|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Answer:

Answer: Total number of outcomes = 2^3=8

Total combinations are : {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

Having two heads and one tail are = {HHT, HTH, TTH} = 3

Probability that two heads and one tail are 1/8+1/8+1/8=3/8= 0.375 = 37.5%

Q4) Two Dice are rolled, find the probability that sum is

Total number of out comes are = 6^2=36

The possible sums of two dice rolls are (2,3,4,5,6,7,8,9,10,11,12)

1. Equal to 1

Answer: 0% Probability of that sum is equal to 1.

1. Less than or equal to 4 Answer: There are 3 possible number of combinations possible when two dies rolled the sum is less than or equal to 4. Probability=1/36+2/36+3/36=1/6

c) Sum is divisible by 2 and 3 Answer: The only two sums(6 and 12) are divisible by 2 and 3. The 6 sum having 5 possible ways and 12 sum having 1 possible way to get, then total possible ways to get the sum which is divisible by 2and 3. Probability=6/36=0.1666=16.66%

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Answer:

Two balls are drawn every time randomly

Total number of balls are 2+3+2 = 7

Number of balls with out blue colour = 2+3 =5

Drawing 2 balls from total of 7 balls is

Drawing 2 balls none of them are blue 5

The probability that none of the balls drawn is blue: 5 is 10/21=47.6%

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

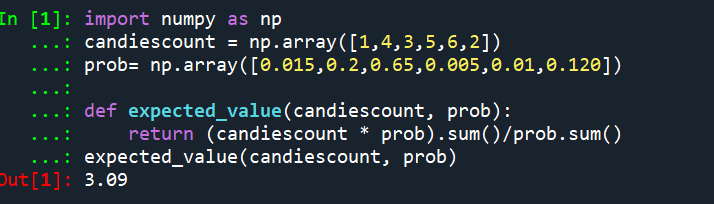
|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Answer: Expected number E(x)=(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)=3.09

Python program compilation using Spyder:



Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

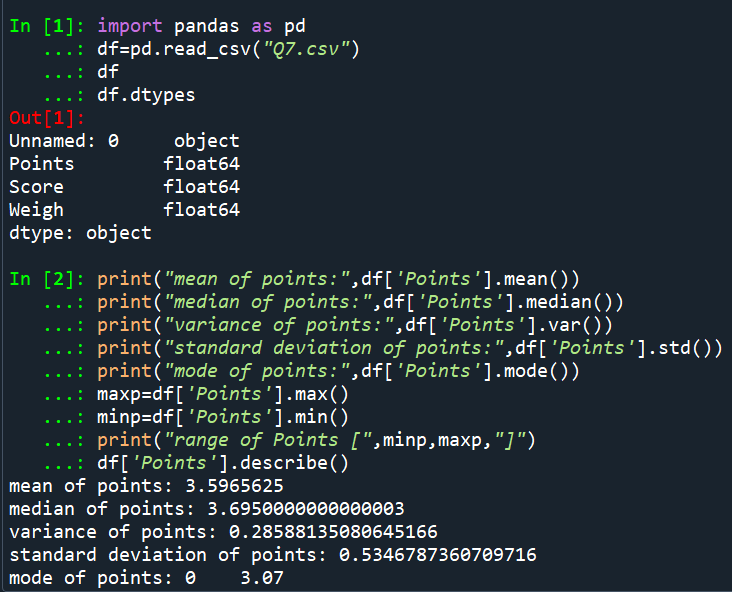
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

