|  |  |
| --- | --- |
| 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 | Classification - Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Classification - Nominal |

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) | Ratio |
| 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 - The probability of getting 2H and 1T in toss of 3 coins is 3/8 or 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 –

1. If two dices were rolled, then total possible cases =36. Total Favorable cases (Having sum =1) = 0. As minimum sum is 2 for outcome (1,1). **So, probability of getting sum equals to 1 is 0.**
2. Probability of an event B occurs is define as: P(B) = N(B) / N(S)

Events of getting Sum less than or equal to 4 are: N(B) = 6

|  |  |  |
| --- | --- | --- |
| Die 1 | Dei2 | Sum |
| 1 | 1 | 2 |
| 1 | 2 | 3 |
| 1 | 3 | 4 |
| 2 | 1 | 3 |
| 2 | 2 | 4 |
| 3 | 1 | 4 |

No of events for rolling two dies are: N(S) = 36

So, Probability for getting sum less than or equal to 4 = 6/36 = 1/6

1. Total number of possible outcomes = 36

Favorable outcomes = sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1, 5), (3, 3), (4, 2), (5, 1), (6, 6)

Therefore,

Number of favorable outcomes = 5

So, Probability for getting sum divisible by 2 and 3 = 5/36.

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 –

Total number of balls = S = (2 + 3 + 2) = 7  
Then, n(S) = Number of ways of drawing 2 balls out of 7 = 7C2​  
= (7 x 6) / (2 x 1) = 42/2 = 21  
Let E = Event of drawing 2 balls, none of which is blue.  
So, n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls = 5C2​  
= (5 x 4) / (2 x 1)​ = 20/2 = 10  
So, Probability of event E = P(E)= n(E) / n(S) ​= 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

Answer - from above we need to calculate expected no. of candies for a randomly selected child

= (1\*0.015+4\*0.2+3\*0.65+5\*0.005+6\*0.01+2\*0.012) = 3.09

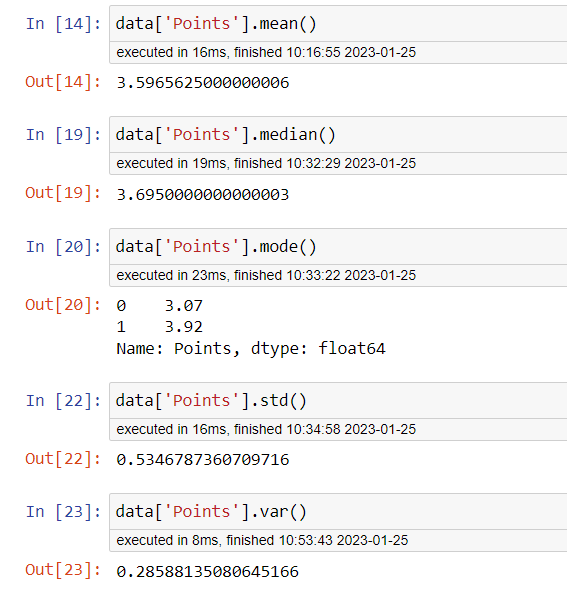
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

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

**Use Q7.csv file**

1. **Points**

****

1. **Score**

****

1. **Weigh**

****

There isn’t too high difference between mean and median values of all parameter. So, the data seems to be approximately normally distributed. Weighs dataset have one outlier.

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 – total no. of patients = 9

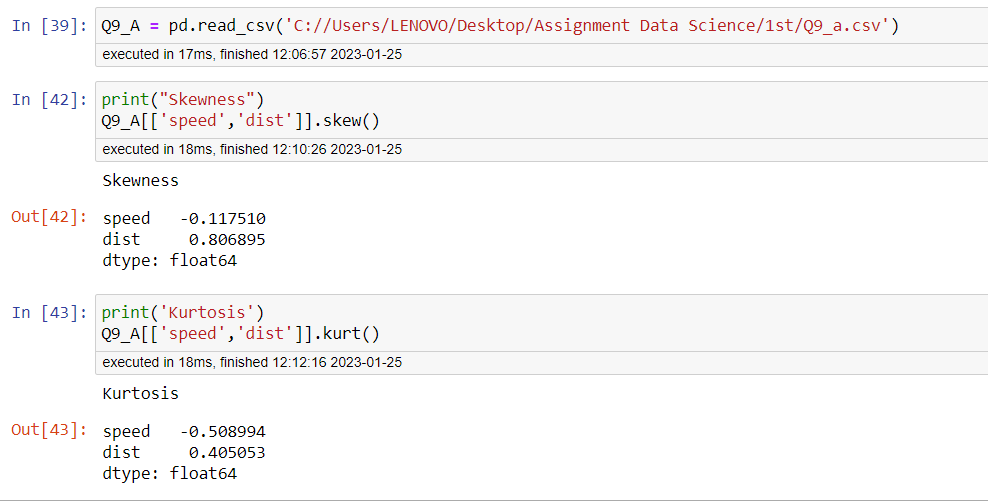
Probability of selecting one person = 1/9

