|  |  |
| --- | --- |
| 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 | 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 | ordinal |
| SAT Scores | interval |
| Years of Education | Ratio |

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

Ans: Probability:number of fav outcomes/total number of outcomes

(THH,HTH,HHT) 2^3=3/8

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

1. Equal to 1

1=0

1. Less than or equal to 4

Number of fav outcomes/total number of outcomes=6/36=1/6

1. Sum is divisible by 2 and 3

=5/38

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?

Ans: number of ways of drawing 2 balls=7c2 =(7\*6)/(2\*1)=21

No od 2 blue balls

n(E) =2+3=5

5c2=(5\*4)/(2\*1)=10

P(E)=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

Expected value=summation of P(E)\*N

(1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)=

0.015+.8+1.95+.025+.06+.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.

MEAN

Points=3.596563

Score=3.217250

Weigh=17.848750

MEADIAN

Points=3.695

Score=3.325

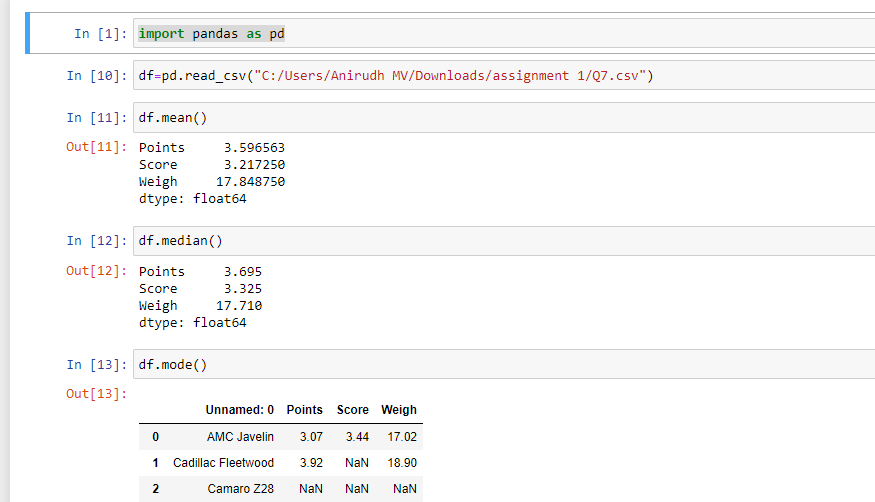
Weigh=17.710

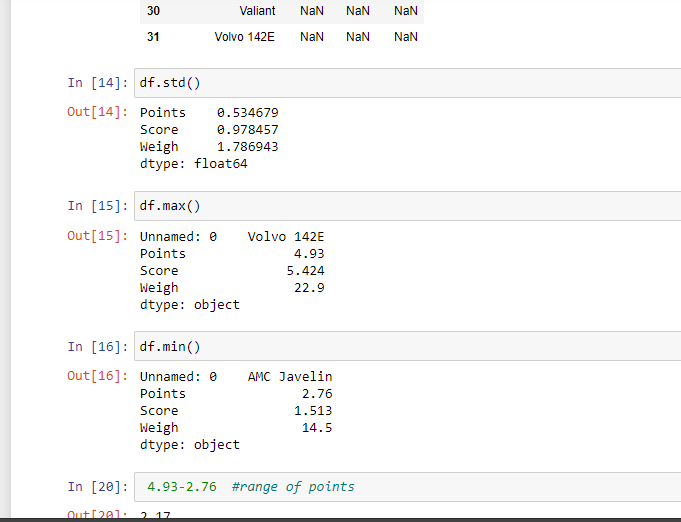
RANGE

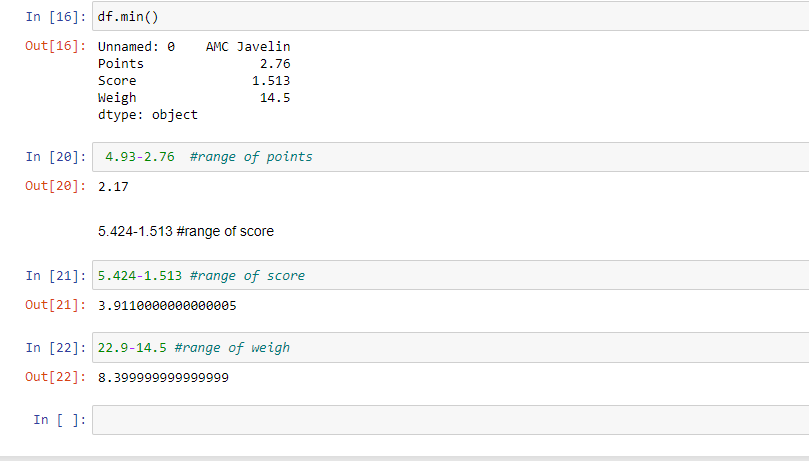
Points=4.93-2.76=2.17

Score=5.424-1.513=3.911

Weigh=22.90-14.50=8.4







**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?

Ans: Expected Value  =  sum of ( probability  \* Value )

