**Student Name:** Kiran Balasaheb Tandale

**Batch:** Data Science weekday Batch – Data science Weekday-Aishwarya-Pune-24th February 2023-01:00 PM to 03:00 PM

**Assignment Name:** Basic Statistics Level 1 (Assignment No.1)

|  |  |
| --- | --- |
| 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) | 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) | 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, Interval |

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

Answer: If three coins are tossed; there are = 8 possible outcomes. Those are

( HHH, HHT, HTT, HTH, TTT, TTH, THH, THT )

So, 3 possible outcomes give two head and one tail is (HHT, HTH, THH)

i.e probability is 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: a) If 2 dice are rolled then total possible outcomes are 6\*6=36

Minimum sum of 2 occurrences is 1+1 = 2

So, Probability of getting sum equal to 1 is 0

b) If two dice are rolled then total possible outcomes are 6\*6 = 36

Chances of getting a sum less than or equal to 4 are (2,2)(1,3)(3,1)

So, Probability of getting sum less than or equal to 4 is

(interested events)/(total events) i.e 3/36

c)If 2 dice are rolled than total possible outcomes are 6\*6 =36

Chances of getting a sum divisible by 2, 3 are (4,2)(2,4)(1,5)(5,1)(6,6)(3,3)

So, Probability of getting sum divisible by 2 and 3 is (interested events)/(total events) i.e 6/36

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 = 7

N (2 balls are drawn randomly from bag) =7! / 2! \* 5!

=(7\*6\*5\*4\*3\*2\*1)/(2\*1)\*(5\*4\*3\*2\*1)

N (Event (2 balls are drawn randomly from bag) = (7\*6)/(2\*1) =21

If none of them drawn 2 balls are blue =7-2 = 5

N (None of the balls drawn is blue) = 5! / 2! \* 3! = (5\*4) / (2\*1)

= 10

P (None of the balls drawn is blue) = N(Event (None of the balls drawn is blue)/

N(Event (2 balls are drawn randomly from

bag)

= 10/21

p= 0.476

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 : 0.0015+0.8+1.95+0.025+0.06+0.24 = 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**

**Answer :**

Mean of Points = 3.59, Score = 3.21 and Weight = 17.84

Median for Points = 3.69, Score = 3.32 and Weight = 17.71

Mode for Points = 3.07, Score = 3.44, and Weight = 17.02

Variance for Points = 0.28, Score 0.95, Weight = 3.19

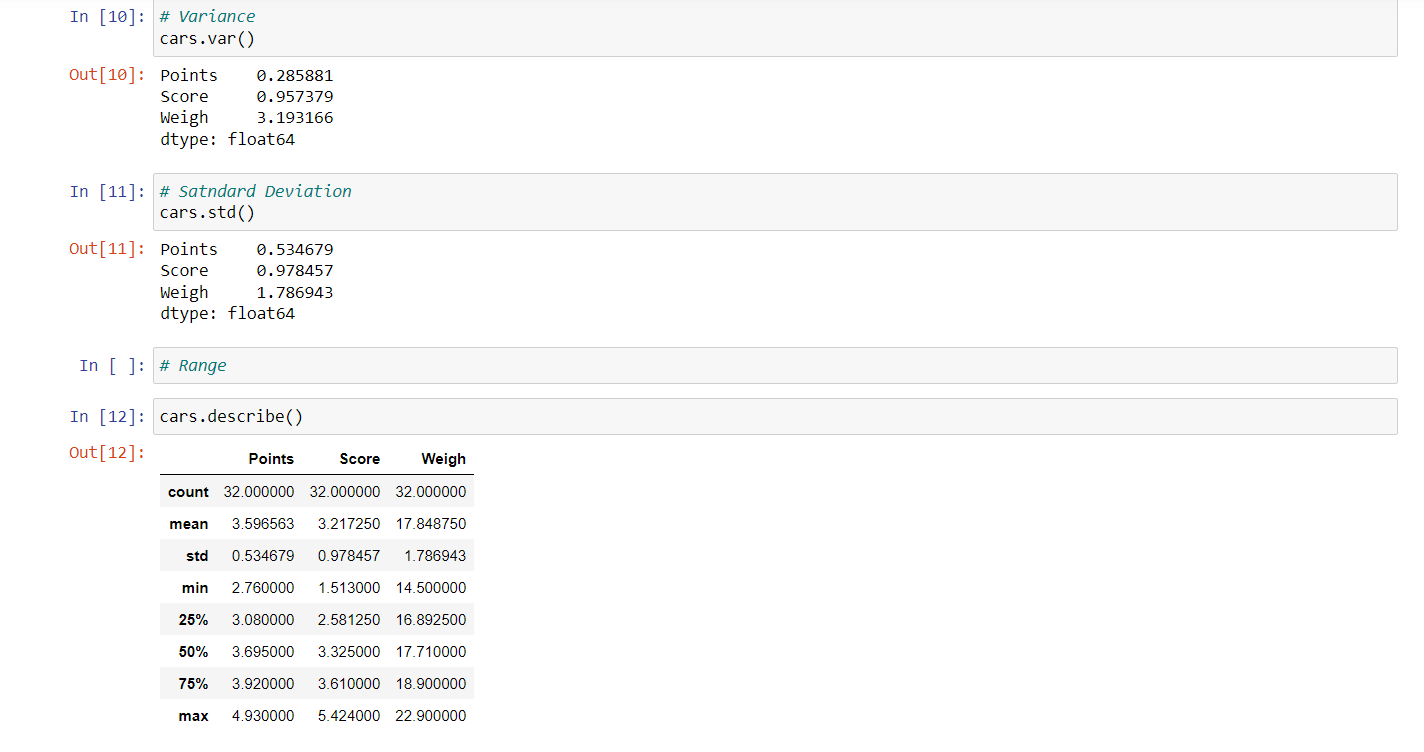
Standard Deviation for Points = 0.53, Score = 0.97, Weight = 1.78