= (1/9) \* (108+110+123+134+135+145+167+187+199) = 145.34

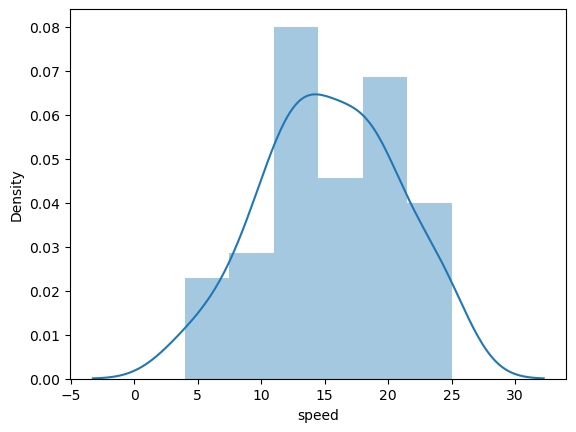
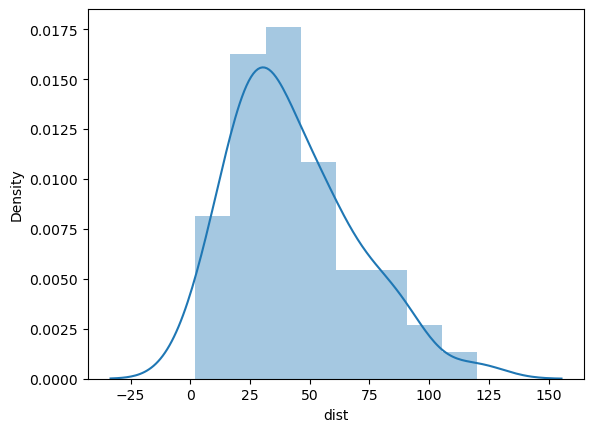
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

****

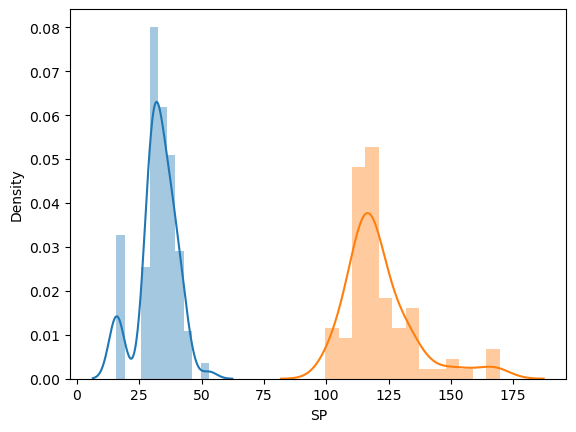
Data is normally distributed. Speed data set is slightly lift skewed whereas Dist data set is slightly right skewed as shown in distplot below.

** **

**SP and Weight (WT)**

**Use Q9\_b.csv**

****

Datasets for SP and WT function are normally distributed and do have positive kurtosis values. ****