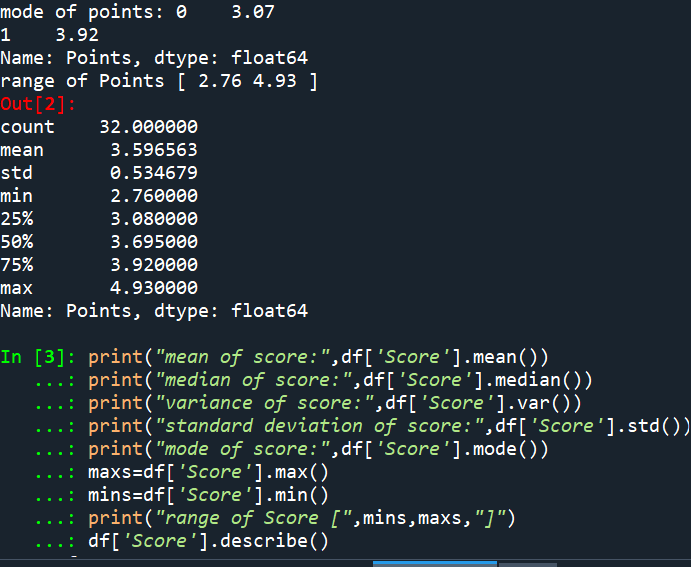
**Use Q7.csv file**

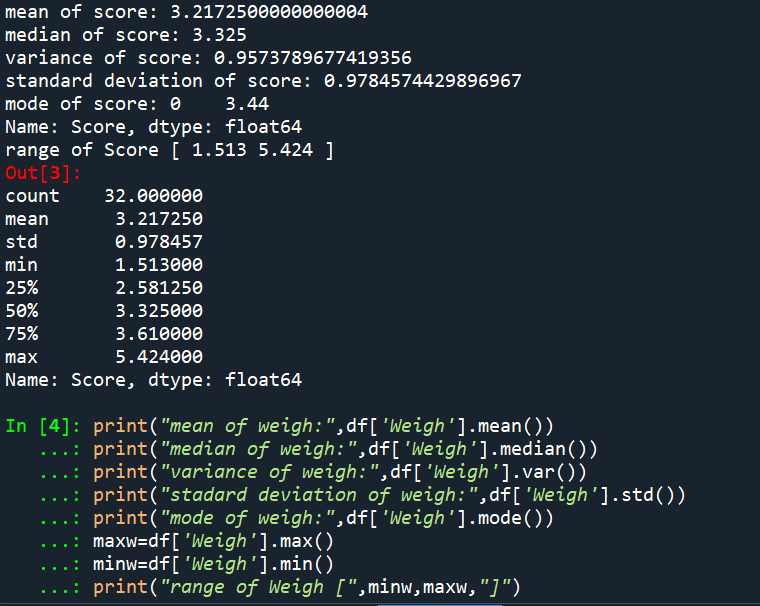
Answer:

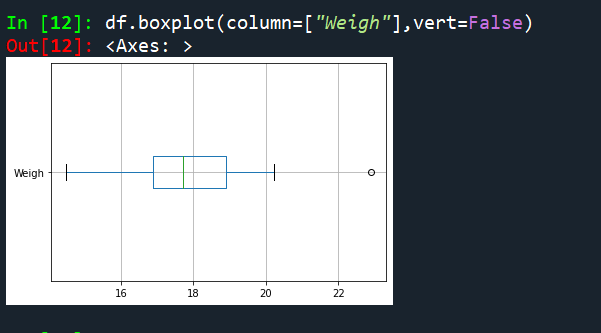
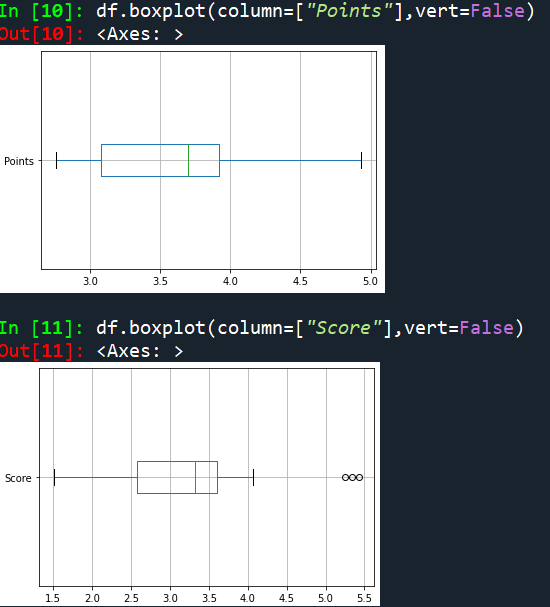
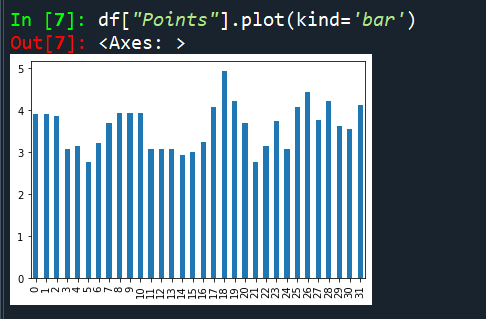
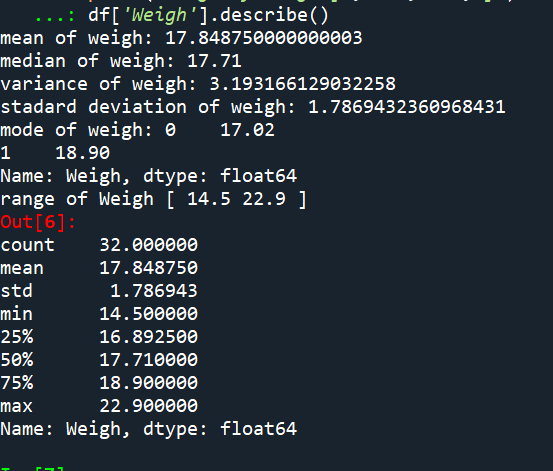
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Data | Mean | Mode | Variance | Standard Deviation | Range | Median |
| Points | 3.596 | 3.07&3.92 | 0.285 | 0.534 | [2.76,4.93] | 3.695 |
| Score | 3.21 | 3.44 | 0.957 | 0.978 | [1.513,5.424] | 3.325 |
| Weigh | 17.848 | 17.02&18.90 | 3.193 | 1.786 | [14.5,22.9] | 17.71 |

