**Assignment 8**

**Roll no.: A-44**

**Subject: DAP**

**Code:**

import pandas as pd

import matplotlib.pyplot as plt

import statsmodels.api as sm

from statsmodels.formula.api import ols

import seaborn as sns

import pandas.tseries

import numpy as np

from statsmodels.stats.multicomp import pairwise\_tukeyhsd

from scipy.stats import ttest\_ind,f\_oneway

plt.style.use('fivethirtyeight')

myData=pd.read\_csv("diet.csv")

print(myData.head(10))

print("Total number of rows in the dataset: ",myData.size)

print(myData.gender.unique())

print(myData[myData.gender == ' '])

f,ax=plt.subplots(figsize=(11,9))

plt.title("Weight distribution among sample")

plt.ylabel('pdf')

sns.distplot(myData.final\_weight)

plt.show()

myData['final\_weight']=pd.to\_numeric(myData['final\_weight'],errors='coerce')

print(myData.groupby('gender')['final\_weight'].agg([np.mean,np.median,np.std,np.count\_nonzero]))

print(myData.groupby(['gender','diet\_type'])['final\_weight'].agg([np.mean,np.median,np.std,np.count\_nonzero]))

mymod\_female=ols('height~diet\_type',data=myData[myData.gender == 'Female']).fit()

aovtable\_female=sm.stats.anova\_lm(mymod\_female,type=2)

print("ANOVA table for female")

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

print(aovtable\_female)

print()

mymod\_male=ols('height~diet\_type',data=myData[myData.gender == 'Male']).fit()

aovtable\_male=sm.stats.anova\_lm(mymod\_male,type=2)

print("ANOVA table for male")

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

print(aovtable\_male)

print()

for group1 in myData['diet\_type'].unique():

for group2 in myData['diet\_type'].unique():

if group1 != group2:

data\_group1=myData[myData['diet\_type']==group1]['height']

data\_group2=myData[myData['diet\_type']==group2]['height']

t\_stats,p\_val=ttest\_ind(data\_group1,data\_group2)

print(f"Pairwise t-test b/w {group1} and {group2}: p-value = {p\_val}")

f\_stat,p\_val=f\_oneway(\*[myData[myData['diet\_type']==diet]['height'] for diet in myData['diet\_type'].unique()])

print(f"ANOVA: F-statistic = {f\_stat}, p-value = {p\_val}")

if p\_val<0.05:

tukey\_res=pairwise\_tukeyhsd(myData['height'],myData['diet\_type'])

print(tukey\_res)

