| Activity | Data Type |
| --- | --- |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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 | Discreate |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Discreate |
| Gender (Male or Female) | Discreate |

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 Scale |
| 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 | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | ordinal |

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

**Sol:**- the possible outcomes HHH,HTH,HHT,HTT,TTT,THT,TTH,THH

TOTAL OUTCOMES= 8

P(x) = Number of outcomes having two head and 1 tail

                              = 3 (as HHT, HTH, THH are having at least one head).

**P(x)=3/8**

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

**Sol:-**

1. If two dice are rolled, then total possible outcome = 36

Total Favourable outcomes (having sum 1) = 0

**P(x) = 0/36**

**= 0**

1. Total outcome =36

favourable outcomes for sum is less then or equal to 4 = 6[which is (1,1),(1,2),(1,3),(1,2),(2,2),(1,3)]

**P(x)= 6/36 = 1/6**

1. Favourable outcomes = sum is divisible by 2 and 3

= 5[which is (5,1),(4,2),(3,3),(5,1),(6,6)]

**P(x) = 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?

Sol :- Total num of balls = 2+3+2 = 7

S = Two balls are drawn at random

n(s) = 7C2 = (7\*6)/2\*1 = 21

X = None of balls drawn is blue

n(x)= 5C2 = (5\*4)/2\*1 = 10

**P(x) = 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

Sol:- Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=    3.090

**Expected number of candies for a randomly selected child  = 3.090**

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

* For Points,Score,Weight>

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

**Use Q7.csv file**

**Sol-**

|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard Deviation** | **Range** |
| --- | --- | --- | --- | --- | --- | --- |
| **POINT** | 3.59 | 3.659 | 3.07 | 0.2858 | 0.534679 | 2.17 |
| **SCORE** | 3.21 | 3.325 | 3.44 | 0.9573 | 0.9784 | 3.91 |
| **WEIGHT** | 17.84 | 17.84 | 17.02 | 3.1931 | 1.78 | 8.39 |

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?

**Sol:-**

**X=** 108, 110, 123, 134, 135, 145, 167, 187, 199

**There are 9 patients so**

| X | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P(X) | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

Expected value = ∑P(X).X

**Expected Value of the Weight of that patient=** (1/9)(108) + (1/9)110  + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

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

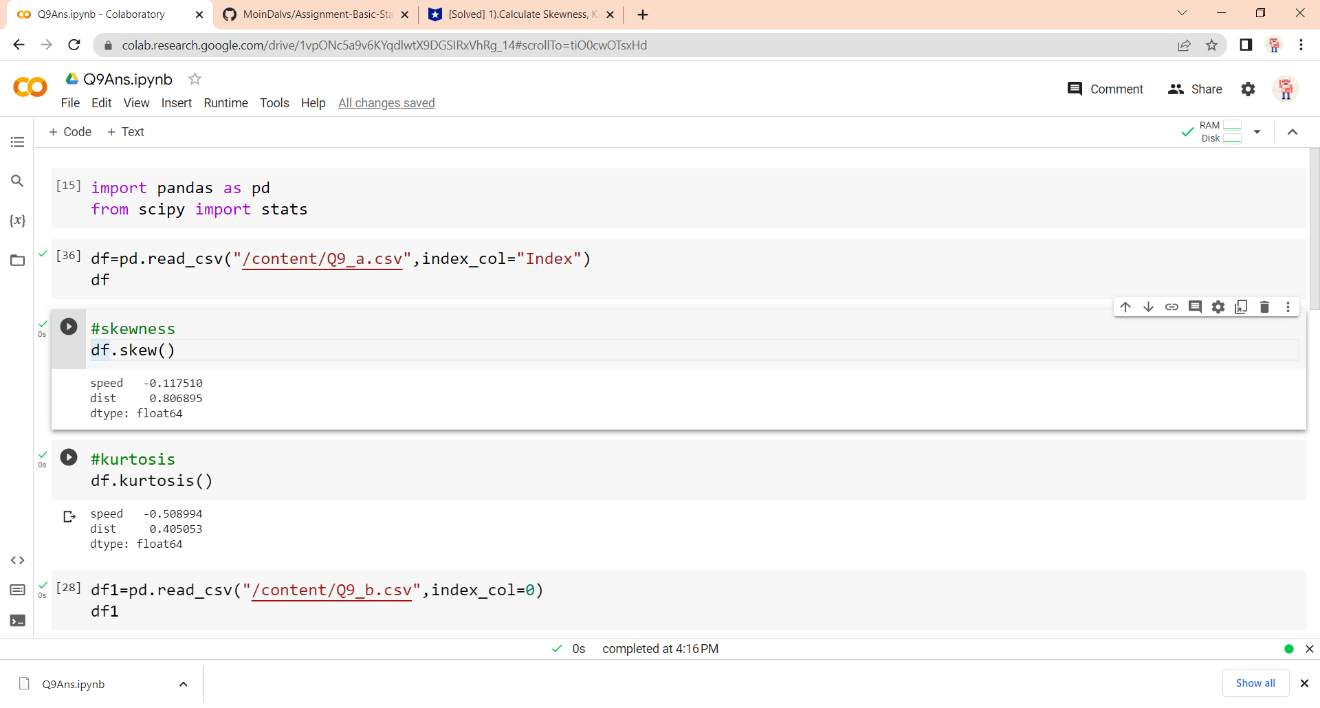
= 1/9(1308)

