following statement(s) is/are **definitely** true?

I) We can obtain the exact value of Y when $X = 4$.
II) The average value of Y is 4 when $X = 8$.
III) The correlation coefficient is 0.25.
IV) The equation provides a non-deterministic relationship between X and Y.

A) I only
B) II and IV only
C) III only
D) IV only

A weak association between 2 variables gives us a non-deterministic relationship since one particular X value can correspond to many Y values. See Unit 8 Slide 3 (pg 89 of script). Hence (IV) is true. The regression equation only gives us the average of the Y values for a given X value. See Unit 8 Slide 13 (pg 99 of script). Thus (I) is false. We cannot draw any conclusions of what values Y can take beyond the range of X values, hence (II) is false See Unit 8 Slide 14 (pg 100 of script). Lastly the correlation is not the gradient of the line in general, so (III) is also false. See Unit 4 Slide 11 (pg 48 of script) and Unit 8 Slide 11 (pg 97 of script).

3) The following plot shows an outlier in both the X and Y directions.



What will happen if we remove the outlier?

(A) The correlation between X and Y will remain the same.
(B) The correlation between X and Y will decrease
(C) The correlation between X and Y will increase.
(D) It is not possible to know what will happen to the correlation between X and Y.

See Unit 5 Slide 9 and 10 (pg 62 and 63 of script) to see how removal of outliers can affect the correlation coefficient.

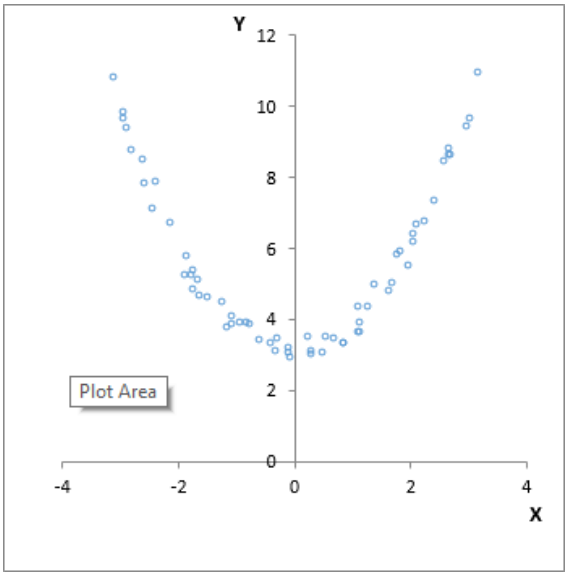
4) Consider two tutorial classes A and B each containing 25 students. Both classes enrolled in an NUS module and sat for two tests: the midterm and the final. For tutorial class A, every student scored 5 more marks for the final compared to the midterm. For tutorial class B, every student scored 3 marks less for the final compared to the midterm. Which of the following statements is/are true?

I) For class A, the correlation between midterm and final is 1.
II) For class B the correlation between midterm and final is -1.
III) The overall correlation (combining class A and B together) between midterm and final would be less than the correlation for Class A.
IV) The overall correlation (combining class A and B together) between midterm and final would be more than the correlation for Class B.

A) I only
B) I and II
C) I and III
D) All of the above

Let X denote the midterm marks and let Y denote the final marks.
For tutorial class A, Y = X + 5, with perfect correlation: r = 1.
For tutorial class B, Y = X – 3, again with perfect correlation: r = 1.
When combining the 2 classes, we are going to get 2 parallel lines of points, therefore the overall correlation has to drop since the combined points cannot all be passing through a single straight line. Hence (I) and (III) are true.

5) The following scatterplot shows data collected from variables X and Y.
Which of the following statements is the **best** description of the plot?



(A) The correlation between X and Y is nearly 0.
(B) The correlation between X and Y is nearly 0.5.
(C) The correlation between X and Y is nearly 0.9.
(D) The plot strongly suggests that X is caused by Y.

See Unit 3 Slide 6 (pg 29 of script)

6) The correlation for 126 pairs of (X,Y) values is 0.8. X ranges from 0 to 1. For the points with X values lying between 0.25 and 0.75, the correlation is 0.9. Is this an example of the attenuation effect?

(A) Yes
(B) No

Attenuation effect is observed when the correlation coefficient is smaller in magnitude due to restricting the range of one of the variables. See Unit 7 Slide 5 (pg 79 of script). In the above scenario, the correlation coefficient INCREASES in magnitude and therefore this cannot be described as attenuation effect.