|  |  |
| --- | --- |
| 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 | Continuous |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer:

*Samples we get: HHT,HHT,HTH,HTT,THH,THT,TTH,TTT*

*P(A)=3/8=0.375*

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

*Answer:*

*Here is the sample space:*

*(1,1),(1,2),(1,3)(1,4)(1,5)(1,6)*

*(2,1),(2,2),(2,3)(2,4)(2,5)(2,6)*

*(3,1),(3,2),(3,3)(3,4)(3,5)(3,6)*

*(4,1),(4,2),(4,3)(4,4)(4,5)(4,6)*

*(5,1),(5,2),(5,3)(5,4)(5,5)(5,6)*

*(6,1),(6,2),(6,3)(6,4)(6,5)(6,6)*

1. Equal to 1

0

b) Less than or equal to 4

6/36=1/6

c) Sum is divisible by 2 and 3

6/36=1/6

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?

*Solution:7c2/5c2*

*N=7\*6/1\*2=21*

*X=5\*4/2\*1=10*

*Ans is P(A)=10/21*

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

*Ans : Solution: 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120*

*Expected No of candies:3.09*

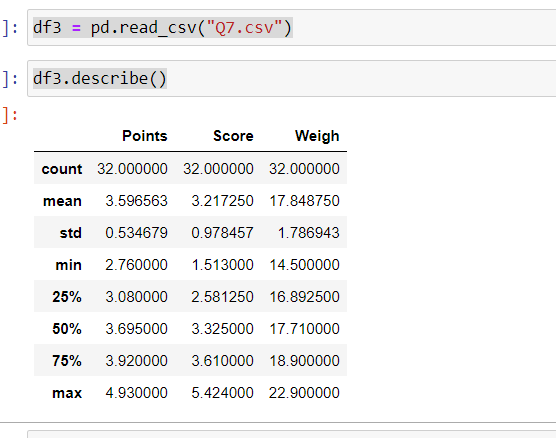
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

Answer:

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Points*** | ***Score*** | ***Weigh*** |
| **mean** | *3.59* | *3.21* | *17.8* |
| **Median** | *3.695* | *3.325* | *17.71* |
| **mode** | *3.92* | *3.44* | *17.02* |
| **var** | *0.276947559* | *0.927461* | *3.09338* |
| **SD** | *0.526258072* | *0.963048* | *1.758801* |
| **max** | *4.93* | *5.424* | *22.9* |
| **min** | *2.76* | *1.513* | *14.5* |
| **range** | *2.17* | *3.911* | *8.4* |
|  |  |  |  |



Above ans in line assignment\_answers\_swati.pynb file->line 37

For the 3 datasets Point,score and weigh , the mean, median and mode are almost near to each other which depicts it could near a symmetrical distribution.

**Use Q7.csv file**

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?

Expected value of the weight of that patient: 108+110+123+134+135+145+167+187+199/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Python file attached: Assignment\_answers\_swati.pynb->line13 to line35**

