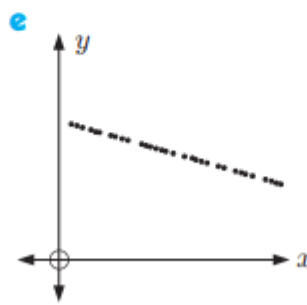
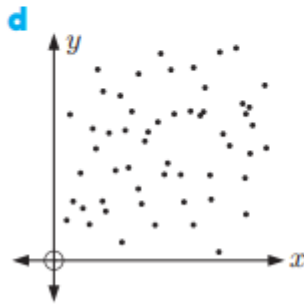
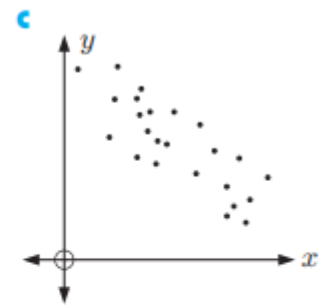
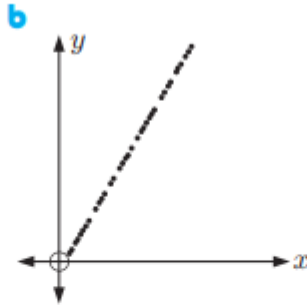
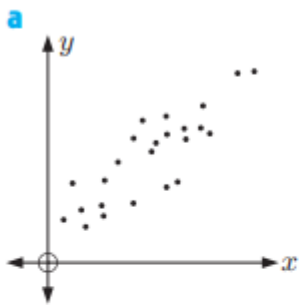


# Correlation Co-efficient and Co-efficient of Determination Questions

1 Match each scatter diagram with the correct value of  $r$ .



**A**  $r = 1$

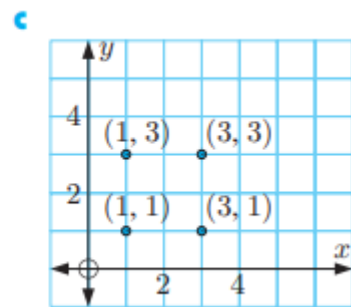
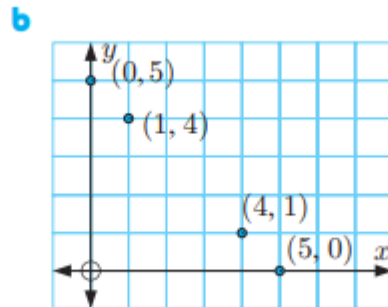
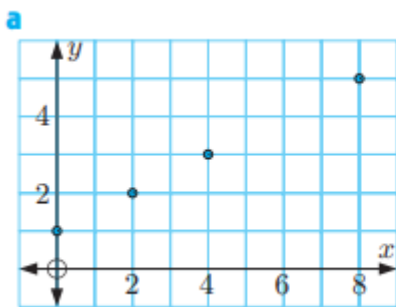
**B**  $r = 0.6$