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

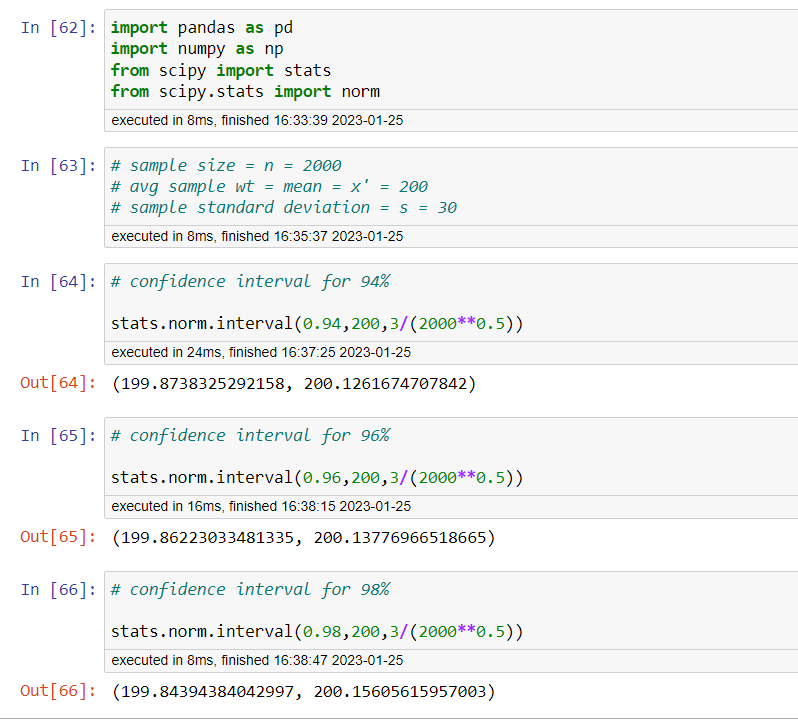


Answer – The histogram shown above is for right skewed dataset (have positive skewness). There is possibility of having some outliers having value above 350.



Answer – The box plot shown above is for right skewed dataset (positive skewness). We have some outliers beyond upper extreme limit.

**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?



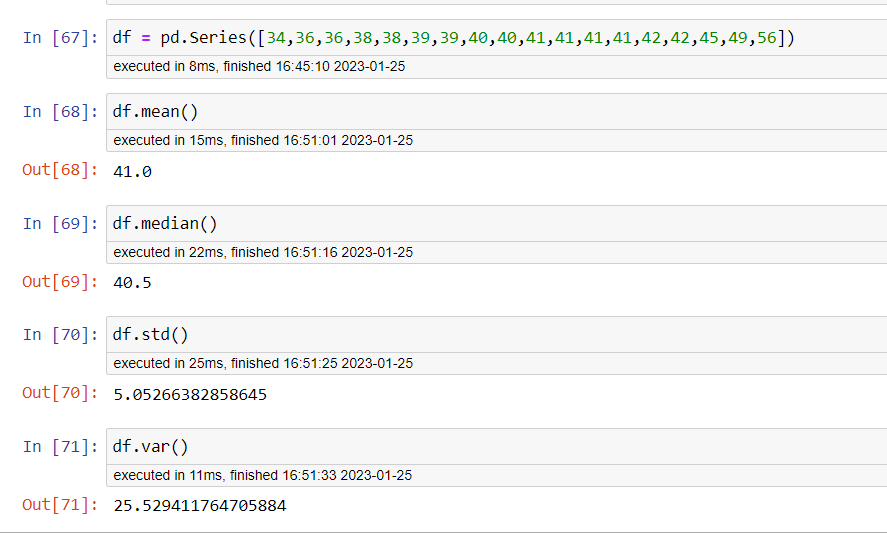
**Q12)** Below are the scores obtained by a student in tests

**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.
2. What can we say about the student marks?

Answer –

1. Mean, Median, Standard Deviation and Variance are calculated in jupyter notebook and the screen shot is attached below.
2. No diff. in mean and median as there is no any outliers available.



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

Answer – When mean and median of data are equal then there is zero skewness in the data set.

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

Answer – Right Skewness or Positive Skewness

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

Answer – Left Skewness or Negative Skewness

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

Answer – Thinner Pick and lighter data distribution at tail.

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

Answer – Flatter Pick and heavy tailed data distribution.

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



What can we say about the distribution of the data?

Answer – Mass of data distributed in left side is heavy. It is not normally distributed. Mean of data is less than Median of Data.

What is nature of skewness of the data?

Answer – the nature of skewness is left or negative skewness.

What will be the IQR of the data (approximately)?   
Answer – IQR = Upper Quartile – Lower Quartile

IQR = 18 – 10 = 8

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

Answer –

1. Both Boxplots are normally distributed.

2. Median is equal for both datasets.

3. No outliers available in both boxplots.

4.Data set against boxplot 1 is having positive kurtosis whereas data sets against boxplot 2 is having negative kurtosis.