**= 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

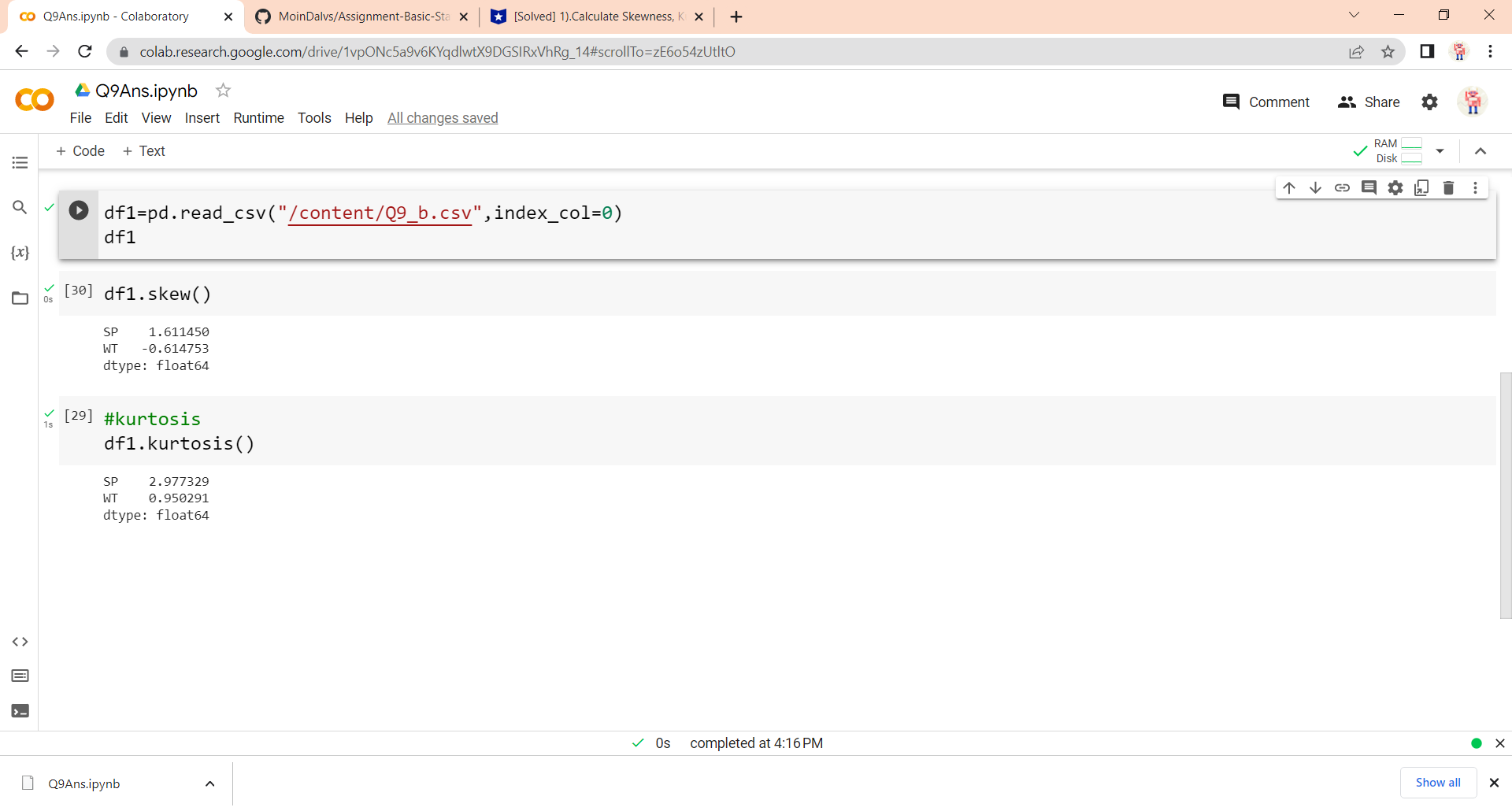
**Sol:-** 

|  | **Skewness** | **Kurtosis** |
| --- | --- | --- |
| **Speed** | **-0.12** | **-0.51** |
| **Distance** | **0.81** | **0.41** |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  | **Skewness** | **kurtosis** |
| --- | --- | --- |
| **SP** | **1.6114** | **2.9773** |
| **WT** | **-0.6147** | **0.9502** |