Ans: **For speed data** skewness is: -0.117

Kurtosis is: -0.508

It’s left skewed as well negative kutosis tells that the distribution is flat

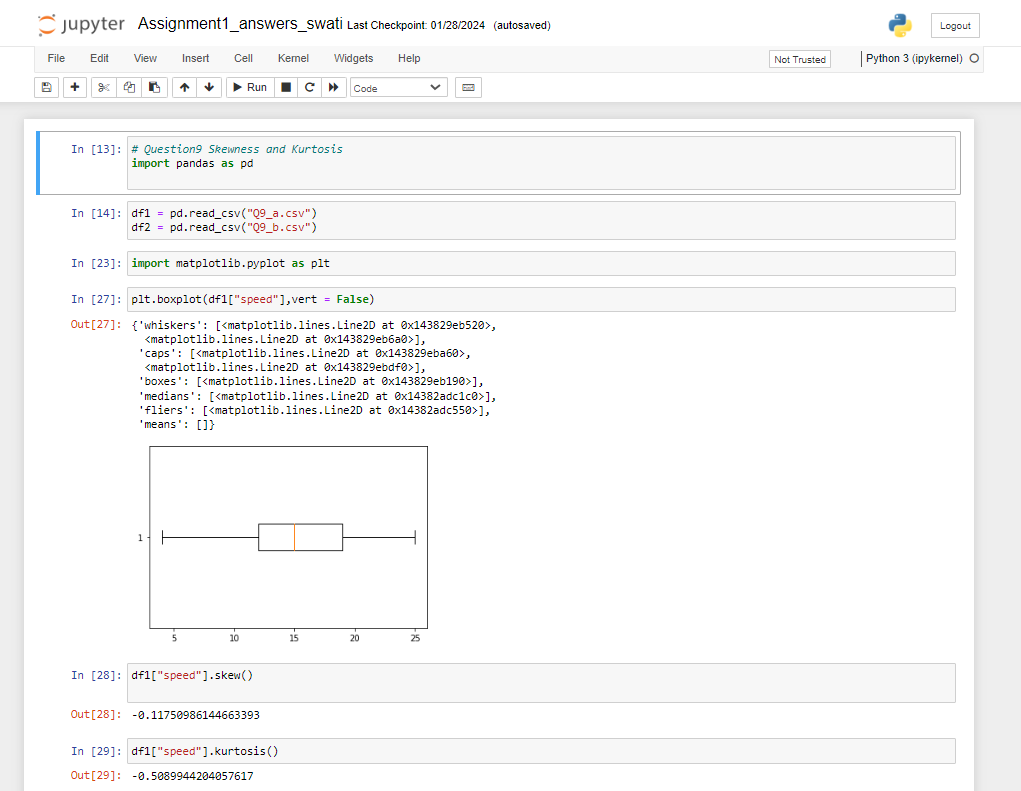
No outliers.

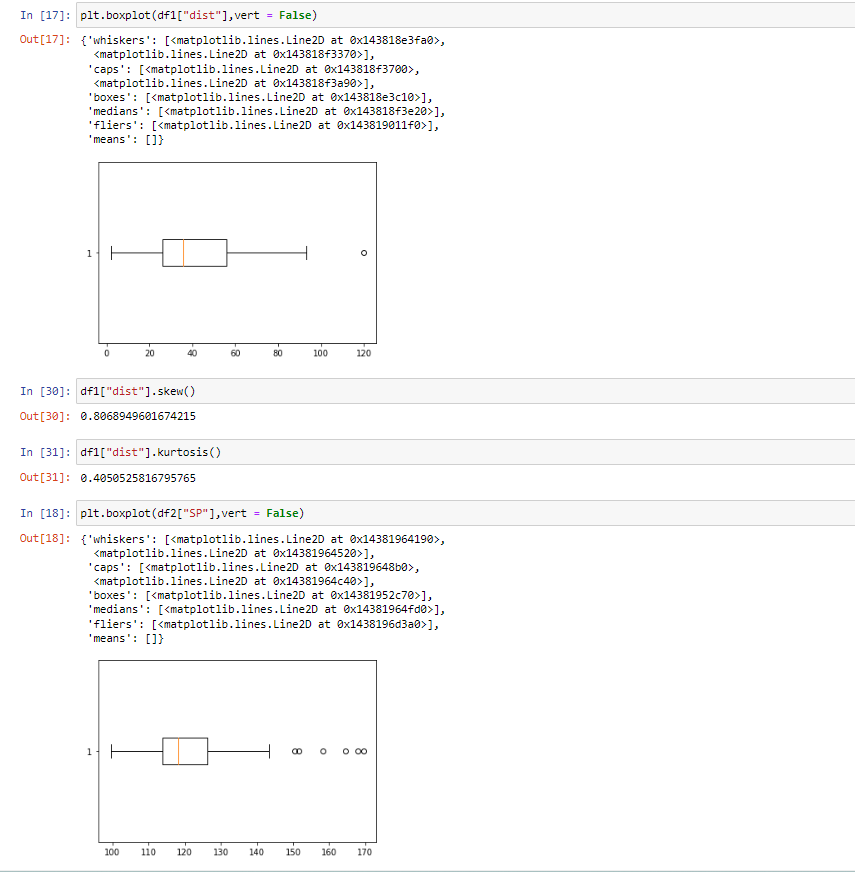
**For distance data:** Skewness: 0.8068949601674215

