Tecnológico Nacional de México
Instituto Tecnológico de Tijuana

Subdirección Académica

Departamento de sistemas y computación

SEMESTRE:
ENERO - JUNIO 2020

Carrera:
Ingeniería en Tecnologías de la Información y Comunicaciones

Datos Masivos

Tarea 2 Variance en scala

Unidad 1

Alumno:
Zamorano García Osvaldo Arturo 15211707
Barraza Sierra Alexis Fernando 16212532

Profesor:
Dr. Jose Christian Romero Hernandez

The variance or variance is a measure of the dispersion of a random variable (values that are required randomly). The variability of said dispersion is specifically specific in the area of express statistics. The variance of a sample or a set of values is the sum of the squared deviations from the average or the average, all this divided by the total number of observations minus 1. In a very general way it can be said that the variance is the standard deviation squared.

The variance, together with the standard deviation, are measures of data dispersion or observations. The dispersion of these data indicates the variety that they present, that is, if all the values in a set of data are equal, then there is no dispersion, but instead, if not all are equal then there is dispersion.

This dispersion can be large or small, depending on how close the average values are. he variance of a sample is symbolized as S2, while the variance of a population symbolizes as σ2. he variance of a sample is used to estimate the variance of a population, which is often unknown. This is why S2 is also commonly considered as a statistic and σ2 as a parameter.

The variance of a sample has the following formula:

$$\frac{\sum_{i=1}^{n}(xi-\overline{x})^2}{n-1}$$

Where, represents the sum of the subtraction between each of the sampled values () and the mean (), squared.

In turn, it represents the total number of observations or sampled data. For very large values the variance is minimal or even negligible.