GER1000 2018 Sem 2
Quiz 2 and solutions

1. Which of the following is true?

The correlation coefficient…
(i) ranges between 0 to 1.
(ii) describes the probability of a linear association being present in the data.
(iii) describes how closely the data points cluster around a line.

(A) (i) only
(B) (ii) and (iii)
(C) (iii) only
(D) (i), (ii), and (iii)
(D) None of the above

See Unit 3 Slide 3 (pg 26 of script).
The correlation coefficient ranges between -1 to 1. The coefficient describes the strength and direction of the linear association between two variables. The correlation coefficient describes how close the data points are around a line - see Unit 3 Slide 9.

2. Three father-son pairs had their heights measured. The following table shows their heights:

Pair	Father (inches)	Son (inches)
A	68	72
B	70	71
C	66	70

Using these three data points, standard deviation for the fathers would be 1.6, and for the sons it would be 0.8.

From the table, what is the standard unit for the son from pair A?

(A) -1.25
(B) 0
(C) 1.25
(D) 1.88
(E) None of the above

See Unit 4 Slide 3 (pg 40 of script). The sons' average = 71. SU = (72-71)/0.8 =1.25

3. Five father-son pairs had their heights measured. The following table shows their heights.

Father (inches)	Son (inches)
68	72
67	70
70	68
72	72
68	69

Find the correlation coefficient. (Hint: You can use Excel)
The correlation coefficient lies within the range:

(A) -1.0 to -0.5
(B) -0.5 to -0.1
(C) -0.1 to 0.1
(D) 0.1 to 0.5
(E) 0.5 to 1.0

See Unit 4 Slide 7 (pg 44 of script). Input the table into Excel, and use the function "CORREL". r = 0.2.

4. A student is studying the relationship between height of teenagers (in inches), and their reported amount of physical activity done in a usual week (in minutes). He gathered the data by surveying 100 teenagers, and computed a correlation coefficient of r = 0.7.

Which of the following situation/s will definitely not change the correlation coefficient?

(i) He realizes an error in the measuring ruler, and adds 2 inches to each of the teenagers' data.
(ii) He changes his height data from inches into centimeters.
(iii) 10 of the teenagers had incomplete survey forms and he removed them from the data set.

(A) (i) only
(B) (ii) only
(C) (iii) only
(D) Both (i) and (ii)
(E) All of the above

See Unit 4 Slide 15 (pg 52 of script). (i) is adding a number to all values of a variable. (ii) is multiplying a positive number to all values of a variable. These will not affect the correlation coefficient. In (iii), by removing the data, it is necessary to recalculate the correlation coefficient.

5. In country X, a **two-year** military national service is compulsory for all male citizens. In the year 2000 onwards, the graduating enlistees were given an option to 'sign-on' and continue their military service as a career (i.e., the 1998 enlistees were the first batch to be given the option when they graduated in 2000). A study was done on country X males to find out if there is an association between the number of enlistees in a batch and the number of them who choose to sign on **two years later**.

The following table shows the number of enlistees and sign-ons each year:

	A	B	C
1	Year	Number of enlistees (thousands)	Number of sign-ons
2	1998	100	-
3	1999	110	-
4	2000	100	700
5	2001	90	800
6	2002	95	900
7	2003	90	1000
8	2004	95	900

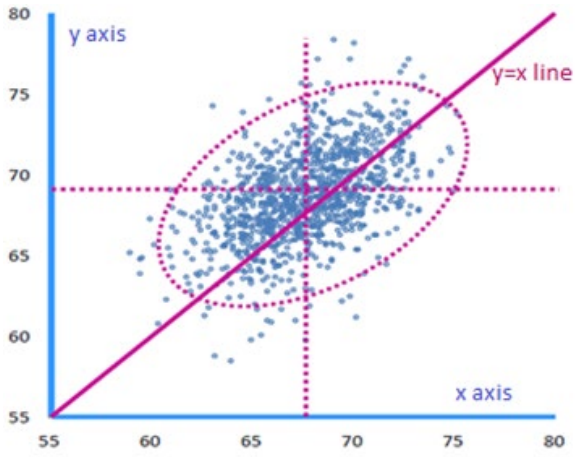
Using only the data from the table, what is the most suitable way to calculate the correlation coefficient for the study?

(A) CORREL(A4:A8,B4:B8)
(B) CORREL(C4:C8,B4:B8)
(C) CORREL(B4:B8,C4:C8)
(D) CORREL(B2:B8,C4:C8)
(E) CORREL(B2:B6,C4:C8)

See Unit 4 Slide 10 (pg 47 of script). Input for the CORREL function requires 2 arrays of the same size. For the 1998 batch, they only graduated in 2000, hence the data in B2 should be paired with C4. B3 should be paired with C5, etc.

6. In the scatter diagram shown below, the dotted straight lines mark the average values of X and Y. From the diagram, we can say that

(A) The Y=X line cuts through the data points in half, with 50% of the data points on either side of the line.
(B) The average of Y is larger than the average of X.
(C) The average of X is larger than the average of Y.



The intersection of average lines of X and Y lie above the Y=X line. This shows that the average of Y variable is higher than the average of X. The Y=X line does not cut through the data points in half. Many more points are above the line than below it. See Unit 2 Slide 10 (pg 20 of script).