**GER1000 2020 Sem 1**

**Quiz 2 and solutions**

1. In a study of 100 mother-daughter pairs, their heights were measured and plotted in a scatter diagram - mothers’ heights at the horizontal axis, and their respective daughters’ heights at the vertical axis. All the data points are above the 45-degree line passing through the origin. This means that

(A) Every daughter is taller than her mother in the dataset.

(B) Every daughter is shorter than her mother in the dataset.

(C) None of the other options.

*Explanation: See Chapter 2 Slide 20.*

2. In a study of 100 mother-daughter pairs, their heights were measured and plotted in a scatter diagram - mothers’ heights at the horizontal axis, and their respective daughters’ heights at the vertical axis. All the data points are above the 45-degree line passing through the origin. This means that the correlation between mothers’ height and daughters’ height is

(A) negative.

(B) positive.

(C) zero.

(D) There is insufficient information to choose any of the other options.

*Explanation: See Chapter 2 Slide 20. Just because the scatterplot points are above the 45-degree line does not imply that the correlation must be positive/negative/zero.*

3. There are 50 students in class and the average (or mean) number of pens per student is 4. Not all students in this class have the same number of pens. For a student with 4 pens, what is the standard unit for the number of pens for this student?

(A) -1

(B) 0

(C) 1

(D) The standard deviation is required to answer this question.

*Explanation: See Chapter 2 Slide 40. Note that standard deviation is non-zero here, since not all students in this class have the same number of pens: there is a spread around the mean 4.*

4. Which of the following correlation coefficients suggests the strongest linear association between two variables in the data set?

(A) -0.9

(B) -0.3

(C) 0.65

(D) 1.3

*Explanation: See Chapter 2 Slide 32. The correlation coefficient always lies between -1 and 1. The closer it is to 1 or -1, the stronger the linear association between the 2 variables. Hence, the strength of linear association is indicated by the magnitude of the correlation coefficient, i.e. ignoring whether it is positive or negative.*

5. In a study of 100 mother-daughter pairs, their heights were measured and plotted in a scatter diagram - mothers’ heights at the horizontal axis, and their respective daughter’s heights at the vertical axis. The current scale to measure height is in centimeters. A researcher then converted all height of mothers and daughters into inches. The statement “The correlation coefficient stays the same” is

(A) True

(B) False

*Explanation: See Chapter 2 Slide 52.*

6. The commuting times (in minutes) and daily screen times (in minutes) of 150 working adults were collected, and the correlation coefficient *r* was calculated. Which of the following statement(s) is/are correct?

(I) The unit of the standard deviation of the commuting times is in minutes.

(II) The unit of *r* is in minutes.

(A) Both (I) and (II)

(B) Only (I)

(C) Only (II)

(D) Neither (I) nor (II)

*Explanation: See Chapter 2 Slide 39 & 52.*

7. Use Microsoft Excel to calculate the correlation coefficient for X and Y in the data set below. (All NUS students can use the free online version of Microsoft Excel by signing in with their NUS email address on [www.office.com](http://www.office.com). The offline version is also available for free.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 49 | 73 | 38 | 87 | 15 | 39 |
| Y | 3 | 52 | 96 | 75 | 3 | 66 |

Which of the following numbers is closest to the correlation coefficient?

(A) -1

(B) -0.5

(C) 0

(D) 0.5

(E) 1

*Explanation: See Chapter 2 Slide 44 to 48. The correlation coefficient is 0.386, which is closest to 0.5. For your own learning, try to plot a scatter plot for this data set, and also swap the variables and observe the same correlation coefficient, like in Slide 51.*

8. Consider 50 points in a scatter diagram, and for every point (X,Y), we have Y = 2 – 5X. Which of the following statement(s) is/are correct?

(I) The correlation coefficient is 1.

(II) There is a perfect negative correlation between X and Y.

(A) Both (I) and (II)

(B) Only (I)

(C) Only (II)

(D) Neither (I) nor (II)

*Explanation: See Chapter 2 Slide 31. You can try a few points that follow this deterministic relationship to convince yourself that the correlation coefficient should be -1. All the data points must lie on the line Y = 2 – 5X*.

9. Pick the option that fills in the respective blanks in the sentence: “The \_\_\_\_\_\_\_\_ is used to measure how widely spread the data points are about its \_\_\_\_\_\_\_\_\_\_.”

(A) Correlation coefficient, mean

(B) Mean, correlation coefficient

(C) Standard deviation, mean

(D) Mean, standard deviation

*Explanation: See Chapter 2 Slide 16.*