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) = 0.3475
  2. P(MPG<40) = 0.7293

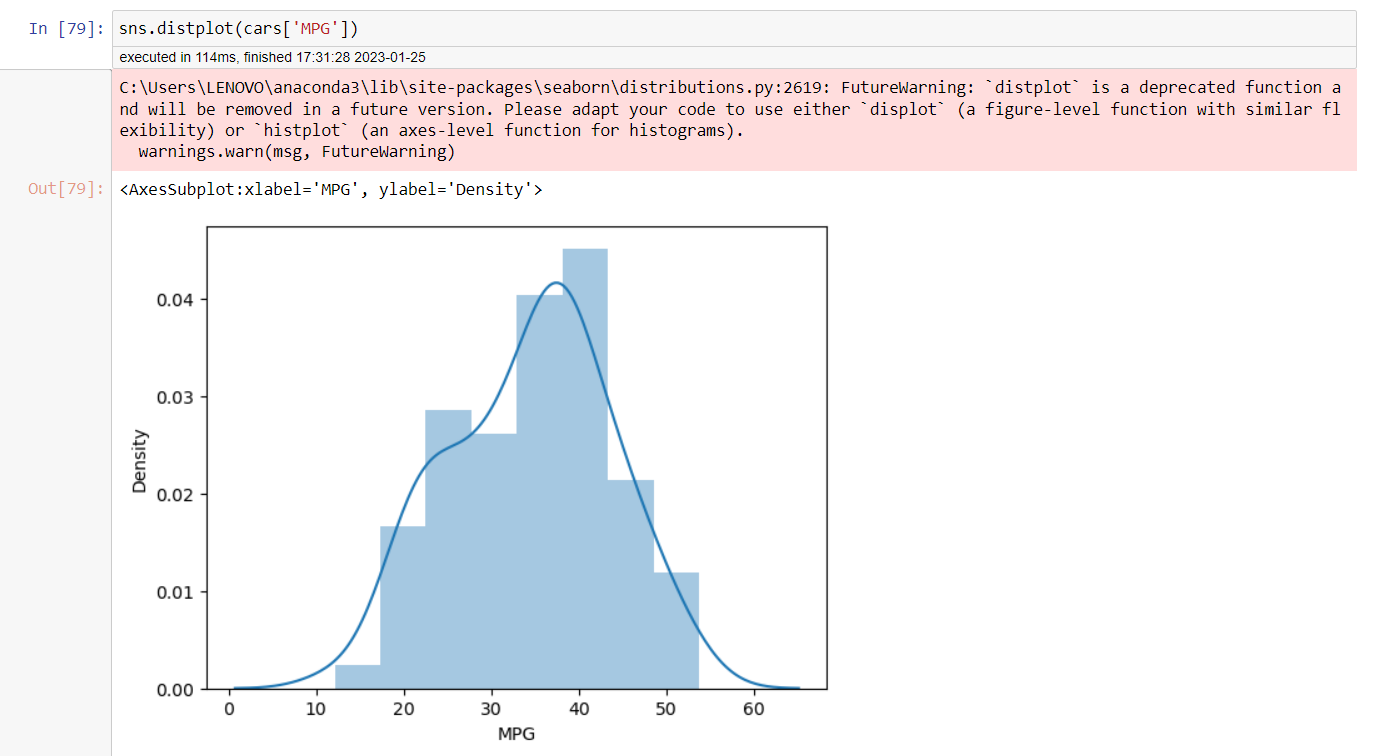
c. P (20<MPG<50) = 0.8988



