Summary

- Correlation is the degree to which two or more quantities are linearly associated. In a two-dimensional plot, the degree of correlation between the values on the two axes is quantified by the correlation coefficient.
- It is the scale-less measure of the relationship between X and Y. It is obtained by divided the covariance by the product of the two standard deviations.
- Causation means 'cause and effect' relationship. No discussion of correlation would be complete without a discussion of causation. It is possible for two variables to be related (correlated), but not have one variable cause another.
- The following points are the accepted guidelines for interpreting the correlation coefficient:
    - 0 indicates no linear relationship.
    - +1 indicates a perfect positive linear relationship
    - -1 indicates a perfect negative linear relationship
- The correlation coefficient requires that the underlying relationship between the two variables under consideration is linear. If the relationship is known to be linear, or the observed pattern between the two variables appears to be linear, then the correlation coefficient provides a reliable measure of the strength of the linear relationship. If the relationship is known to be nonlinear, or the observed pattern appears to be nonlinear, then the correlation coefficient is not useful
- Following are the properties of correlation coefficient
    - Correlation coefficient is independent of unit of measurement of the variables.
    - If correlation is present, then coefficient of correlation would lie between ±1.
    - Correlation coefficient is independent of change of origin and scale.
    - If X and Y are random variables and a, b, c, d are any numbers provided that $a \neq 0, c \neq 0$ then $r(aX + b, cY + d) = \frac{ac}{|ac|} r(X, Y)$
- The variables X and Y are connected by the equation aX+bY+c=0. Then, the correlation between them is -1 if the signs of a and b are alike and +1 if they are different.