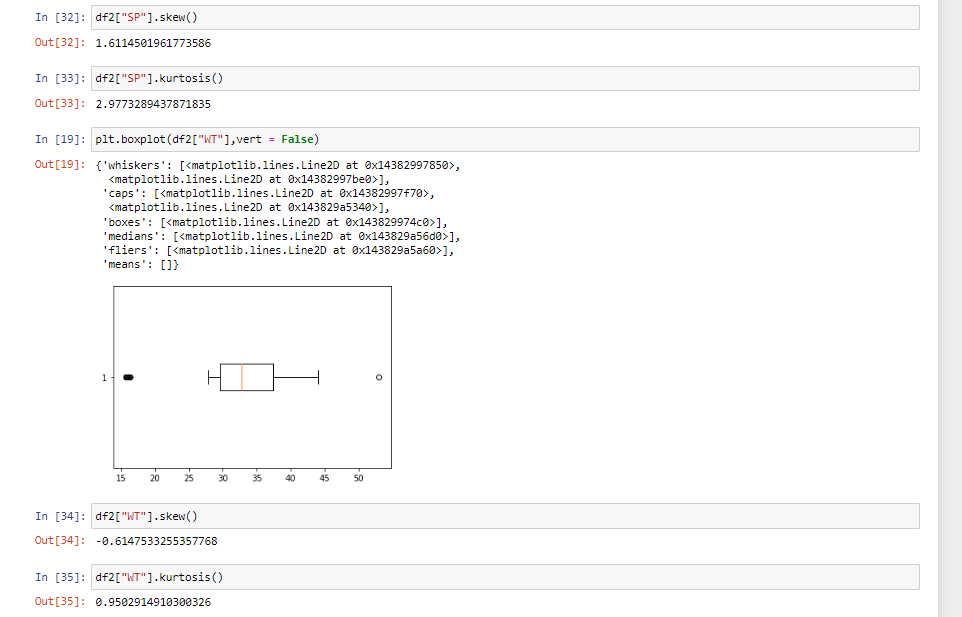
Kurtosis: 0.4050525816795765

Data is right skewed and distribution has more peakedness.

It has outliers present.







**SP and Weight(WT)**

Use Q9\_b.csv

For SP data:

Skewness: 1.6114501961773586

Kurtosis: 2.9773289437871835

Data is right skewed and distribution has more peakedness.

It has outliers present.

**For WT data:**

Skewness: -0.6147533255357768

Kurtosis:0.9502914910300326

Data is skewed to the left and has more peakedness as kurtosis is +ve

It has outliers present.

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



Ans: It’s a frequesncy distribution plot against weight which tells data is right skewed. There are extreme values present in lower ends of the range especially from 250 to 400 which depicts that there could be outliers in the data.

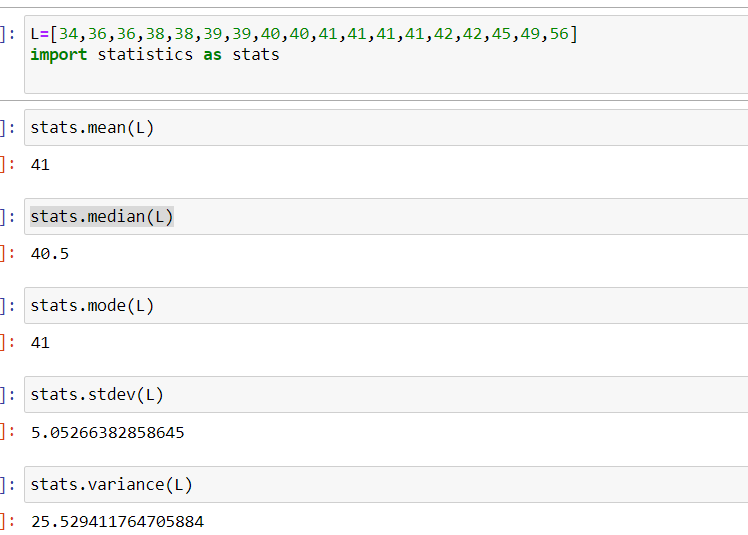
For interval between 50-100 we have highest numbers of observations.

**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?

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

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.



1. What can we say about the student marks?

Ans: The avg score of students 41 and most of the students got 41 marks.

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

symmetrical

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

Right skewed

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

Left skewed

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

More peakedness

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

Flat Data distribution

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



What can we say about the distribution of the data?

Whiskers are long here and median(Q2) is 15. Q2-Q1>Q3-Q2.

Q1 is 10, Q3 is 18

What is nature of skewness of the data?

Left skewed

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

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.