Q 21) Check whether the data follows normal distribution

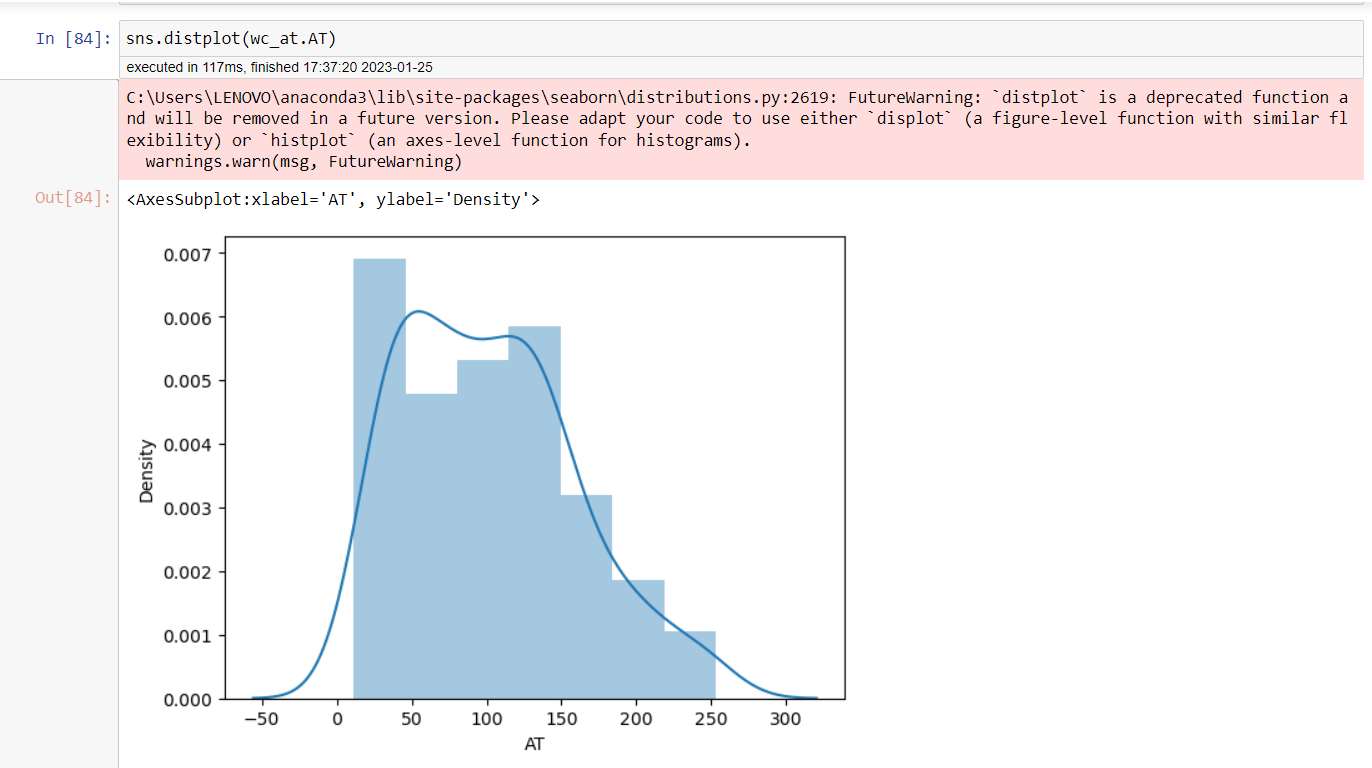
1. Check whether the MPG of Cars follows Normal Distribution

