Graphical Representation on Parkinson's Patient Experiment

IPE-205 Poster Submission

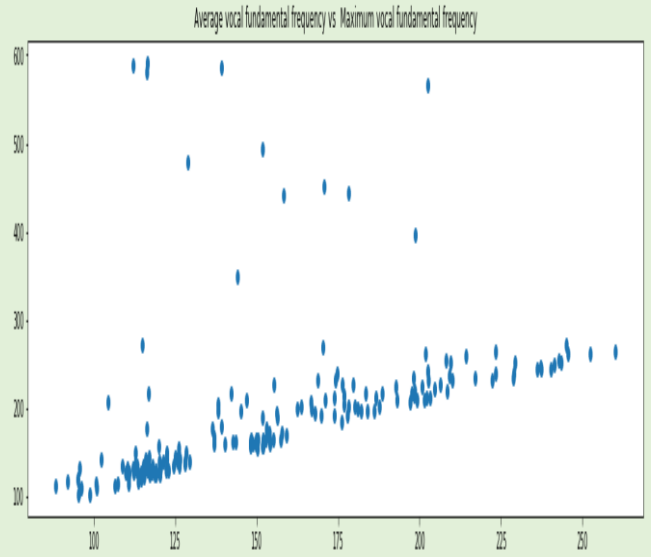
Submitted To:
Assistant Professor Tanmoy Das
Submitted By :
201736021,201736022,
201736027,201736047

Datasets

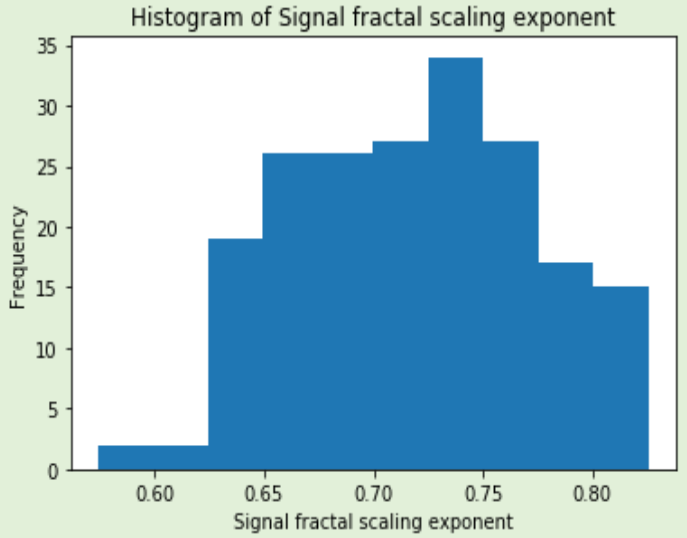
name	MDVP:Fo(H	MDVP:Fhi(I	MDVP:Flo(I	MDVP:Jitte	MDVP:Jitte	MDVP:RAP	MDVP:PPQ	Jitter:DDP
phon_R01_	119.992	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109
phon_R01_	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394
phon_R01_	116.682	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633
phon_R01_	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505
phon_R01_	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966
phon_R01_	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388
phon_R01_	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466
phon_R01_	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431
phon_R01_	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088
phon_R01_	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803
phon_R01_	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763
phon_R01_	91.904	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844
phon_R01_	136.926	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355
phon_R01_	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496
phon_R01_	152.845	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364
phon_R01_	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471
phon_R01_	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632
phon_R01_	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853
phon_R01_	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092
phon_R01_	156.405	189.398	142.822	0.00768	0.00005	0.00372	0.00399	0.01116
phon_R01_	153.848	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285
phon_R01_	153.88	172.86	78.128	0.0048	0.00003	0.00232	0.00267	0.00696
phon_R01_	167.93	193.221	79.068	0.00442	0.00003	0.0022	0.00247	0.00661

