**TEST DATA**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sno | name | sex | bmi | totalsleeptime | apneascount | hypopneascount | ahtotalcount | apneasindex | hypopneasindex | ahi |
| 1 | Rukum Bai | F | 38.4 | 247.5 | 66 | 69 | 135 | 16 | 16.7 | 32.7 |
| 2 | N.Sanjay | M | 24.09 | 245 | 22 | 120 | 142 | 5.4 | 29.4 | 34.8 |
| 3 | G.Swapna | F | 29.67 | 387.5 | 1 | 14 | 15 | 0.2 | 2.2 | 2.3 |
| 4 | N.venkateswarulu | M | 28.4 | 336 | 297 | 197 | 494 | 53 | 35.2 | 82.2 |
| 5 | E.Vivek kumar | M | 37.1 | 320.5 | 43 | 311 | 354 | 8 | 52.8 | 66.3 |
| 6 | Charanjeet singh | M | 46.44 | 390.5 | 213 | 269 | 482 | 32.7 | 41.3 | 74.1 |
| 7 | Devesh devangan | M | 30.02 | 421.5 | 345 | 34 | 379 | 49.1 | 4.8 | 54 |
| 8 | G.S.S.Rajasekhar | M | 19.84 | 222.5 | 1 | 1 | 2 | 0.3 | 0.3 | 0.5 |
| 9 | George | M | 27.34 | 322.5 | 63 | 148 | 211 | 11.7 | 27.5 | 39.3 |
| 10 | mohammed aadam | M | 38.42 | 229 | 166 | 139 | 305 | 43.5 | 36.4 | 79.9 |
| 11 | N.sanjeev reddy | M | 29.75 | 405.5 | 83 | 235 | 318 | 12.3 | 34.8 | 47.1 |
| 12 | P.S.shastri | M | 37.09 | 214.5 | 199 | 75 | 274 | 55.7 | 21 | 76.6 |
| 13 | Ramachandra Rao | M | 25.95 | 375 | 36 | 62 | 98 | 5.8 | 9.9 | 15.7 |
| 14 | S.Mallesh | M | 31.64 | 294 | 14 | 63 | 77 | 2.9 | 12.9 | 15.7 |
| 15 | Satyanayana reddy | M | 33.65 | 324.5 | 111 | 212 | 323 | 20.5 | 39.2 | 59.7 |
| 16 | P.Rajesh | M | 32.48 | 380 | 168 | 209 | 377 | 26.5 | 33 | 59.5 |
| 17 | Y.janakiram | M | 30.46 | 305.5 | 375 | 28 | 403 | 73.6 | 5.5 | 79.1 |
| 18 | Vijaya Bharathi.B | F | 38.54 | 290 | 0 | 34 | 34 | 0 | 7 | 7 |
| 19 | Devamani medoju | F | 31.64 | 301 | 90 | 182 | 272 | 17.9 | 36.3 | 54.2 |
| 20 | J.Uma sundari | F | 55.62 | 310 | 53 | 29 | 82 | 10.3 | 5.6 | 15.9 |
| 21 | ammaji saladi | M | 28.51 | 388.5 | 105 | 175 | 280 | 16.2 | 27 | 43.2 |
| 22 | Arpita Jalui | F | 45.01 | 258.5 | 0 | 39 | 39 | 0 | 9.1 | 9.1 |
| 23 | D.L.tulasamma | F | 30.85 | 399.5 | 75 | 130 | 205 | 11.3 | 19.5 | 30.8 |
| 24 | D.Bhanumathi | F | 33.98 | 277.5 | 358 | 302 | 660 | 77.4 | 65.3 | 142.7 |
| 25 | Jana Symala | F | 24.67 | 362 | 5 | 24 | 29 | 0.8 | 4 | 4.8 |
| 26 | K.V.Suseela | F | 33.74 | 221 | 67 | 183 | 250 | 18.2 | 49.7 | 67.9 |
| 27 | M.Venkata Laxmi | F | 37.46 | 262.5 | 74 | 117 | 191 | 16.9 | 26.7 | 43.7 |