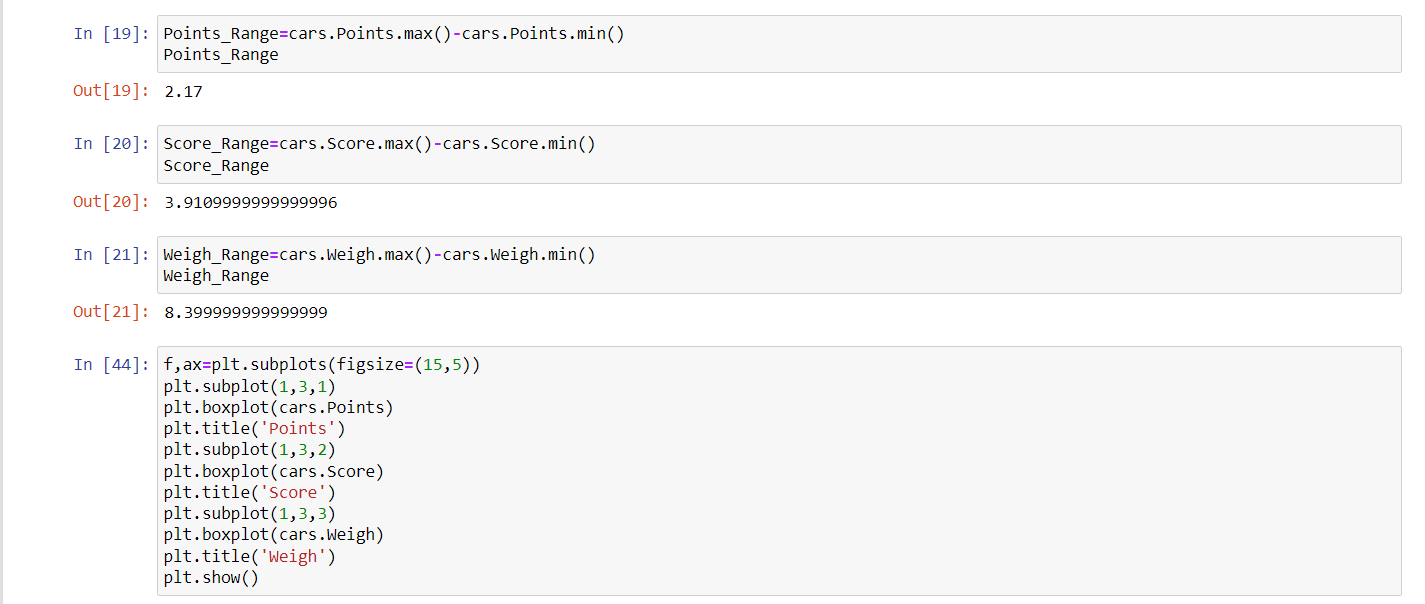
Range [Min-Max] for Points [3.59-4.93], Score [3.21-5.42] and Weight [17.84-22.9

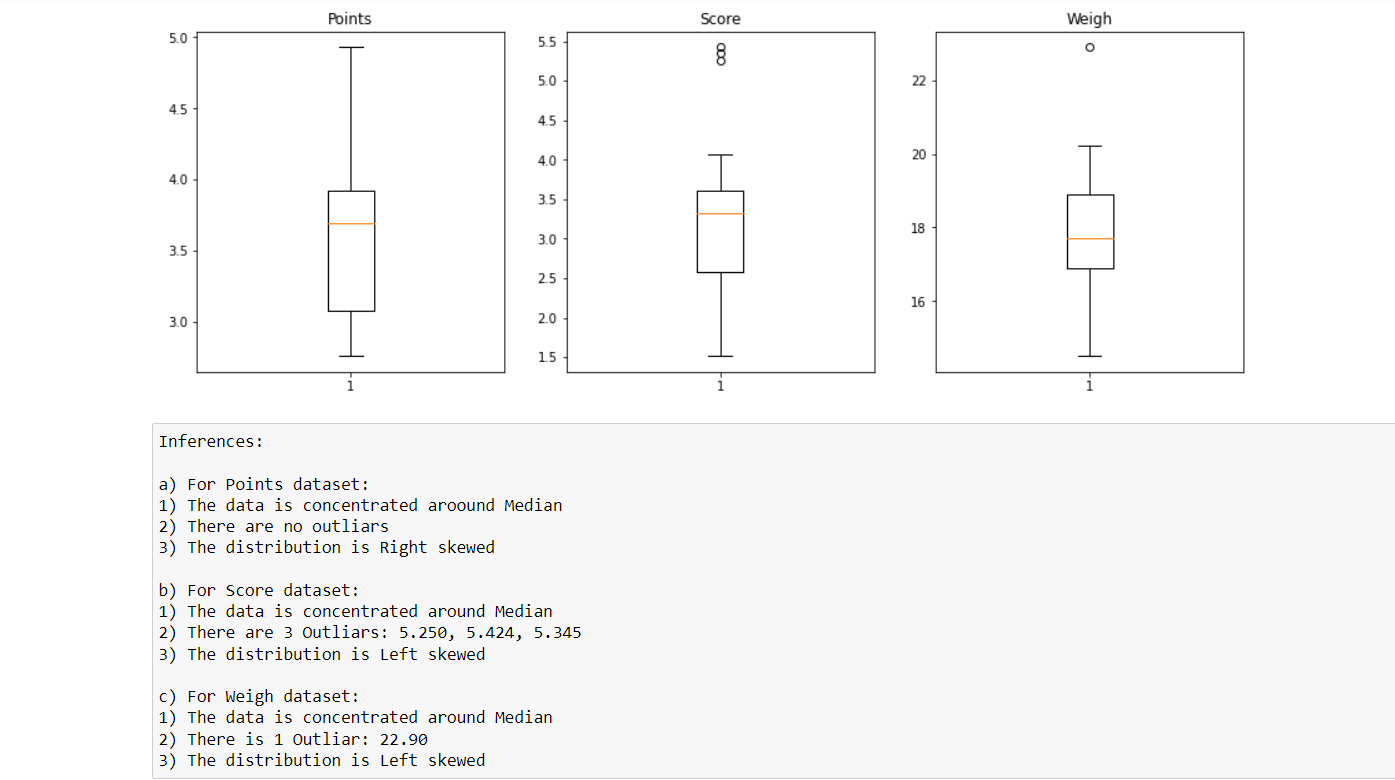
Draw Inferences

From above values of points , score, weight of mean, median, mode is likely same so when mean=median=mode then distribution is normal distribution.









