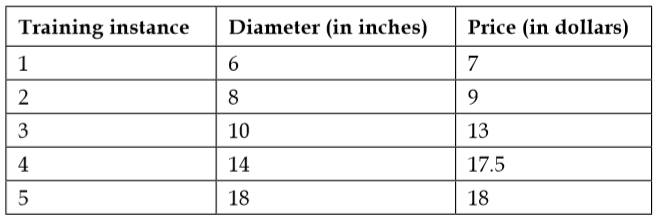
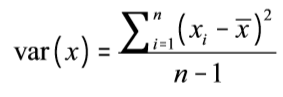
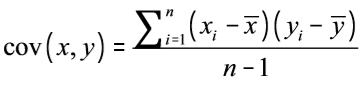
**Question 1 (Linear regression):**

Consider the following training dataset having the price of a pizza based on its diameter:

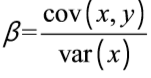
**Table 1:** Pizza price based on diameter of pizza



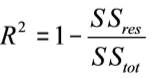
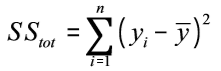
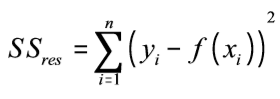
Find the regression line y=α+βx using least square method. To do so, you will calculate the variance of x and covariance of x and y, as follows:

Using the following formula, α and β can be calculated for the regression line:

Using a python program, estimate the regression line. Plot the dataset and regression line using matplotlib. To evaluate your model, find the R2 value using the following formula:

Interpret the effectiveness of your model using your python program.

**Hint:**

1. An r-squared score of one indicates that the response variable (y) can be predicted without any error using the model.
2. An r-squared score of one half indicates that half of the variance in the response variable can be predicted using the model.

**Question 2 (Clustering)**

Generate two clusters of dataset of random numbers using numpy uniform distribution as follows:

C1 between 0.5 and 1.5 of size (2, 10)

C2 between 3.5 and 4.5 of size (2, 10)

Stack the dataset into a third numpy array and then plot it. Find the centroid (mean) of each cluster and draw on the same plot.