| 28 | M.V.Lakshmi | F | 32 | 371 | 234 | 95 | 329 | 37.8 | 15.4 | 53.2 |
| 29 | Mani Manchmshetty | F | 28.39 | 290.5 | 210 | 228 | 438 | 43.4 | 47.1 | 90.5 |
| 30 | S.Shanwaz Begum | F | 38.28 | 317.5 | 300 | 147 | 447 | 56.7 | 27.8 | 84.5 |
| 31 | V.Nagamani | F | 36.75 | 271 | 206 | 370 | 576 | 45.6 | 81.9 | 127.5 |
| 32 | Rama Murthy | M | 43.89 | 162 | 7 | 173 | 180 | 2.6 | 64.1 | 66.7 |
| 33 | M.Bhagyashree | F | 33.2 | 464.5 | 23 | 69 | 92 | 3 | 8.9 | 11.9 |
| 34 | Sridevi | F | 31.14 | 395.5 | 32 | 48 | 80 | 10 | 15 | 25 |
| 35 | Laxmana Charry | M | 29.1 | 316 | 9 | 62 | 71 | 1.7 | 11.8 | 13.5 |
| 36 | Mahboob Ali | M | 26.72 | 398 | 135 | 353 | 488 | 20.4 | 53.2 | 73.6 |
| 37 | Mahamed Farooq | M | 25 | 410.5 | 92 | 95 | 187 | 13.4 | 13.9 | 27.3 |
| 38 | Ratan Lal Agarwal | M | 29.8 | 405.5 | 123 | 133 | 256 | 18.2 | 19.7 | 37.9 |
| 39 | S.Sambaiah | M | 31.4 | 401.5 | 33 | 136 | 169 | 4.9 | 20.3 | 25.3 |
| 40 | T.Venkat Narsimhulu | M | 23.9 | 269.5 | 35 | 81 | 116 | 7.8 | 18 | 25.8 |
| 41 | G.Nagendra | M | 22.9 | 426 | 3 | 38 | 41 | 0.4 | 5.4 | 5.8 |
| 42 | Kannakaiah | M | 24.39 | 338 | 127 | 132 | 259 | 22.5 | 23.4 | 46 |
| 43 | Kantha Rao Chadalavada | M | 32.41 | 366.5 | 60 | 109 | 169 | 9.8 | 17.8 | 27.7 |
| 44 | Venkateshwar reddy | M | 32.6773 | 354.5 | 16 | 78 | 94 | 2.7 | 13.2 | 15.9 |
| 45 | Giridhar | M | 23.3 | 282.5 | 65 | 108 | 173 | 13.8 | 22.9 | 36.7 |
| 46 | Rajaiah | M | 30.5 | 204.5 | 56 | 78 | 134 | 16.4 | 22.9 | 39.3 |
| 47 | chandra sekhar | M | 32.6773 | 237 | 306 | 39 | 345 | 91.8 | 11.7 | 103.5 |
| 48 | . K. Ramakrishna | M | 39.3 | 298 | 25 | 55 | 80 | 5.6 | 12.4 | 18 |
| 49 | Aisha Banu | F | 15.9 | 283 | 8 | 57 | 65 | 2.3 | 16.4 | 18.8 |
| 50 | Aruna | F | 41.66 | 303.3 | 24 | 283 | 307 | 5.1 | 60.4 | 65.6 |
| 51 | Rajanna | M | 31.2 | 244 | 31 | 92 | 123 | 8.1 | 24.1 | 32.2 |
| 52 | G. Narender | M | 42.7 | 270 | 45 | 152 | 197 | 11.8 | 40 | 51.8 |
| 53 | M. Sunil | M | 33.3 | 240 | 69 | 51 | 120 | 24.4 | 18 | 42.4 |
| 54 | Anantha Laxmi | F | 26.2 | 226 | 140 | 193 | 333 | 37.2 | 51.2 | 88.4 |
| 55 | Adi lakshmi | F | 28.13 | 155.5 | 48 | 24 | 72 | 18.5 | 9.3 | 27.8 |
| 56 | P.S.shekhar | M | 38.28 | 193 | 54 | 98 | 152 | 16.8 | 30.5 | 47.3 |
