Omkar Gaikwad
31126

Page No.
Date

Assignment 3

Title: Descriptive statistics measure of central tendency and variance

Problem statement:

Perform following on nba & iris dataset

1. Provide summary statistics (mean, std, min, max etc) for a dataset (age, income etc) with numeric variables grouped by categorical variables for eg. if categorical column is team then you may group other numeric columns by team. then provide summary on grouped data.
2. Create a python program to describe statistical data about iris dataset grouped by species column. provide output & explain everything you do.

Objective:

1. provide statistical summary for sample data.
2. To identify some basic statistical details about some column.

Outcome

Student will be able to:

1. Calculate mean, median, min, max, std dev etc
2. Identify basic statistical details of some data.

H/w and S/w requirements:

64-bit Windows OS, 8GB RAM

python, jupyter notebook.

Page No.
Date

Theory:

central tendancy

value that gives/ summarizes some sample data. concerns with average of values

Variability:

study of spread of data points about it's central values is called as variability

Measure of central tendency

1. Mean: defined as average of data points it is calaulated as sum of data by the total number of data points.

$$\bar{x} = \frac{\sum x_i}{N}$$

2. Median: central value of data observations

$x_1, x_2, \underline{x_3}, x_4, x_5$

3. Mode: data value occuring maximum number of times

4. measure of variability

Range The lowest & highest values of a data sample.

5. standard deviation

defined as the root of average of difference of datapoints from mean's square.

$$\sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

Page No
Date

Analysis of datasets

nba.csv

- player who play in center(C) have highest salary.
- max salary paid to small forward (SF)
- the mean age of players is almost similar (27 years)
- maximum & minimum age of players in NBA is 40 & 19
- players in age group 30-39 have maximum mean salary and in age group 10-19 are given less salary.

iris.csv

- average values of iris-setosa is more only in sepal-width in comparison to other varities although avg sepalW of versicolor > virginica, the avg petalLW of vesicolor < virginica

Conclusion:

Students learned descriptive analysis, Measures of central tendency and variability by performing various operations of given dataset using pandas.