**C**  $r = 0$

**D**  $r = -0.7$

**E**  $r = -1$

2. Looking at the points shown, estimate the value of  $r$ .



3. Calculate the proportion of the variance of Y which

(a) can be predicted from (or explained by) the variance of X if

- i.  $r = 0.8$       ii.  $r = -0.9$

(b) cannot be predicted from (or explained by) the variance of X if

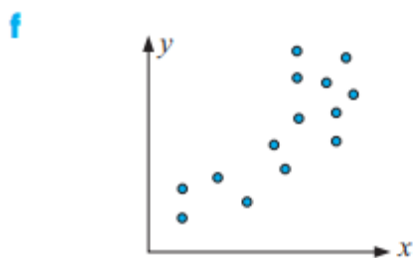
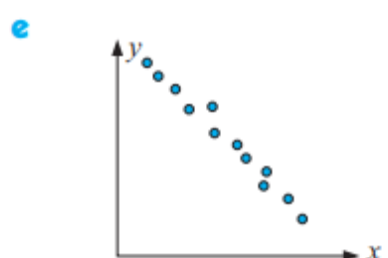
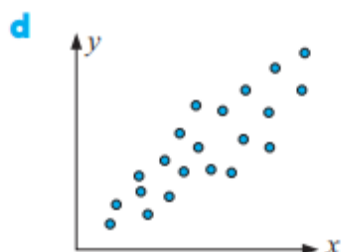
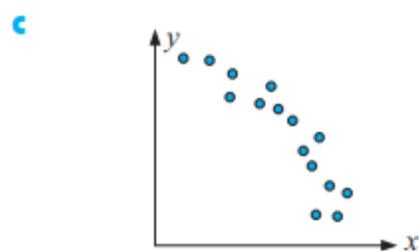
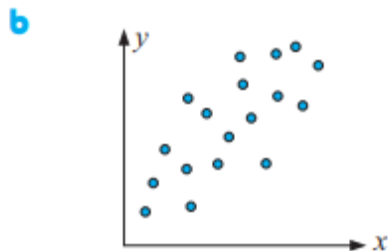
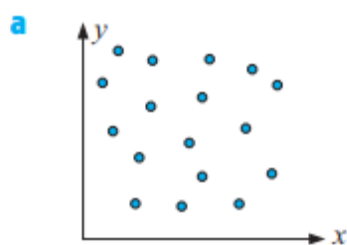
- i.  $r = 0.7$       ii.  $r = -0.6$

4. How many times is a difference in predictive capacity between correlations of 0.7 and 0.8 greater than between correlations of 0.2 and 0.3?

5. What correlation between X and Y is required in order to state that 85% of the variance of X depends on the variance of Y?

6 For the following scatterplots comment on:

- i the existence of any *pattern* (positive, negative or no association)
- ii the relationship *strength* (zero, weak, moderate or strong)
- iii whether or not the relationship is linear
- iv whether or not there are any outliers.



7 What is meant by causation?

The following are multiple choice questions. Select the best answer.

8. A research study has reported that there is a correlation of  $r = -0.59$  between the eye colour (blue, green, brown) of an experimental animal and the amount of nicotine that is fatal to the animal when consumed. This indicates:
- a) Nicotine is less harmful to one eye colour than the others.
  - b) The lethal dose of nicotine goes down as the eye colour of the animal changes
  - c) One must always consider the eye colour of animals in making statements about the effect of nicotine consumption
  - d) The researchers need to do further study to explain the causes of this negative correlation.
  - e) The researchers need to take a course in statistics because correlation is not an appropriate measure of association in this situation.
9. The correlation co-efficient provides:
- a) A measure of the extent to which changes in one variable cause changes in another variable.
  - b) A measure of the strength of the linear association between two categorical variables.
  - c) A measure of the strength of association (not necessarily linear) between two categorical variables.

- d) A measure of the strength of the linear association between two quantitative variables.
- e) A measure of the strength of the linear association between a quantitative variable and a categorical variable.

10. On May 11<sup>th</sup>, 50 randomly selected subjects had their systolic blood pressure (SBP) recorded twice – the first time at about 9:00 am and the second time at about 2:00 pm. If one were to examine the relationship between the morning and the afternoon readings, then one might expect:

- a) The correlation to be near zero, as the morning and afternoon readings would be independent of one another.
- b) The correlation to be high and positive, as those with relatively high readings in the morning will tend to have relatively high readings in the afternoon.
- c) The correlation to be high and negative, as those with relatively high readings in the morning will tend to have relatively low readings in the afternoon.
- d) The correlation to be near zero, as the correlation measures the strength of the linear association.
- e) The correlation to be near zero, as blood pressure readings should follow a bell curve.

## Solutions

1 a B      b A      c D      d C      e E

2 a  $r = 1$       b  $r = -1$       c  $r = 0$

3. a. i. 64% ii. 81% b. i. 51% ii. 64%

4. 3 times

5. 0.922

- 6 a i no association ii zero iii non-linear  
iv no outliers
- b i positive association ii weak iii roughly linear  
iv no outliers
- c i negative association ii strong iii non-linear  
iv no outliers
- d i positive association ii moderate iii linear  
iv no outliers
- e i negative association ii strong iii linear  
iv no outliers
- f i positive association ii moderate iii non-linear  
iv no outliers

7. Causation is that one thing is CAUSED by another. Eg. Smoking causes cancer. It is very different than correlation – that two things increase together. Eg. Icecream and murder rate. To prove that you have causation, and not merely correlation, you need to use PPDAC. Especially, your study needs to try to have sufficient replication, random assignment, and random sampling.

8. e

9. d

10. b