There are four scatter plots labelled C5, C6, C7, and C8. Which of the following is true for the above scatter plots?

1. A) C5 has r close to -1, C6 has r close to 1 and the remaining two have r close to 0
   B) C5 has r close to -1, C6 has r close to 1, C7 has r close to .6 and C8 has r close to 0
   C) C5 has r close to 1, C6 has r close to -1, C7 has r close to .6 and C8 has r close to 0
   D) C5 has r close to 1, C6 has r close to -1 the remaining two have r close to 0
The mean height of American women is about 63.6 inches and the standard deviation is about 1.7 inches. The mean height of American men is about 67.6 inches with a standard deviation of 3.3 inches. The correlation between the heights of husbands and wives is about 0.5. Suppose we transform the data to centimeters, instead of inches (1 inch is approximately 2.3 cms). Then which of the following statements are correct:

A. correlation will remain the same for both the original data and the transformed data
B. The slope of the regression line will remain the same for both sets
C. The intercept of the regression line will remain the same for both sets

2. A ☐ A and C
   B ☐ None of them
   C ☐ A and B
   D ☐ A, B and C
   E ☐ B and C
Michigan State University researchers want to investigate how rainfall affects the yield of crops in East Lansing. The researchers found that the mean amount of rainfall is about 250 inches and the standard deviation is about 20 inches. The mean yield of crops in East Lansing is about 300 tonnes with a standard deviation of 45 tonnes. The correlation between the amount of rainfall and yield of crops is about 0.4.

2 pt  The slope of the regression line of yield of crop on amount of rainfall is

3. A 0.18
   B 196.00
   C 0.90
   D 75.00

2 pt  The intercept of the appropriate regression line is

4. A 75.00
   B 0.90
   C 0.18
   D 196.00
5) We will now look at some transformation properties. We transform a random variable $X$ to a variable $N=aX+b$, where $a>0$. Let,

$Z_N$ and $Z_X$ denote the Z scores of $N$ and $X$ respectively. 
$\bar{N}$ denotes the mean of $N$ 
$S_N$ denotes the standard deviation of $N$ 

Study the following example where we calculate $Z_N$. We know that the formula for calculating Z-score is as follows:

$$Z_N = \frac{N-\bar{N}}{S_N} = \frac{aX+b-(a\bar{X}+b)}{|a|S_X} = \frac{aX-a\bar{X}}{|a|S_X} = \frac{aX-a\bar{X}}{aS_X} = \frac{X-\bar{X}}{S_X} = Z_X$$

Here $a>0$ so $|a|=a$

a) Now we look at the following transformation

$M=-X+b$, i.e now $a=-1$

Using above steps as reference show that (4 points)

$Z_M = -Z_X$
Thus from this exercise we can see that
\[ Z_{aX+b} = \text{sign}(a)Z_X \]

Now we turn to correlation \( N=aX+b, \ a>0 \)
The correlation formula is given as:
\[
\text{Corr}(X, Y) = \frac{\sum_{i=1}^{n} Z_{x_i}Z_{y_i}}{n - 1}
\]
The following calculation tell us that correlation between \( N \) and \( Y \) is the same as correlation between \( X \) and \( Y \).
\[
\text{Corr}(N, Y) = \frac{\sum_{i=1}^{n} Z_{n_i}Z_{y_i}}{n - 1} = \frac{\sum_{i=1}^{n} Z_{x_i}Z_{y_i}}{n - 1} = \text{Corr}(X, Y)
\]
(Here we use the fact that \( Z_X = Z_N \) that we proved earlier)

b) \( M=-X+b \). Now using above correlation calculation as reference
Show that \( \text{corr}(M, Y) = -\text{corr}(X, Y) \) (4 points)
Thus we see that $\text{Corr}(aX + b, Y) = \text{sign}(a)\text{Corr}(X, Y)$

Now we look at the correlation of the variable with itself

\[
\text{Corr}(X, X) = \frac{\sum_{i=1}^{n} Z_{x_i} Z_{x_i}}{n - 1} = \left( \frac{1}{n - 1} \right) \sum_{i=1}^{n} Z_{x_i}^2
\]

\[
= \left( \frac{1}{n - 1} \right) \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{S_x} \right)^2 = \frac{1}{S_x^2} \left( \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) = \frac{1}{S_x^2} (S_x^2) = 1
\]

c) Referring to the above calculations show that $\text{Corr}(X, -X) = -1$ (4points)

Hint:

\[
\text{Corr}(X, -X) = \frac{1}{n - 1} \sum_{i=1}^{n} Z_{x_i} Z_{-x_i} = \left( \frac{1}{n - 1} \right) \sum_{i=1}^{n} Z_{x_i} Z_{-x_i}
\]

\[
= \left( \frac{1}{n - 1} \right) \sum_{i=1}^{n} \left( \frac{x_i - \bar{x}}{S_x} \right) \left( \frac{-x_i - \text{blank}}{\text{blank}} \right)
\]
Suppose in an exam there are 10 questions. For any student let \( X \) be the number of correct answers and let \( Y \) be the number of incorrect answers. Then the correlation \((r)\) between \( X \) and \( Y \) is:

7. A\( \bigcirc \) 1  
   B\( \bigcirc \) -1  
   C\( \bigcirc \) 0  
   D\( \bigcirc \) Not enough information to find \( r \)

(hint: draw picture or use \( Y=10-X \) and above properties )