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 : Expected value = Sum (X \* Probability of X)

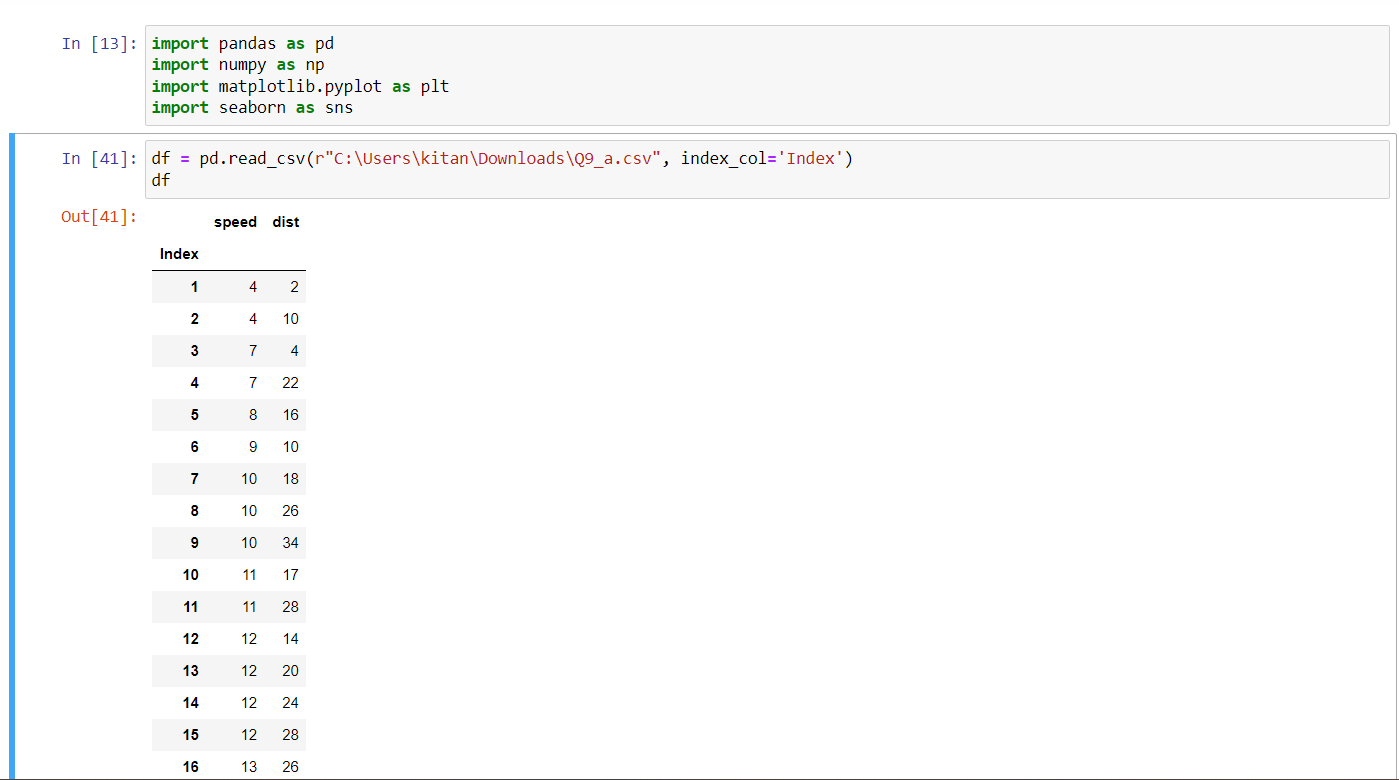
= (1/9)(108)+(1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+

(1/9)(187)+ (1/9)(199)