Ans: Both are similar and symmetrical distributions where Q3-Q2=Q2-Q1

IQR for Boxplot1:Q3-Q1:275-250=25

IQR for Boxplot2:Q3-Q1:300-225=75

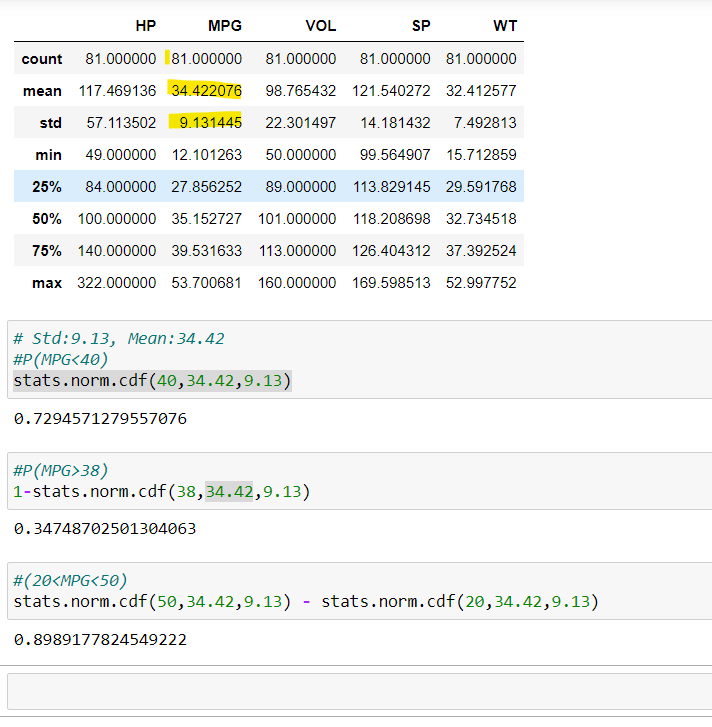
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

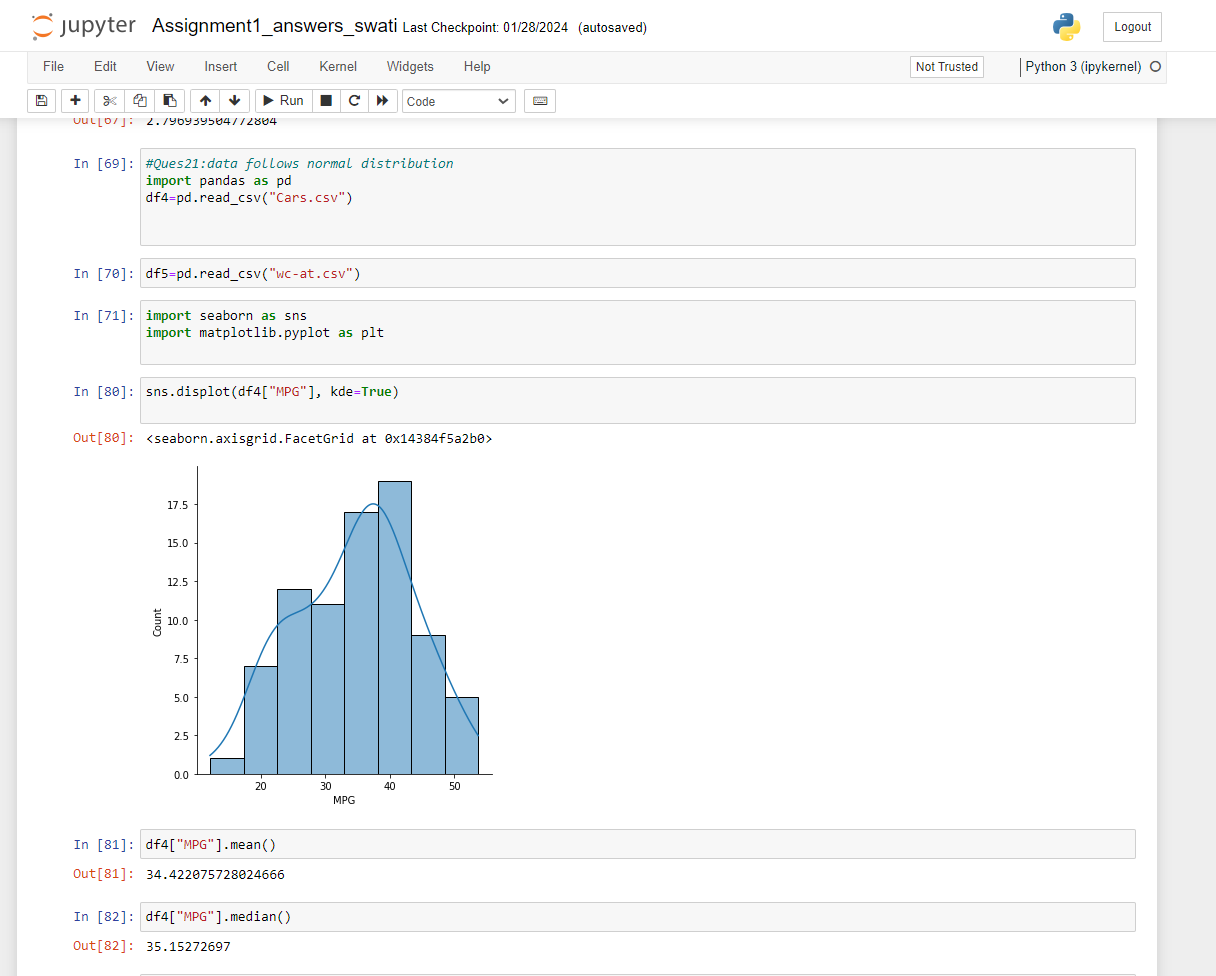
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

