

5. a. Use the diabetes data set from UCI and Pima Indians Diabetes data set for performing the following:

Univariate analysis: Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis.

Aim:

Analysis the various univariate functions like Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis on dataset like Pima Indian diabetes dataset.

Procedure

- Download dataset like Pima Indian diabetes dataset. Save them in any drive and call them for process.
- The `mean()` function can be used to calculate mean/average of a given list of numbers.
- The `median()` method calculates the median (middle value) of the given data set.
- The mode of a set of data values is the value that appears most often.
- The `var()` method calculates the variance for each column.
- Standard deviation `std()` is a number that describes how spread out the values are.
- The `skew()` method calculates the skew for each column. Skewness refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data.
- **Kurtosis:**
- It is also a statistical term and an important characteristic of frequency distribution. It determines whether a distribution is heavy-tailed in respect of the normal distribution. It provides information about the shape of a frequency distribution.

Program:

```
import pandas as pd
from scipy.stats import kurtosis
import pylab as p
df = pd.read_csv (r'd:\\diabetes.csv')
print (df)
df1 = pd.DataFrame(df, columns= ['Age','Glucose'])
print (df1)
df1.mean()
df1.median()
df1.mode()
print(df1.var())
df1.std()
print(df1.skew())
print(kurtosis(df, axis=0, bias=True))
```

Dataset download link

<https://github.com/npradaschnor/Pima-Indians-Diabetes-Dataset/blob/master/Pima%20Indians%20Diabetes%20Dataset.ipynb>

Result:

The various univariate functions like Frequency, Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis on dataset Pima Indian diabetes dataset are successfully executed.