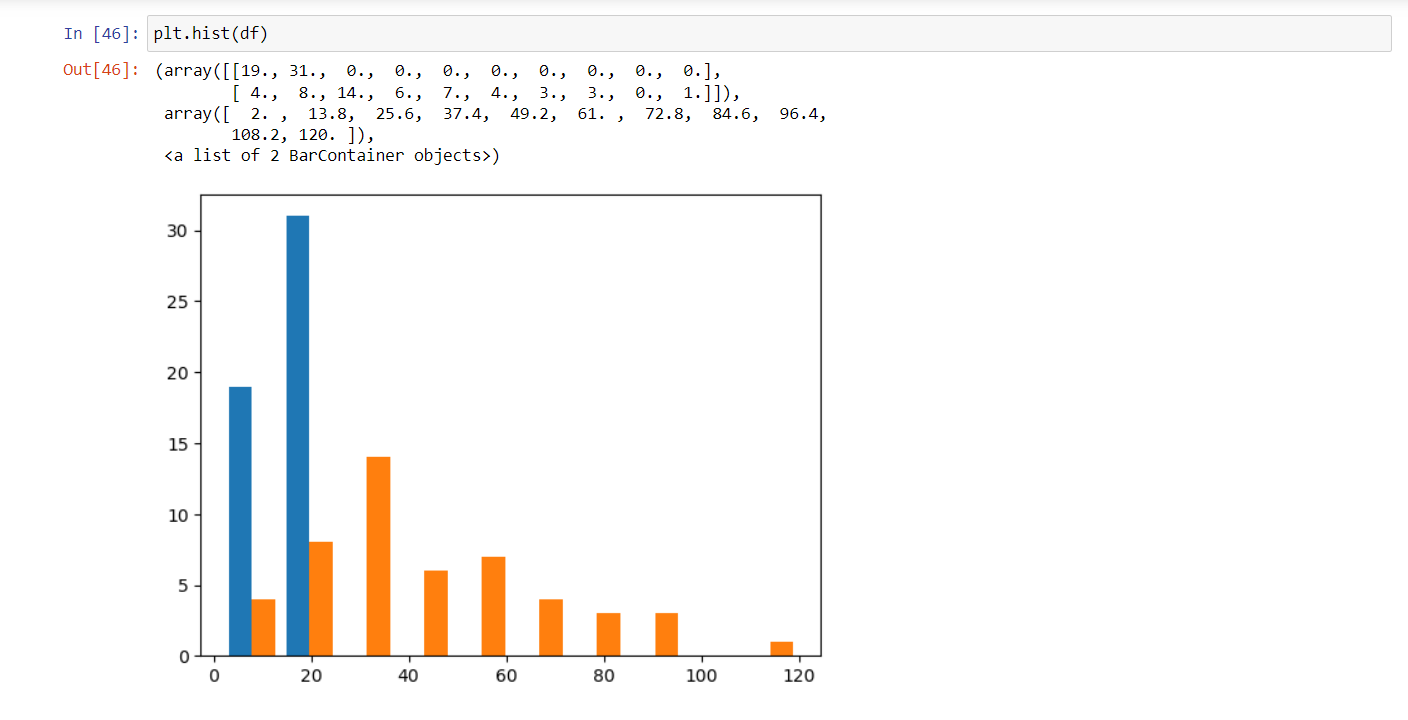
= 145.33

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

**Cars speed and distance**



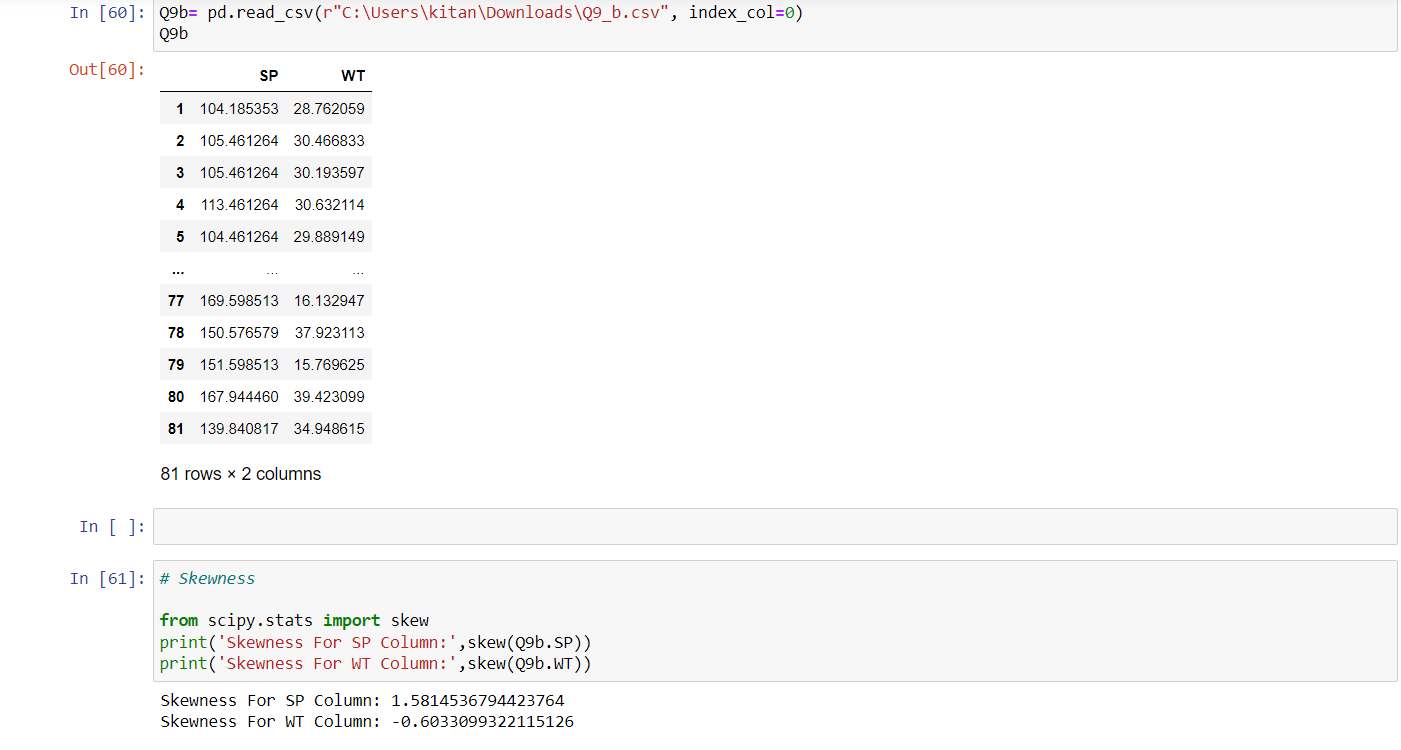


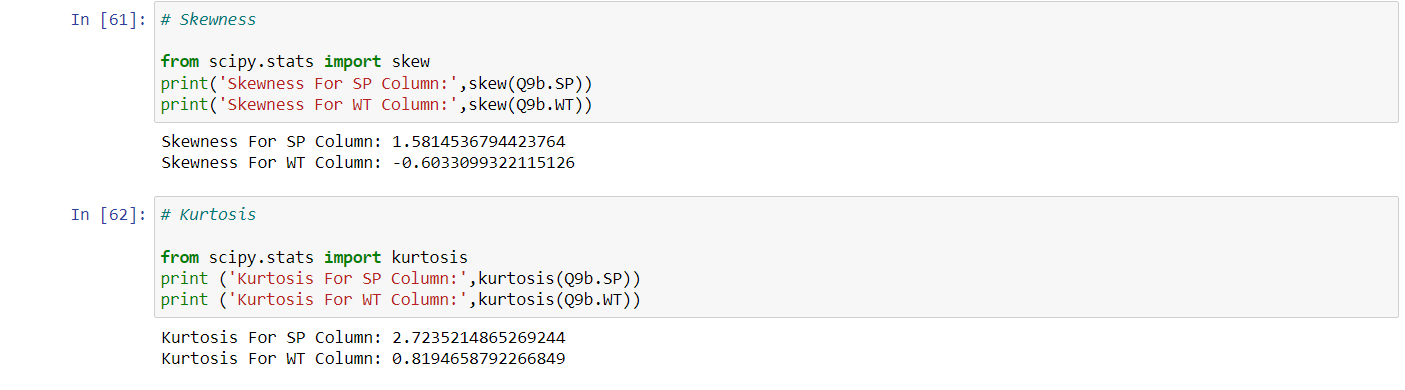
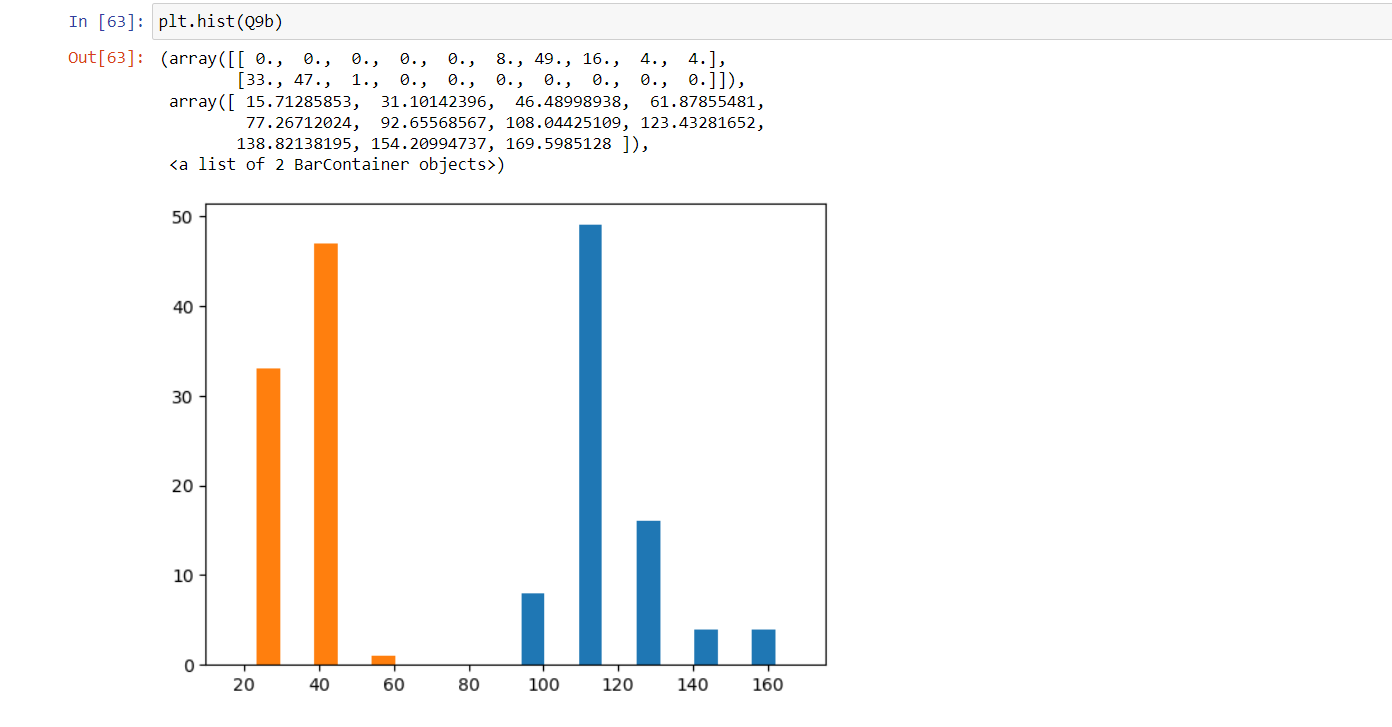


**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**





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



Answer: The histogram peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

**Ans :**  Histogram is positively skew. We can predict that there is outlier above upper extreme. Most of data point is present in between 50-150 wt.



Answer : The boxplot has outliers on the maximum side.

**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 : conf\_94 stats.t.intarval(alpha = 0.94, df =1999, loc=200,scale=30/np.sqrt(2000))

Print(np.round(conf\_94,0))

Print(conf\_94)

For 94% confidence interval Range is [198.73 – 201.26]

For 98% confidence interval Range is [198.43 – 201.56]

For 96% confidence interval Range is [198.62 – 201.37]