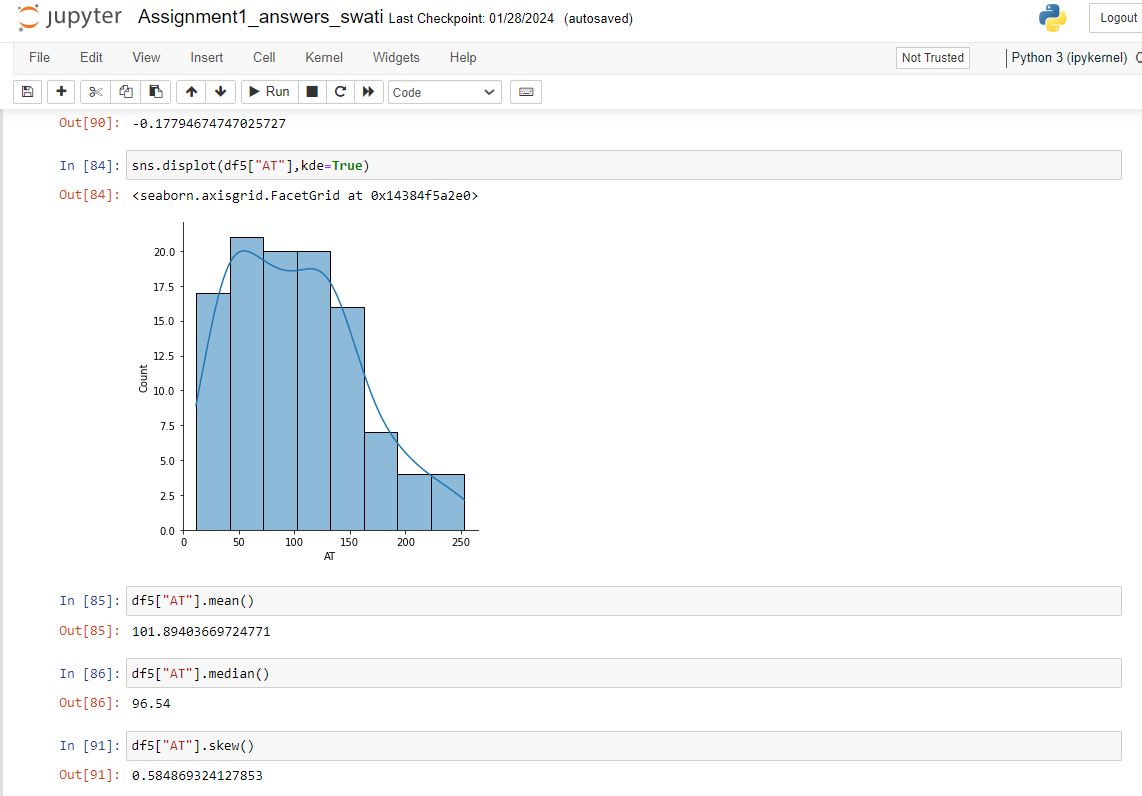
Dataset: Cars.csv

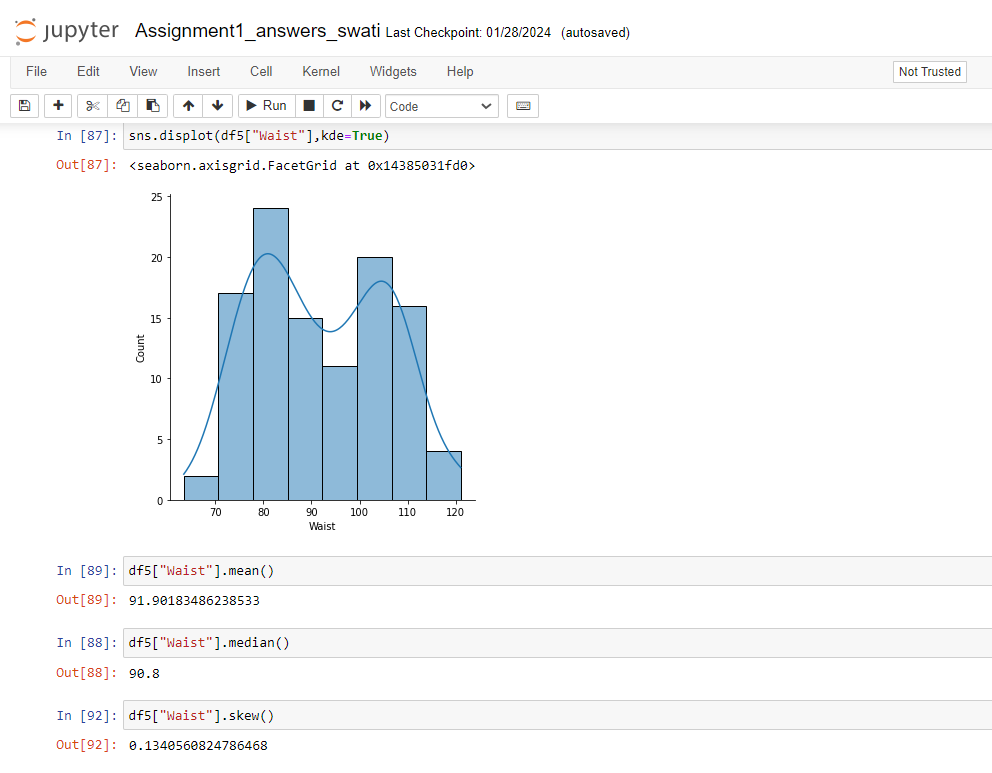




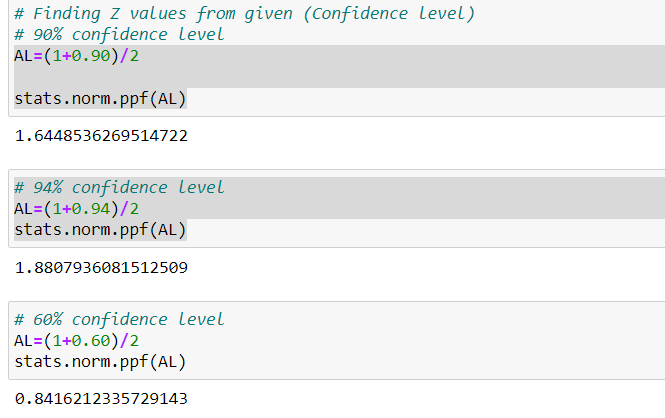
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

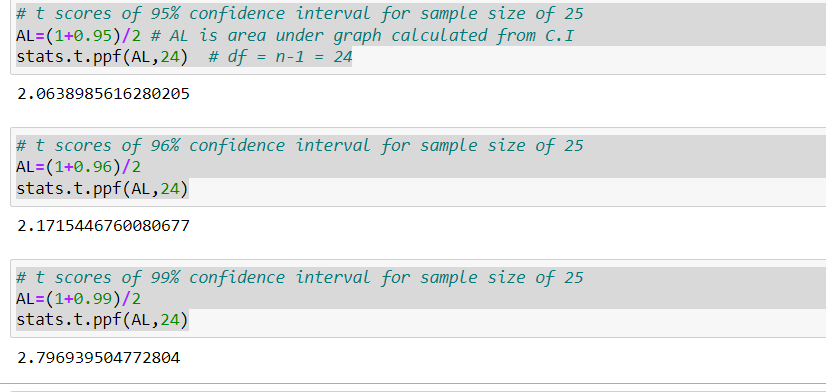




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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



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

*Ho= Avg life of bulb is=270*

*Ha= Avg life of bulb less than 260*

*Given values Mean=260, SD=90 n=18*

*Df=18-1=17*

*It has probability value of 0.32*