**Output**

id gender age height diet\_type initial\_weight final\_weight

0 1 Female 22 159 A 58 54.2

1 2 Female 46 192 A 60 54.0

2 3 Female 55 170 A 64 63.3

3 4 Female 33 171 A 64 61.1

4 5 Female 50 170 A 65 62.2

5 6 Female 50 201 A 66 64.0

6 7 Female 37 174 A 67 65.0

7 8 Female 28 176 A 69 60.5

8 9 Female 28 165 A 70 68.1

9 10 Female 45 165 A 70 66.9

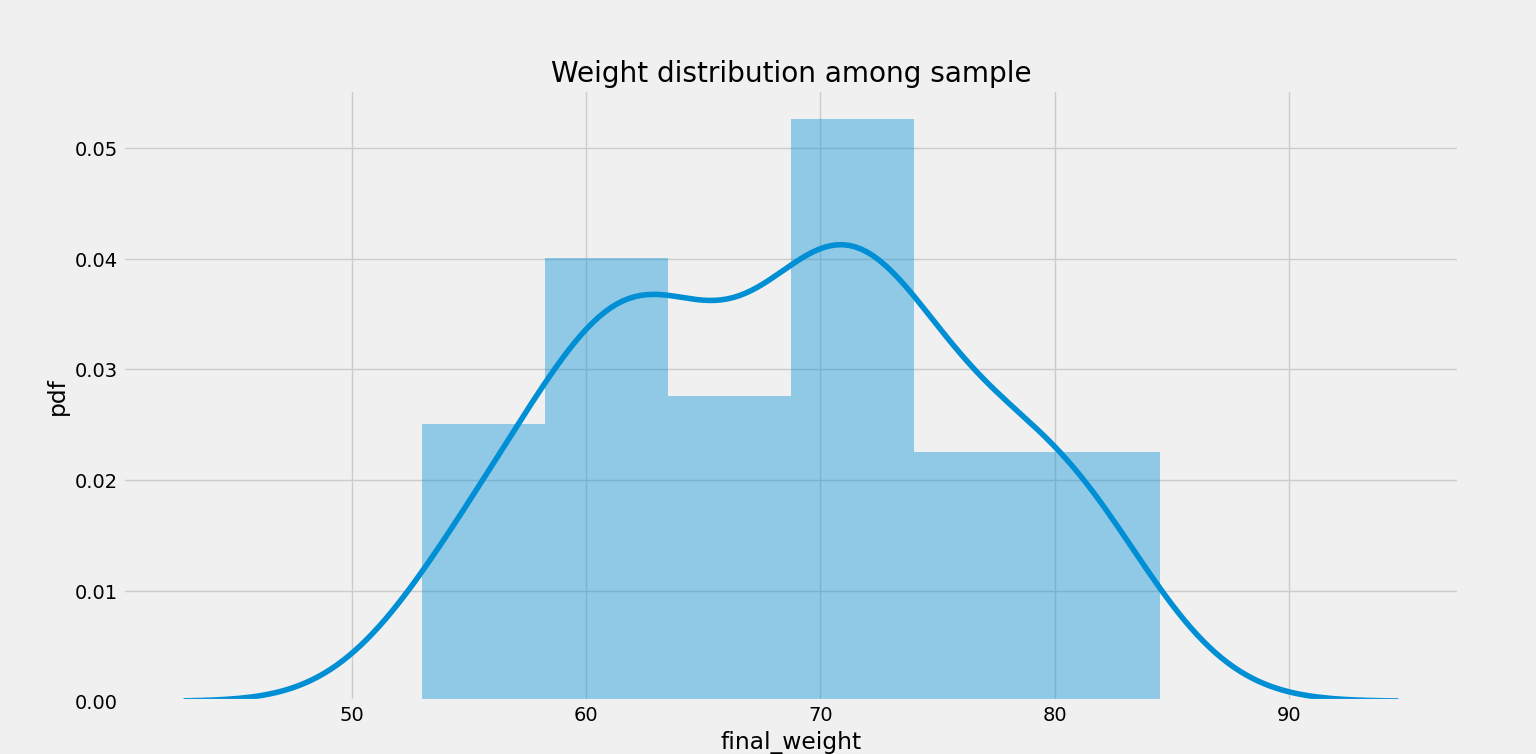
Total number of rows in the dataset: 532

['Female' 'Male']

Empty DataFrame

Columns: [id, gender, age, height, diet\_type, initial\_weight, final\_weight]

Index: []



mean median std count\_nonzero

gender

Female 63.223256 62.4 6.150874 43

Male 75.015152 73.9 4.629398 33

mean median std count\_nonzero

gender diet\_type

Female A 64.878571 64.50 6.877296 14

B 62.178571 61.15 6.274635 14

C 62.653333 61.80 5.370537 15

Male A 76.150000 75.75 5.439414 10

B 73.163636 72.70 3.818448 11

C 75.766667 76.35 4.434848 12

ANOVA table for female

-----------------------

df sum\_sq mean\_sq F PR(>F)

diet\_type 2.0 565.938870 282.969435 3.548397 0.038139

Residual 40.0 3189.828571 79.745714 NaN NaN

ANOVA table for male

-----------------------

df sum\_sq mean\_sq F PR(>F)

diet\_type 2.0 1367.893939 683.946970 5.87219 0.007044

Residual 30.0 3494.166667 116.472222 NaN NaN

Pairwise t-test b/w A and B: p-value = 0.16541391613468584

Pairwise t-test b/w A and C: p-value = 0.29957666257576565

Pairwise t-test b/w B and A: p-value = 0.16541391613468584

Pairwise t-test b/w B and C: p-value = 0.015303633764582497

Pairwise t-test b/w C and A: p-value = 0.29957666257576565

Pairwise t-test b/w C and B: p-value = 0.015303633764582497

ANOVA: F-statistic = 3.225113125905158, p-value = 0.04547909842728667

Multiple Comparison of Means - Tukey HSD, FWER=0.05

=====================================================

group1 group2 meandiff p-adj lower upper reject

-----------------------------------------------------

A B 4.7483 0.2981 -2.8376 12.3343 False

A C -3.0324 0.5954 -10.4795 4.4147 False

B C -7.7807 0.0361 -15.1486 -0.4129 True

-----------------------------------------------------