Python program compilation using Spyder:



****





Q8) Calculate Expected Value for the problem below

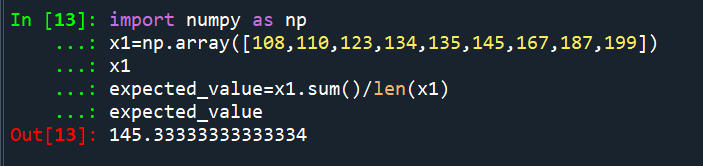
1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Answer:

Python code compilation using spyder:



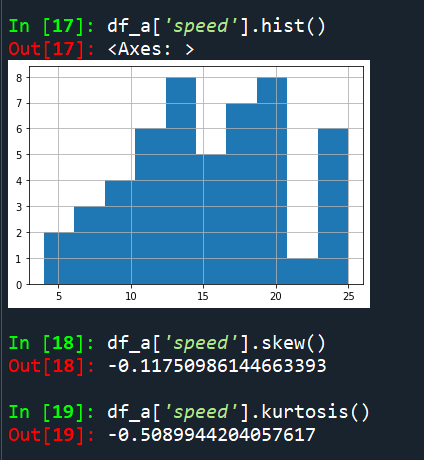
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

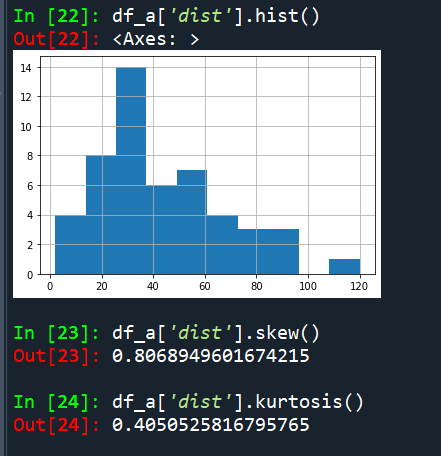
**Cars speed and distance**

**Use Q9\_a.csv**

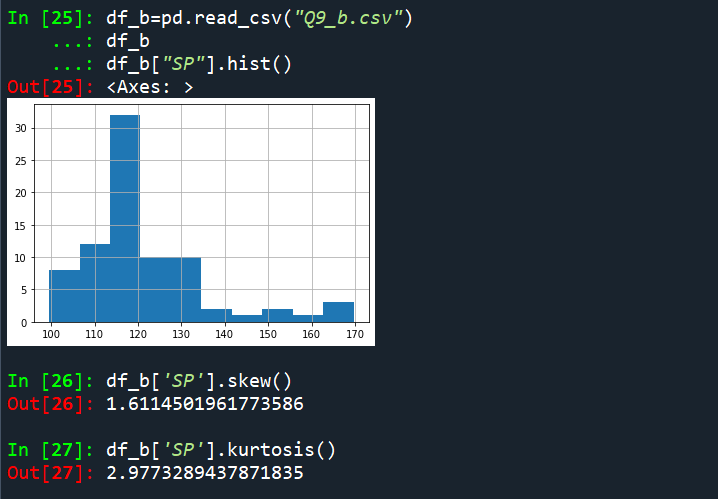
Answer:

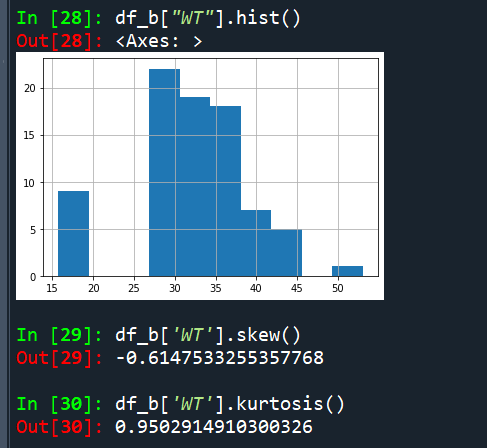
Python code compilation using spyder:



****

**Use Q9\_b.csv**

****

****

**Q10) Draw inferences about the following boxplot & histogram**



Answer: 50-100 weight having more frequency around 180 and 350-400 weight having very less frequency around 5. Data is right skewed not normally distributed means follows positive skewness .



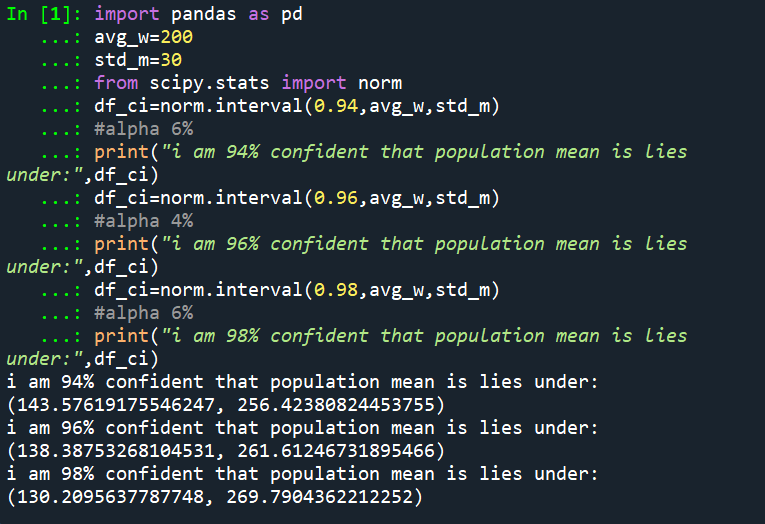
Answer: Data is having 7 Outliers and not normally distributed that is right skewed means follows positive skewness.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Answer:

|  |  |
| --- | --- |
| %confidence | Confidence interval |
| 94% | (143.576, 256.423) |
| 96% | (138.387, 261.612) |
| 98% | (130.209, 269.790) |

Python code compilation using spyder:



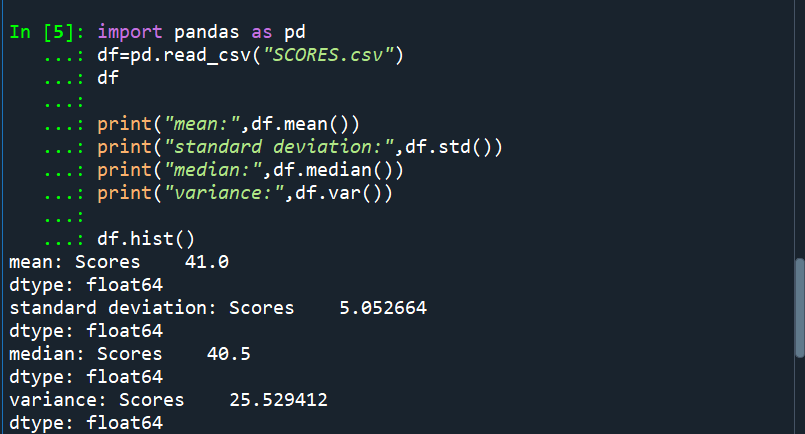
**Q12)** Below are the scores obtained by a student in tests