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



**Sol:-** This data is right-skewed because it’s longer on the right side of its peak. There is long tail on right. Most of data is between 50 - 200. Right skew: mean > median



**Sol:-** The boxplot has outliers at upper side and the data is right skwed data.

**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?

Sol:- population= 3,000,000

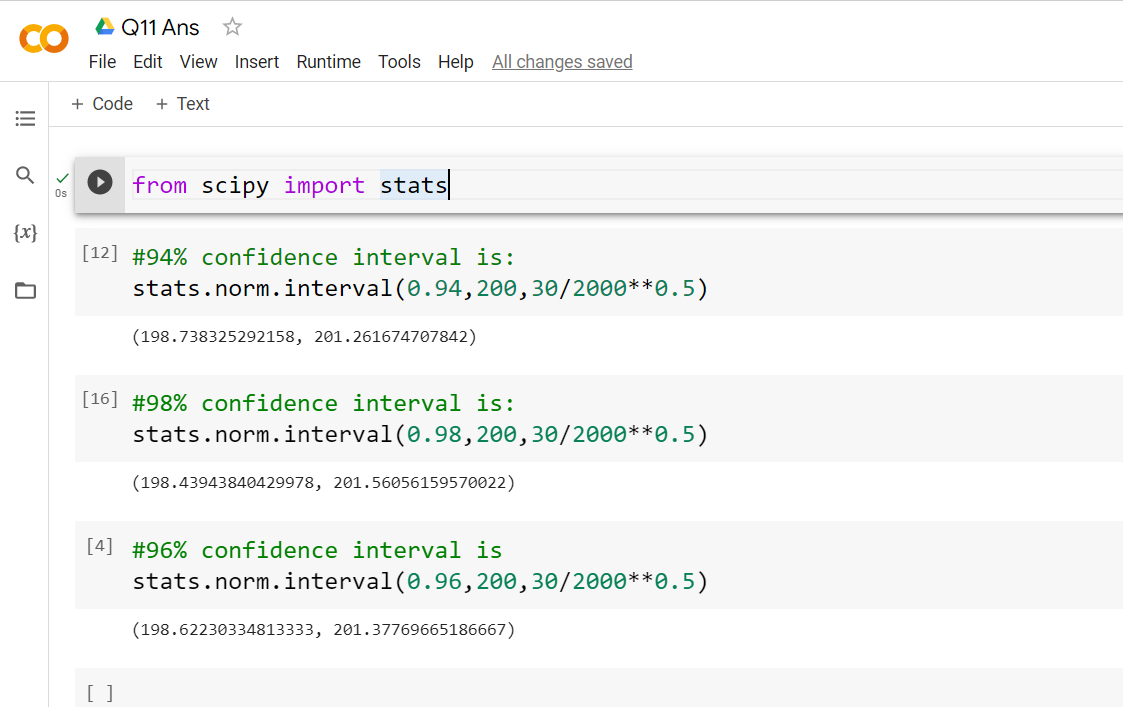
Sample size(n)= 2000

Sample mean(x̄)= 200

Standard deviation(s)= 30

Using the t – distribution

x̄ ± t



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

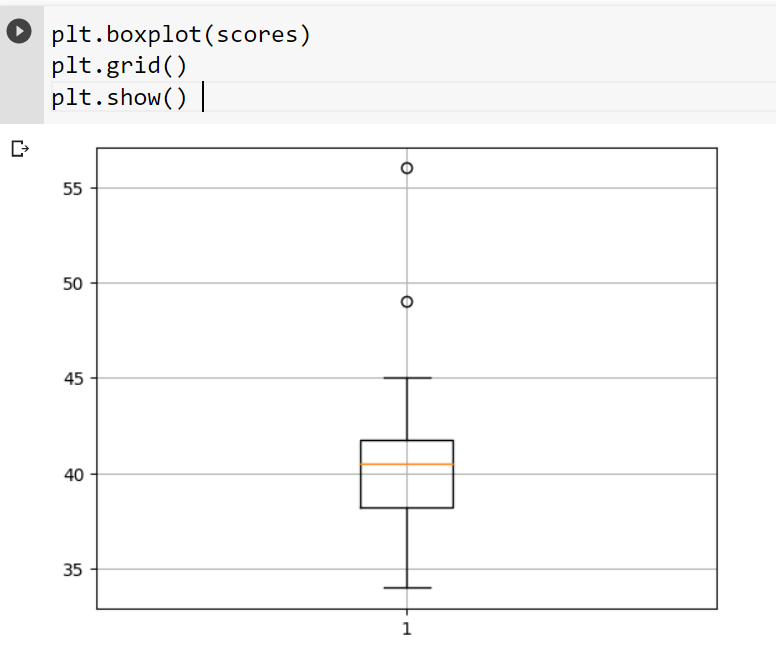
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

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

**Sol:-**

| **mean** | **median** | **variance** | **Standard deviation** |
| --- | --- | --- | --- |
| **41** | **40.5** | **25.6** | **5.05** |

1. What can we say about the student marks?

Sol:-

From above plot we can say that most of the student in between 41-42. And there are two outliers 49 and 46.