| 57 | N.Prasad Rao | M | 28.22 | 71 | 5 | 57 | 62 | 4.2 | 48.2 | 52.4 |
| 58 | B.Narsimulu | M | 25.46 | 213 | 306 | 29 | 335 | 86.2 | 8.2 | 94.4 |
| 59 | D.Hari Shankar | M | 28.39 | 182.5 | 110 | 18 | 178 | 52.6 | 5.9 | 58.5 |
| 60 | G.S.Ashok | M | 31.23 | 121.5 | 32 | 173 | 205 | 15.8 | 85.4 | 101.2 |
| 61 | Moirza Ashaq Ali | M | 34.08 | 199.5 | 268 | 16 | 284 | 80.6 | 4.8 | 85.4 |
| 62 | Syed Mohammad Zakiullah | M | 28.71 | 192.5 | 6 | 151 | 157 | 1.9 | 47.1 | 48.9 |
| 63 | Fariyad | F | 30.11 | 189.5 | 238 | 34 | 272 | 75.4 | 10.8 | 86.1 |
| 64 | K.Swasanthra | F | 33.05 | 213.5 | 55 | 247 | 302 | 15.5 | 69.4 | 84.9 |
| 65 | krishna bai | F | 45.16 | 165 | 390 | 36 | 426 | 141.8 | 13.1 | 154.9 |
| 66 | M.ratna sri | F | 41.55 | 141 | 94 | 81 | 175 | 40 | 34.5 | 74.5 |
| 67 | B.rama devi | F | 32.46 | 160.5 | 136 | 42 | 178 | 50.8 | 15.7 | 66.5 |
| 68 | Sayeed | M | 32.6773 | 150.5 | 23 | 65 | 88 | 9.2 | 25.9 | 35.1 |
| 69 | C.H.prasad | M | 30.5 | 184.5 | 27 | 92 | 119 | 8.8 | 29.9 | 38.7 |
| 70 | B.Vishvam | M | 32.6773 | 201 | 278 | 22 | 300 | 83 | 6.6 | 89.6 |
| 71 | chandrasekhar | M | 32.6773 | 200 | 306 | 39 | 345 | 91.8 | 11.7 | 103.5 |
| 72 | G.srinivas | M | 32.6773 | 194.5 | 59 | 100 | 159 | 18.2 | 30.8 | 49 |
| 73 | Mandapati ravi | M | 36.57 | 228.5 | 282 | 11 | 293 | 74 | 2.9 | 76.9 |
| 74 | Lathifa begum | F | 44 | 198 | 52 | 124 | 176 | 15.8 | 37.6 | 53.3 |
| 75 | seema | F | 26.7 | 212.5 | 34 | 30 | 64 | 9.6 | 8.5 | 18.1 |
| 76 | Mona | F | 32.6773 | 219.5 | 72 | 44 | 116 | 19.7 | 12 | 31.7 |
| 77 | nalini verma | F | 26.3 | 232 | 258 | 90 | 348 | 66.7 | 23.3 | 90 |
| 78 | C.srinivas | M | 24.22 | 426 | 12 | 12 | 24 | 1.7 | 1.7 | 3.4 |
| 79 | SK.Samir Uddin | M | 28.71 | 487 | 0 | 3 | 3 | 0 | 0.6 | 0.6 |
| 80 | Sanjay agarwal | M | 44 | 456 | 50 | 153 | 203 | 9.4 | 28.9 | 38.3 |
| 81 | Nikita | F | 30.11 | 435 | 17 | 66 | 83 | 3.4 | 13.3 | 16.8 |
| 82 | V.Venkamma | F | 32.45 | 312 | 84 | 90 | 174 | 26.3 | 28.1 | 54.4 |
| 83 | Md.Miraj Ahmed | M | 28.34 | 257 | 366 | 15 | 381 | 100.7 | 4.1 | 104.9 |
| 84 | Subba Rao Aduri | M | 46.77 | 307 | 316 | 15 | 331 | 82.3 | 3.9 | 86.2 |
| 85 | B.Nirmala | F | 51.93 | 203 | 101 | 91 | 192 | 41.2 | 37.1 | 78.4 |
| 86 | Kamalamma | F | 31.99 | 425 | 7 | 2 | 9 | 1.5 | 0.4 | 1.9 |