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

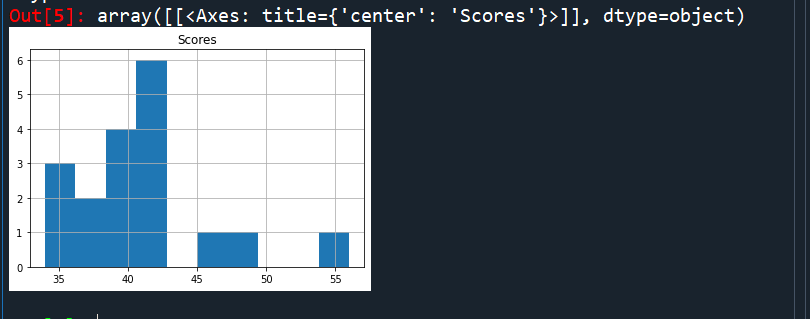
Answer:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Mode | 41 |
| Standard Deviation | 5.05 |
| Variance | 25.529 |

Python code compilation using spyder:



Student marks are right(positively) skewed.



Q13) What is the nature of skewness when mean, median of data are equal?

Answer: Shape of the data is follows symmetrical distribution with zero skewness.

Q14) What is the nature of skewness when mean > median ?

Answer: Positive Skewness

Q15) What is the nature of skewness when median > mean?

Answer: Negative Skewness

Q16) What does positive kurtosis value indicates for a data ?

Answer: Data has Highest Peak than normally distributed data.

Q17) What does negative kurtosis value indicates for a data?

Answer: Data has Lowest peak than normally distributed data.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Answer: There are no outliers

Q1(25th Quartile):10

Q2(50th Quartile):14

Q3(100th Quartile):18

What is nature of skewness of the data?

Answer:positive skewness

What will be the IQR of the data (approximately)?

Answer:IQR=Q3-Q1=18-10=8  
  
  
  
Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Answer: Box plot-2 having high spread in data and whisker lengths compared with Box Plot-1 and both the boxes having same median.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)

Answer: 1-cdf(38)= 0.347

* 1. P(MPG<40)

Answer: cdf(40)= 0.729

* 1. P (20<MPG<50)

Answer: cdf(50)-cdf(20)=0.898

Python code:

import pandas as pd

df=pd.read\_csv("Cars.csv")

df

#H0: Data is normal

#H1:Data is not normal

from scipy.stats import shapiro

test,pvalue=shapiro(df['MPG'])

pvalue

alpha=0.05

if pvalue<alpha:

print("H0 is rejected and H1 is accepted")

else:

print("H0 is accepted and H1 is rejected")

mean\_mpg=df['MPG'].mean()

std\_mpg=df['MPG'].std()

#NORMAL DISTRIBUTION

from scipy.stats import norm

nd=norm(mean\_mpg,std\_mpg)#mean,sd

#P(MPG<38)

nd.cdf(38)

#p(MPG>38)

1-nd.cdf(38)

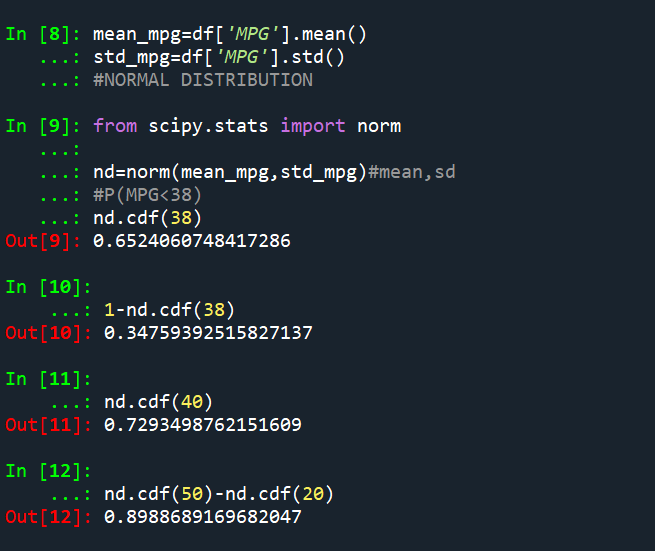
#P(MPG<40)

nd.cdf(40)

#P(20<MPG<50)

nd.cdf(50)-nd.cdf(20)

Python code compilation using spyder:



Q 21) Check whether the data follows normal distribution

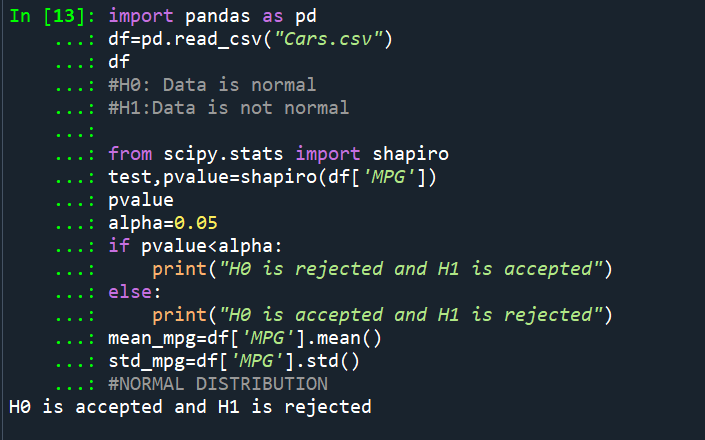
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: By test of Hypothesis ->We got H0 is accepted and H1 is rejected

Then we can say that MPG follows Normal Distribution

Python code compilation using spyder:



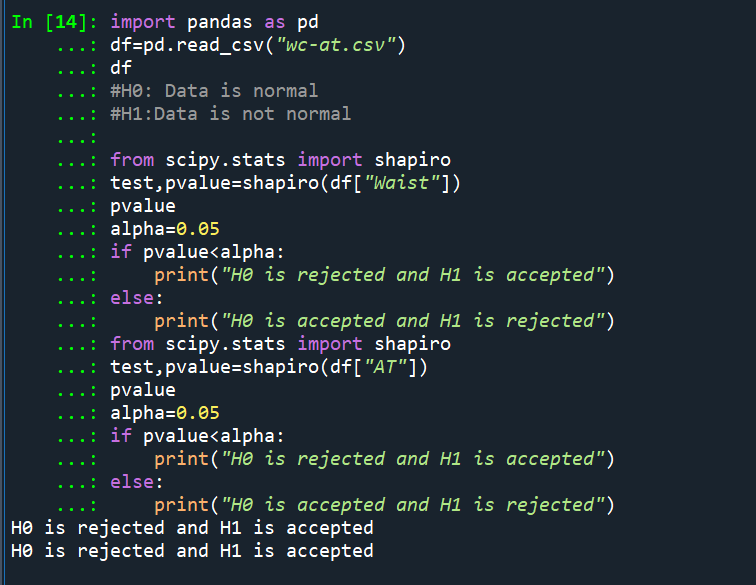
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer: By test of Hypothesis ->We got H1 is accepted and H0 is rejected for AT and Waist

Then we can say that both not follows Normal Distribution.

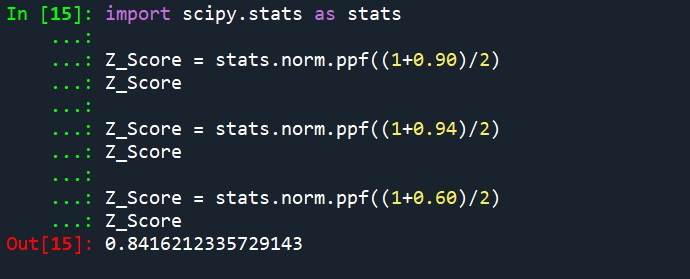
Python code compilation using spyder:



Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Answer:

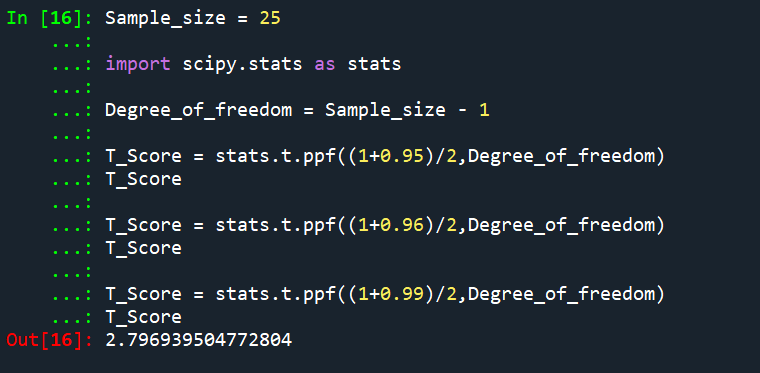
Python code and compilation using spyder:



Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Answer:

Python code and compilation using spyder:



Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Answer: the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.3217

Python code and compilation using spyder:

