Assignment -3

Roll : 31444   DOC : 27/1/22
Name : Mehul Oswal   Dos : 1/2/22

TITLE :

Descriptive Statistics - Measure of central tendency and variability

PROBLEM STATEMENT :

Perform following operations on any open source dataset (provided nba.csy , iris.csv)

① Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income, etc) with numeric values grouped by one of the qualitative (categorical variable). Create a list that contains a numeric value for each response to the categorical variable

② Write 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

LEARNING OBJECTIVE :

1) Understand Descriptive Statistics
2) Learn and apply the steps to measure central tendency and variability.

THEORY

What is central tendency?

→ A measure of central tendency is a summary statistics that represents the center point or typical value of a dataset. These measure indicates where most values in a distribution falls and are also refered to as the central location of a distribution

Basic Statistic Terms:

① **Mean :** Finding an average is often called 'mean' value.
Generally,

$$mean = \frac{sum}{count\ (n)} = \frac{\sum f(N)}{N}$$

② **Median :** Can be defined as the middle observation in a set of data.
Generally,

For odd : → The number in the middle

For even : → Average of the two middle obs.

③ **Mode :** Most frequent observation

④ **Min, Max and Range**

⑤ Range : Numerical indication of the span of our data.

⑥ Standard deviation : One way of measuring the dispersion of a given set of data. It is calculated by the square root of the average squared deviation

Analysis and Approach :

Part 1 :
① Import modules (pandas, numpy)
② Read the CSV
③ Preprocessing of the data
④ Group the data on the basis of age
⑤ Paint statistics based on age

Part 2 :
① Import modules
② Read the csv
③ Preprocessing of data
④ Group data based on species
⑤ Paint Statistics based on species

Conclusion
Understood the measures of central tendency and variability and ways to implement them on the dataframe.