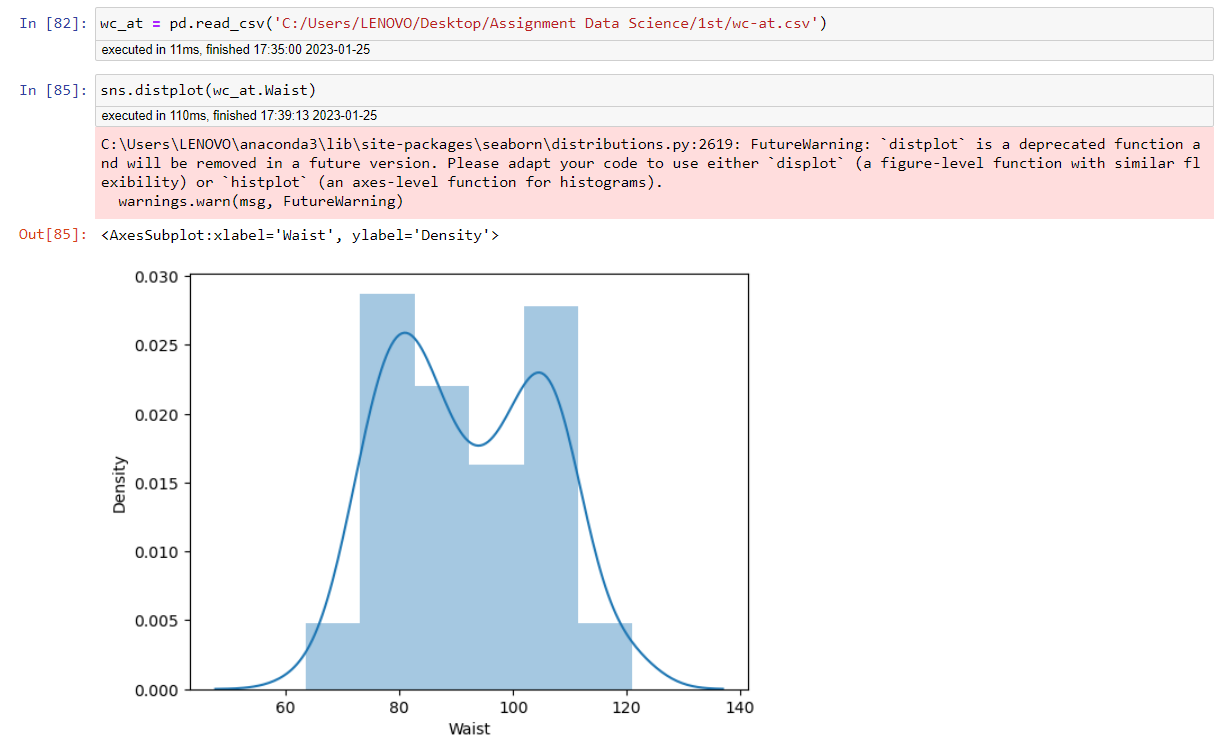
Dataset: Cars.csv



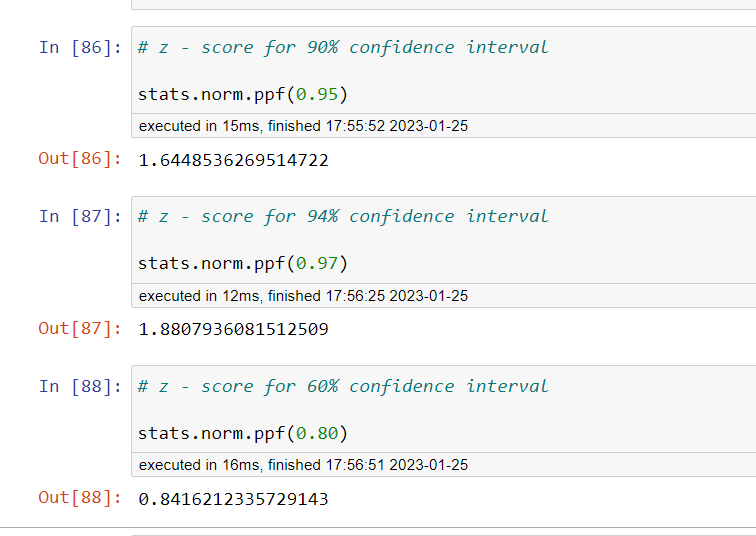
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

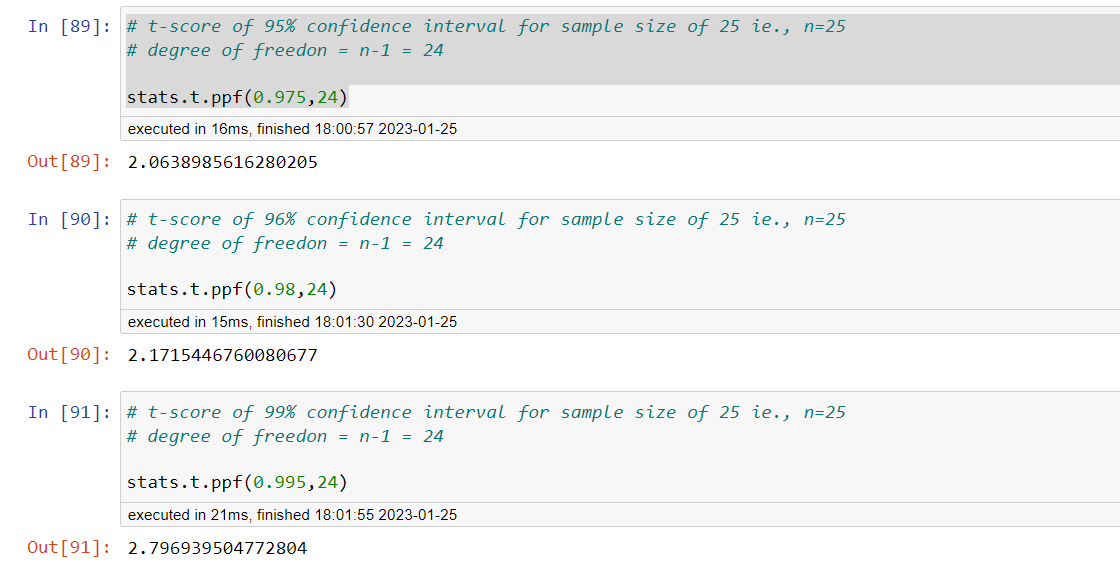




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