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

**Sol:-** the mean is equal to median and the distribution has zero skewness.

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

**Sol:-** The distribution is right-skewed because it’s longer on the right side of its peak. There is a long tail on the right.

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

**Sol:-** The distribution is left-skewed because it’s longer on the left side of its peak. The [mean](https://www.scribbr.com/statistics/central-tendency/#mean) of a left-skewed distribution is almost always less than its [median](https://www.scribbr.com/statistics/central-tendency/#median).

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

**Sol:-** Positive excess values of kurtosis (> 3) indicate that distribution is peaked and possesses thick tails. Leptokurtic distributions have positive kurtosis values.

A leptokurtic distribution has a higher peak (thin bell) and taller (i.e., fatter and heavy) tails than a normal distribution.

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

**Sol:-** Negative excess values of kurtosis (<3) indicate that the distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

**Ans:-** The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

**Ans:-** The whisker range of minimum value is greater than maximum and the data is a skewed towards left

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

**Sol:-** IQR= UQ – LQ

= 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:-** We can say that they both have same median and and there are no outliers. they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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) = 0.3474
  2. P(MPG<40) = 0.7294
  3. P (20<MPG<50) = -0.0570

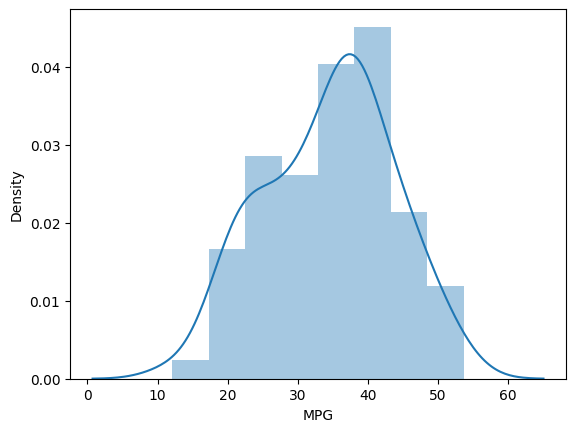
Ans:- Q20Ans.pynb files

**21)** Q Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

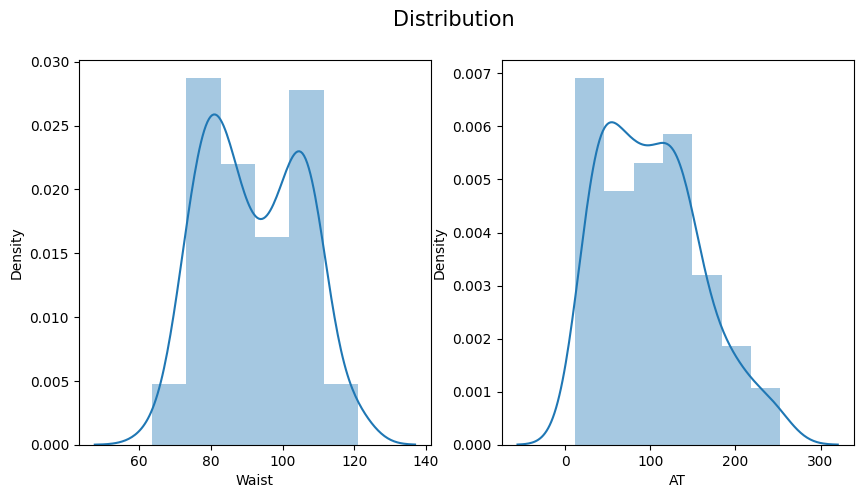
**Ans:-** MPG of cars follows normal distribution



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

Dataset: wc-at.csv

**Ans:**- Adipose Tissue (AT) and Waist does not follow Normal Distribution



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

Ans:- Z score of 90% CI:- 1.6449

Z scores of 94% CI:- 1.880

Z scores of 60% CI:- 0.8416



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

**Ans:-** t scores of 95% confidence interval for sample size of 25 :- 2.063

t scores of 95% confidence interval for sample size of 25 :- 2.1715

t scores of 95% confidence interval for sample size of 25 :- 2.7969

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

Ans:- µ= 270 , n=18 , x̄= 260 , s = 90

H0: average life of bulb>= 260

Hα: average life of bulb< 260

So P>α So So we reject H0 .