**DATA MODEL**

#!/usr/bin/env python

# coding: utf-8

# In[1]:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import scipy.stats as stats

import statsmodels.formula.api as smf

import statsmodels.api as sm

import pylab

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

# In[2]:

snore = pd.read\_csv("C:\\Users\\kumar\\snore12.csv")

# In[6]:

snore

# In[7]:

x = snore.iloc[:, :-1].values ## independent variables

y=snore.iloc[:,10] ## dpendent variables

# # exploratory data analysis

# In[8]:

# 1st business moment decision---central tendency

snore.mean()

# In[9]:

snore.median()

# In[10]:

snore.describe()

# In[11]:

# # 2-->Variance,Standard Deviation and Range:

# In[226]:

print("Variances:")

print("-------------------------------")

print(np.var(snore))

print("\nStandard Deviation:")

print("-------------------------------")

print(np.std(snore))

# In[9]:

""" we know data of any distribution lies between mean+sd and mean-sd

we can analyse where most data is distributed using variance and mean

"""

# In[10]:

snore.columns

# #### 3rd business movement decision skewness

# In[12]:

for c in snore.columns:

if c == "name" or c=="sno" or c=='sex':

continue

else:

print(c+": "+ str(stats.skew(snore[c])))

# In[12]:

"""we obseve thatskwness is +ve mean>median

"""

**# #### 4th business movement decison --- kurtosis**

# In[13]:

for c in snore.columns:

if c == "name" or c=="sno" or c=='sex':

continue

else:

print(c+": "+ str(stats.kurtosis(snore[c])))

# In[15]:

""" kutosis is less than 3 ,so less outliers,less standard deviation ,data set is perfect or

we can consider this daata set"""

# In[14]:

### distributions based on sex

# In[13]:

for c in snore.columns:

if c == "name" or c=="sno" or c=='sex':

continue

else:

sns.FacetGrid(snore,hue="sex",height=5).map(sns.distplot,c).add\_legend()

plt.show()

# ##### BOX PLOTS TO SHOW OUTLIERS AS IT IS CONTINOUS DATA

# In[14]:

for c in snore.columns:

if c == "name" or c=='sno ' or c=='sex':

continue

else:

plt.boxplot(snore[c])

plt.xlabel(c)

plt.show()

# #### TO SHOW RELATION BETWEEN INDEPENDET VARIABLES

# In[18]:

sns.pairplot(snore.iloc[:,:-1])

# In[19]:

# all are not uniformly distributed ,may contain some outliers

# ##### as inputs from domain expert ahi value only depends on apneas count,hypoapneas count,ah total count and sleep time

#

# ### ahi value does not depend on name,sex so building model removing those

# In[156]:

snore.columns

# ## regression model

# In[157]:

snore

# In[165]:

model=smf.ols('ahi~bmi+totalsleeptime+ahtotalcount+apneascount+hypopneascount+apneasindex+hypopnnore).easindex',data=sfit()

# In[166]:

model.params

# In[167]:

model.summary()

# In[16]:

### model.confint(0.05)

# In[20]:

### apneas index and hypopneas index became less significant

### so lets construct the model ignoring this and check others significance

# In[21]:

model1=smf.ols('ahi~totalsleeptime+apneascount+hypopneascount+bmi',data=snore).fit()

# In[22]:

model1.params

# In[23]:

model1.summary()

# In[24]:

model2=smf.ols('ahi~apneasindex+hypopneasindex+bmi+totalsleeptime',data=snore).fit()

