Data Analytics

D Samanta

Topic 1:

In the given data, we are provided with three columns. First one being the serial number, second the speed and the third column being the distance travelled. Observing the data we can some of the data points as outliers.

When we observe the data we can deduce that the observations are taken such that the time taken is constant and the speed and distance are varied thoroughly. We have learned about the different significance of Arithmetic, Geometric and Harmonic Mean. All are the arithmetic mean in a sense if the variable is transformed in other ways. In this particular data we have varying distance and time so that we can consider the following relations:

Suppose that for nn instances we travel a fixed time tt at velocities v1,…,vn over distances d1,…,dn. Now, we want the total distance conserved. We have

di−vit=0,

Since the relation between the d (distance) and v (speed) is a proportion relation we can conclude the type of mean we can use. We shall be using the Arithmetic mean. From the given data, the following were the results:

Speed Average = 15.4

Distance Average = 42.98

Observing the data we can see some outliers in some data point. So we could have also used the FivePoint Summary evaluation along with ITR method of outlier detection and removal.