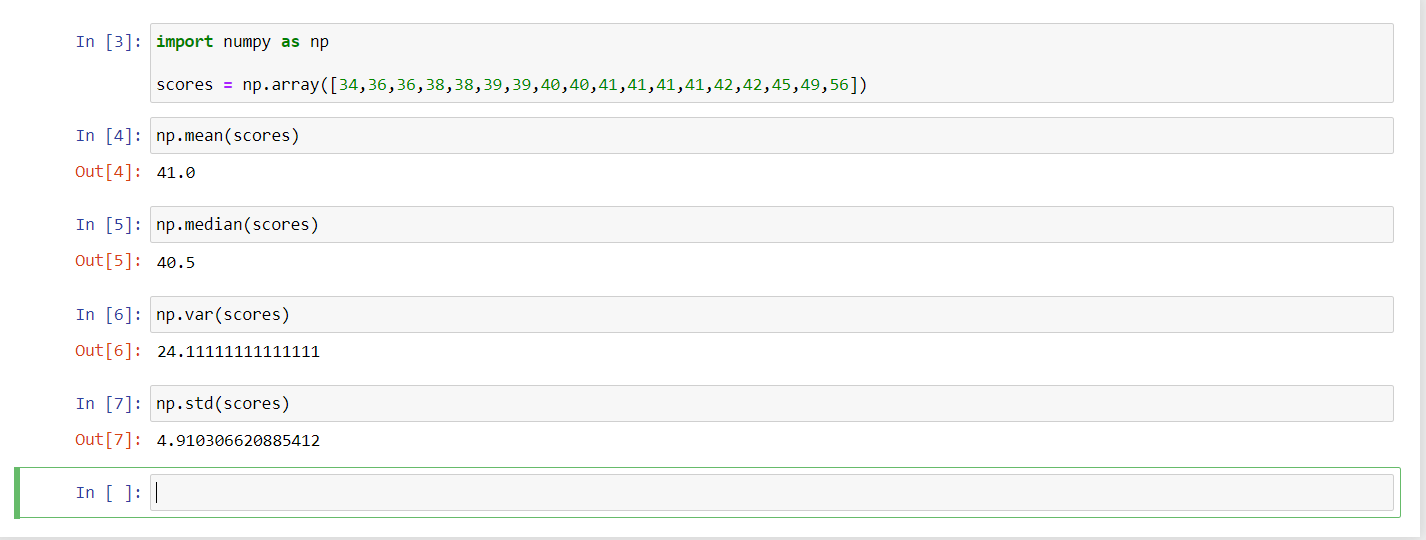
**Q12)** Below are the scores obtained by a student in tests

**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.

Answer: Mean =41,Median =40.5,Vatiance =25.52, and Standard Deviaton =4.91

1. What can we say about the student marks?

Answer: we don’t have outliers and the data is slightly skewed towerds right because mean is greater than median.

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

Ans : No skewness is present we have a perfect symmetrical distribution.

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

Ans: Positive Skewness and tail is towards Right.

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

Ans: Negative Skewness and tail is towards left.

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

Ans: Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Ans: Negative Kurtosis means the curve will be flatter and broader.

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 data is a skewed towards left. The whisker range of minimum value is greater than maximum.

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

Ans: The Inner Quantile Range = Q3 Upper Quartile – Lower Quartile = 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: i) The data is normally distributed because the whisker of both of the sides are symmetrical. Thus, there is no skewness (both boxplot 1 & 2).

ii) There is no outliers in the data (both boxplot 1 & 2)..

iii) Boxplot 1 & 2 has same median.

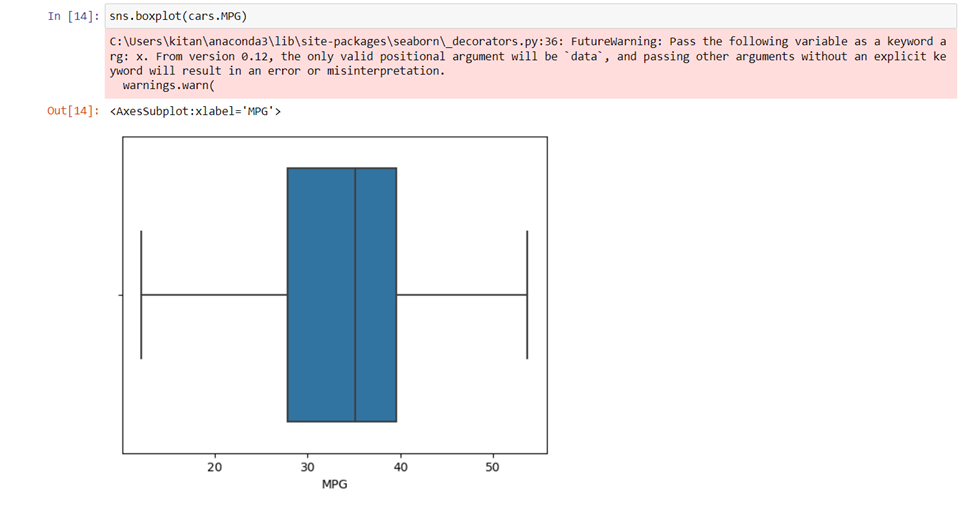
Q 20) Calculate probability from the given dataset for the below cases

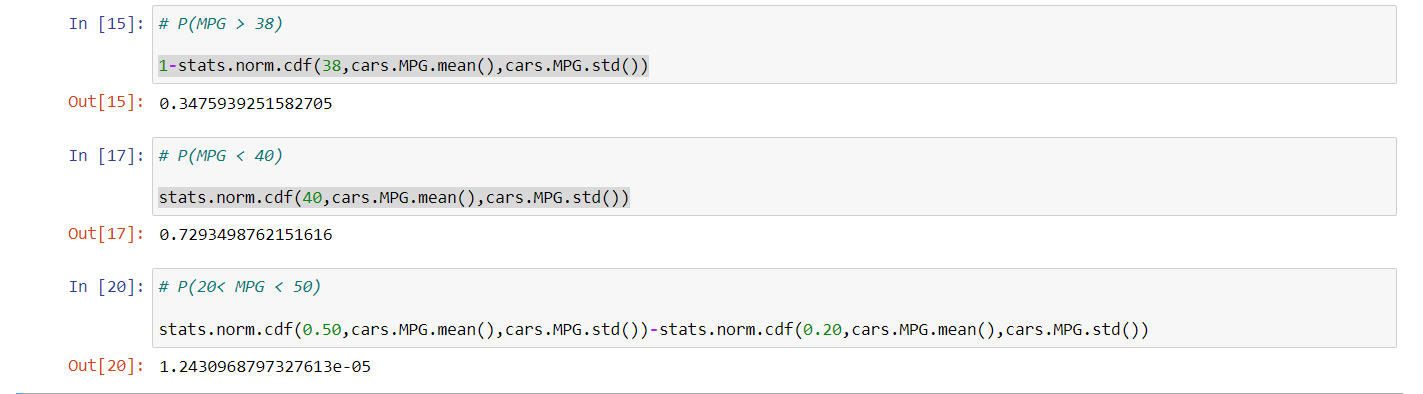
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