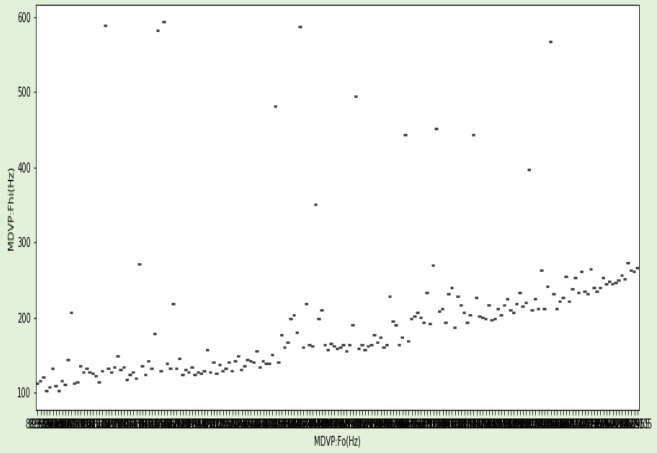
Python Coding

```
import numpy as np # linear algebra
import pandas as pd
import os
print(os.listdir("../input"))
import matplotlib.pyplot as plt
import plotly.plotly as plty
import plotly.graph_objs as go
import seaborn as sns
from sklearn import datasets, linear_model
dataframe = pd.read_csv("../input/parkinsons.data.csv")
print("....Draw a scatter plot:Average vocal fundamental frequency vs  Health status of the subject....")
df = pd.read_csv("../input/parkinsons.data.csv")
print("....Draw a scatter plot:Average vocal fundamental frequency vs  Maximum vocal fundamental frequency....")
fig = plt.figure(figsize=(15,3))
axes = fig.add_axes([0,0,1,1])
x1 = df['MDVP:Fo(Hz)']
y1 = df['MDVP:Fhi(Hz)']
plt.scatter(x1,y1)
plt.title('Average vocal fundamental frequency vs  Maximum vocal fundamental frequency')
plt.show
import statsmodels.api as sm
fig = plt.figure(1, figsize=(15, 6))
ax = fig.add_subplot(111)
sns.boxplot(ax=ax, x=df['MDVP:Fo(Hz)'],y=df['MDVP:Fhi(Hz)'],data=df)
plt.hist(df['DFA'])
plt.xlabel('Signal fractal scaling exponent')
plt.ylabel('Frequency')
plt.title('Histogram of Signal fractal scaling exponent')
from sklearn.model_selection import train_test_split

train, test = train_test_split(df, test_size=0.28, random_state=99)

xtrain = train[['DFA']]
ytrain = train[['MDVP:Fo(Hz)']]

xtest = test[['DFA']]
ytest = test[['MDVP:Fo(Hz)']]
from sklearn import linear_model
regr = linear_model.LinearRegression(fit_intercept=False)
regr.fit(xtrain, ytrain)
y_pred = regr.predict(xtest)
plt.scatter(xtest, ytest,  color='grey')
plt.plot(xtest, y_pred, color='yellow', linewidth=3)
plt.xlabel("DFA")
plt.ylabel("MDVP:Fo(Hz)")
plt.show()
```

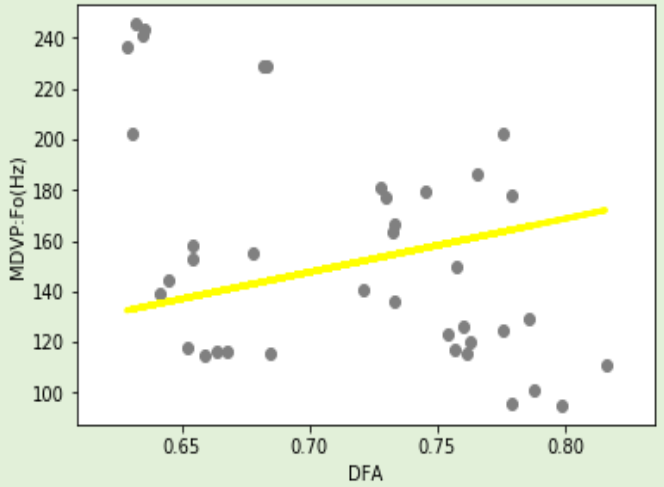


Scatter Diagram





Boxplot



Regression analysis

Summery of the datasets

This dataset is composed of a range of a biomedical voice measurements of 31 people,23 with Parkinson's disease. Each column in the table is a particular voice measure and each row corresponds one of 195 voice recording from these individuals ("name" column). The main aim of the data is to discriminate healthy people from those with PD, according to "status" column which is set to 0 for healthy and 1 for PD.

Reference

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders.