Problem statement -

Perform the following operations on any open source dataset. (e.g. data.csv)

1. Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative variable. For example, if your categorial variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorial variable.
2. write a python program to display some basic statistical details like percentile, mean, standard deviation etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris-versicolor' of iris.csv dataset.

Objective -

theory -

statistics -

It provide and summarize information about your sample data. summary statistics are used to summarize a set of observation, in order to cummunicate the largest amount of information as simply as possible.

Here, some operations that perform on the data. stastics Information we will get.

1. Mean -

It called as the arithmetic mean or average.

simple mean - $\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n}$

population mean - $u = \frac{\sum_{i=0}^{n} x_i}{N}$

2. Median -

Is tells you where the middle of a data set.

if n is odd - median = $\left(\frac{n+1}{2}\right)^{th}$

if n is even -

median = $\frac{(n/2)^{th} + (n/2+1)^{th}}{2}$

3. Minimum -

The minimum is the smallest value in the data set.

4. Maximum

The maximum is the largest value in the data set.

5. standard deviation

Is the measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the value tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

population S.D. = $\sqrt{\frac{1}{N} \sum (x-u)^2} = \sigma$

sample S.D = s = $\sqrt{\frac{1}{n-1} \sum (x-\bar{x})^2}$

6. Percentile -

a percentile is a term that describes how a score compares to other scores from the same set

$$P_x = \frac{x(n+1)}{100}$$

Result -

In this way we have learn about statistical operations . on the data set.