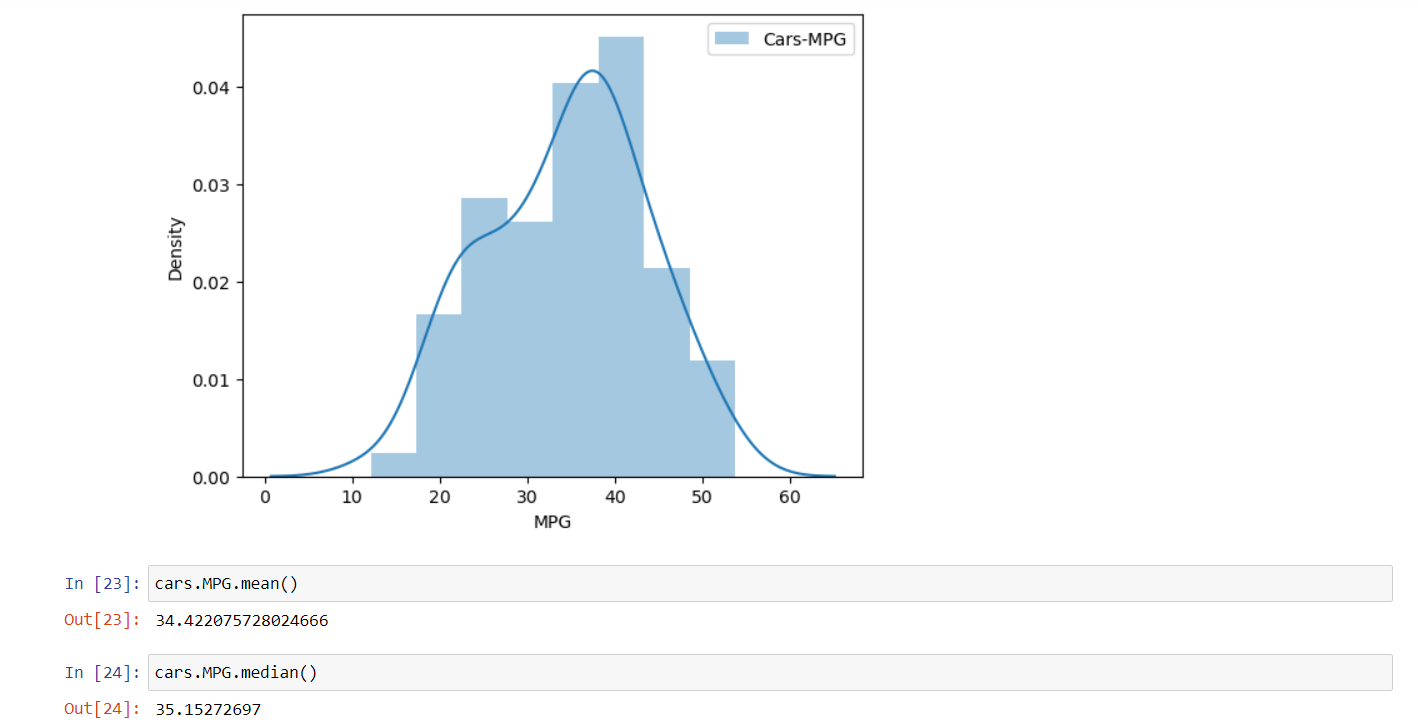




Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

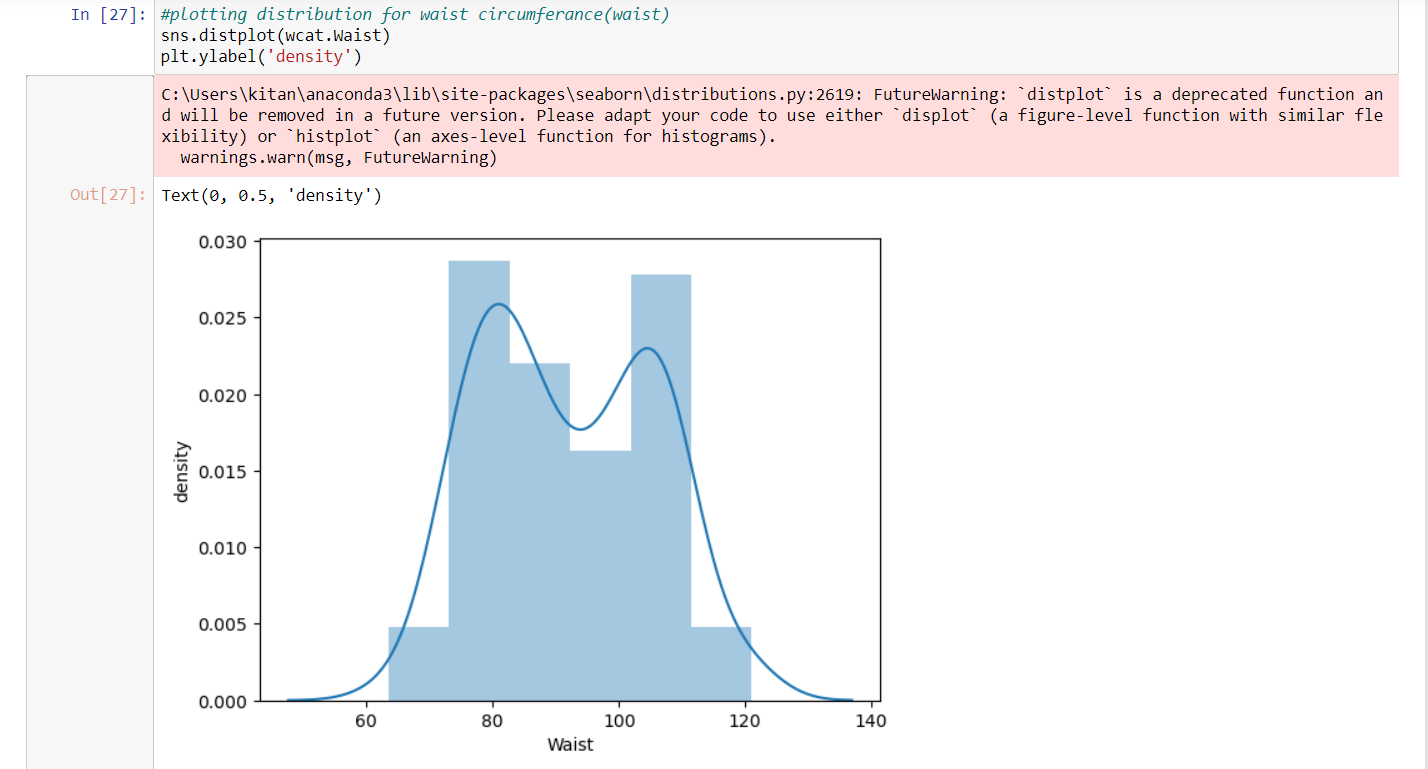
 Dataset: Cars.csv

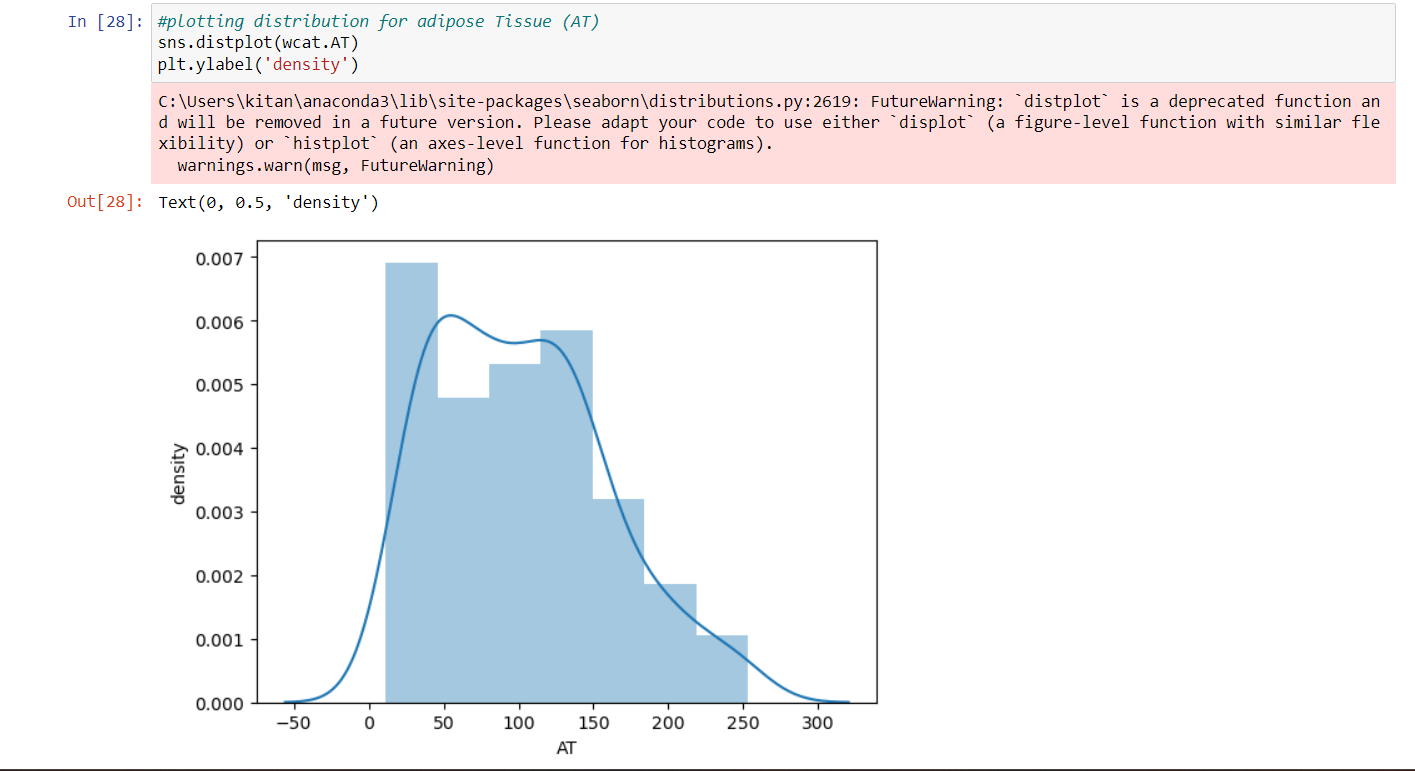


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv









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

Answer :

1. Confidence Interval For 90% = Z-Score = 1.645

2) Confidence Interval For 94% = Z-Score = 1.88

3) Confidence Interval For 60% = Z-Score = 0.84

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

1. Confidence Interval For 95% = T-Score = 2.063

2) Confidence Interval For 96% = T-Score = 2.17

3) Confidence Interval For 99% = T-Score = 2.796

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

