import pandas as pd

import numpy as np

import statistics

import matplotlib.pyplot as plt

#q1

df=pd.DataFrame(pd.read\_csv('pima-indians-diabetes.csv'))

a=list(df['pregs'])

b=list(df['plas'])

c=list(df['pres'])

d=list(df['skin'])

e=list(df['test'])

f=list(df['BMI'])

g=list(df['pedi'])

h=list(df['Age'])

i=list(df['class'])

def part\_1(x):

print('mean-',statistics.mean(x))

print('median-',statistics.median(x))

print('mode-',statistics.mode(x))

print('maximum-',max(x))

print('minimum-',min(x))

print('standard deviation-',statistics.stdev(x))

print('data for attribute pregs'),part\_1(a)

print()

print('data for attribute plas'),part\_1(b)

print()

print('data for attribute pres'),part\_1(c)

print()

print('data for attribute skin'),part\_1(d)

print()

print('data for attribute test'),part\_1(e)

print()

print('data for attribute BMI'),part\_1(f)

print()

print('data for attribute pedi'),part\_1(g)

print()

print('data for attribute Age'),part\_1(h)

#q2a

def part\_2a(x):

plt.scatter(h,x)

plt.show()

part\_2a(a)

part\_2a(b)

part\_2a(c)

part\_2a(d)

part\_2a(e)

part\_2a(f)

part\_2a(g)

#q2b

def part\_2b(x):

plt.scatter(f,x)

plt.show()

part\_2b(a)

part\_2b(b)

part\_2b(c)

part\_2b(d)

part\_2b(e)

part\_2b(g)

part\_2b(h)

#q3a

def standev(m):

return((((len(a))/(len(a)-1))\*\*0.5)\*np.std(m))

def corr\_coeff(x,y):

n = 0

for i in range(len(x)):

n += (x[i]-np.average(x))\*(y[i]-np.average(y))

covariance = n/(len(a)-1)

z=covariance/(standev(x)\*standev(y))

return(z)

print('Correlation between Age and pregs:',corr\_coeff(h,a))

print()

print('Correlation between Age and plas:',corr\_coeff(h,b))

print()

print('Correlation between Age and pres:',corr\_coeff(h,c))

print()

print('Correlation between Age and skin:',corr\_coeff(h,d))

print()

print('Correlation between Age and test:',corr\_coeff(h,e))

print()

print('Correlation between Age and BMI:',corr\_coeff(h,f))

print()

print('Correlation between Age and pedi:',corr\_coeff(h,g))

print()

#q3b

print('Correlation coefficient between BMI and pregs:',corr\_coeff(f,a))

print()

print('Correlation between BMI and plas:',corr\_coeff(f,b))

print()

print('Correlation between BMI and pres:',corr\_coeff(f,c))

print()

print('Correlation between BMI and skin:',corr\_coeff(f,d))

print()

print('Correlation between BMI and test:',corr\_coeff(f,e))

print()

print('Correlation between BMI and pedi:',corr\_coeff(f,g))

print()

print('Correlation between BMI and Age:',corr\_coeff(f,h))

print()

#q4

# plt.hist(a,d)

# plt.show()

#q5

#q6

plt.boxplot(a)

plt.boxplot(b)

plt.boxplot(c)

plt.boxplot(d)

plt.boxplot(e)

plt.boxplot(f)

plt.boxplot(g)

plt.boxplot(h)