# In[25]:

model2.summary()

#

# In[26]:

sm.graphics.influence\_plot(model1,criterion="cooks")

plt.plot()

# In[30]:

from statsmodels.formula.api import ols

m=ols('ahi~bmi+totalsleeptime+apneascount+hypopneascount+apneasindex+hypopneasindex',data=snore).fit()

infl=m.get\_influence()

sm\_fr=infl.summary\_frame()

# In[31]:

sm\_fr

# ### cooks\_d values are less. .. there are no influential points

# In[35]:

### so apneas index and hypopneas index can be removed as they tend to overfit the model and have less significace

# ### vif values are also high and removing those attributes

# In[34]:

## so our final model contains only bmi,apneascount,hypopneascount,totalsleeptime

# In[ ]:

# lets remove bmi and check adjusted r-squared

# In[36]:

model3=smf.ols('ahi~apneascount+hypopneascount+totalsleeptime',data=snore).fit()

# In[37]:

model3.summary()

# In[38]:

### adjusted r squared for the model cotaining bmi,apneascount and hypopneas count is 0.907

# In[39]:

# including bmi

# In[40]:

model4=smf.ols('ahi~bmi+apneascount+hypopneascount+totalsleeptime',data=snore).fit()

# In[41]:

model4.summary()

# In[46]:

bmi\_new=new\_snore['bmi']

apneascount\_new=new\_snore['apneascount']

hypopneascount\_new=new\_snore['hypopneascount']

totalsleeptime\_new=new\_snore['totalsleeptime']

ahi\_new=new\_snore['ahi']

# In[47]:

model4=smf.ols('ahi\_new~bmi\_new+apneascount\_new+hypopneascount\_new+totalsleeptime\_new',data=new\_snore).fit()

# In[48]:

model4.summary()

# In[49]:

### after removing influential points bmi became signifiacnt

# In[50]:

### final dtaset ----new\_snore

# In[51]:

model\_final=smf.ols('ahi\_new~bmi\_new+apneascount\_new+hypopneascount\_new+totalsleeptime\_new',data=new\_snore).fit()

# In[53]:

model\_final.summary()

# #### multicollinearity is not exist in our final model

# In[59]:

model\_final\_pred=model\_final.predict(new\_snore)

# In[60]:

model\_final\_pred

# In[79]:

x=new\_snore.iloc[:,3:9].values

# In[80]:

y=new\_snore.iloc[:,10].values

# In[81]:

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size = 0.2, random\_state=0)

# In[82]:

new\_snore.columns

# In[83]:

new\_snore

type(new\_snore)

# In[93]:

from sklearn.linear\_model import LinearRegression

regressor=LinearRegression()

regressor.fit(x\_train,y\_train)

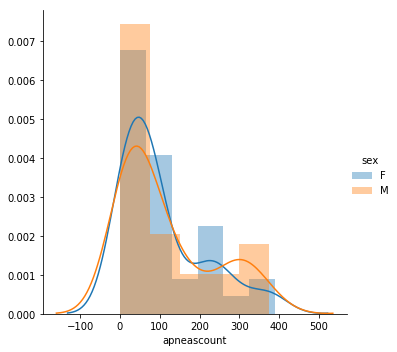
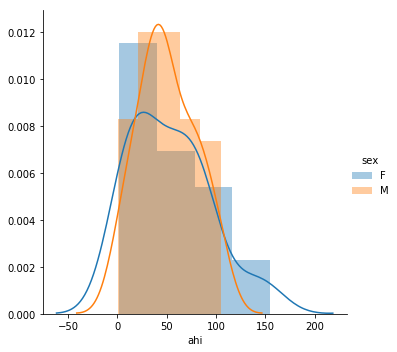
accuracy=regressor.score(x\_test,y\_test)

print((accuracy\*100))

# In[94]:

print("final model accuracy is 94.3%")

**Accuracy of the model is 94.2%**



A close up of a computer

Description automatically generated