 Sum of(P(x).E(x))

there are 9 patients

Probability of selecting each patient = 1/9

E(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

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

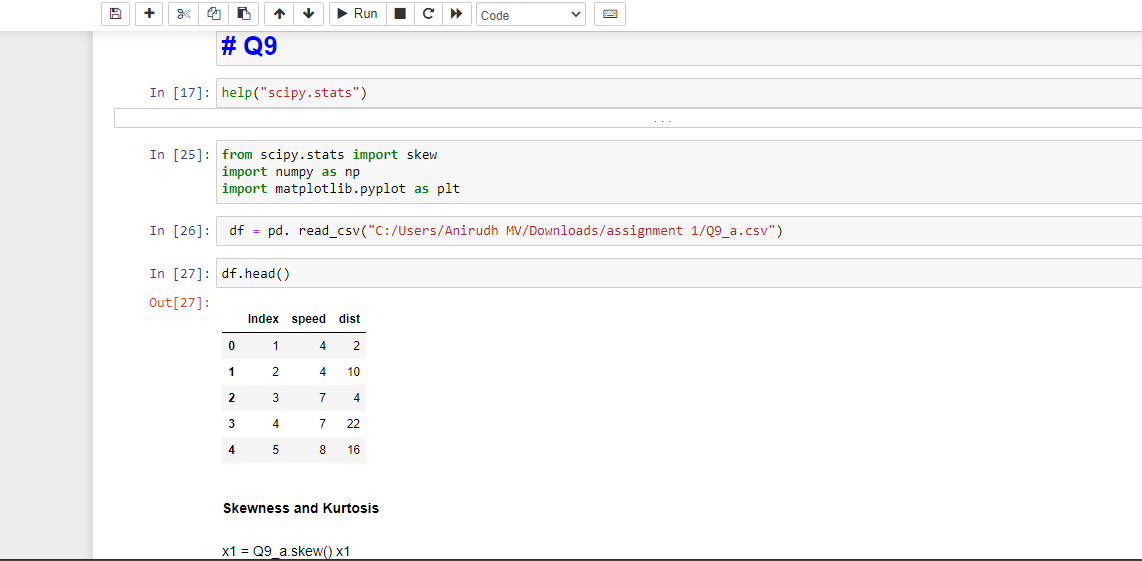
= (1/9) \*(1308)

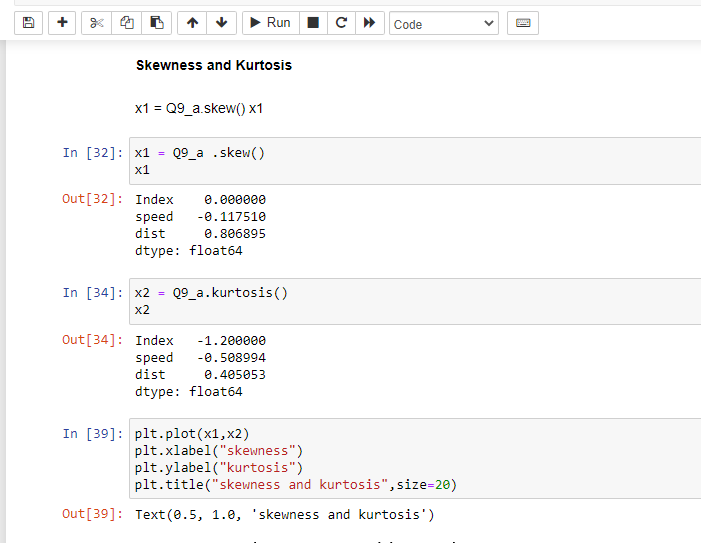
= 145.32

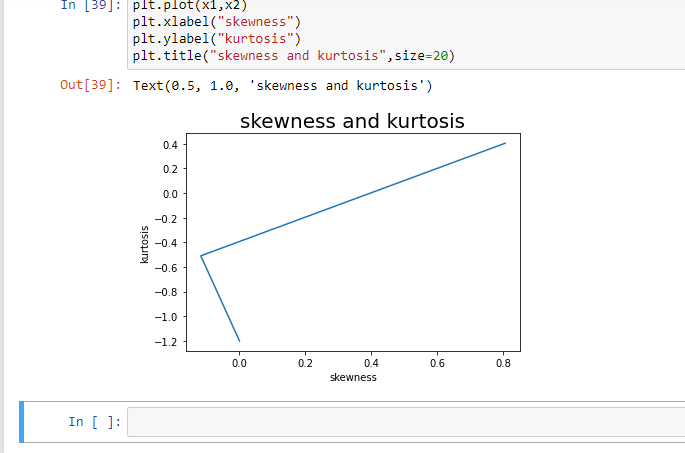
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Ans:**



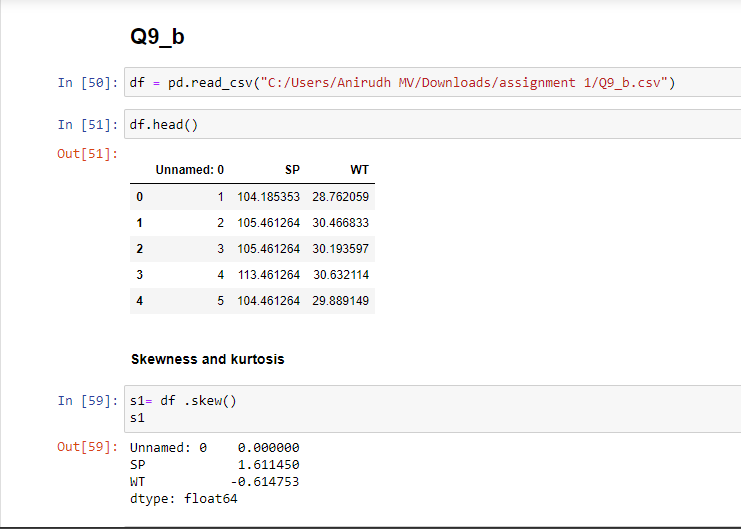


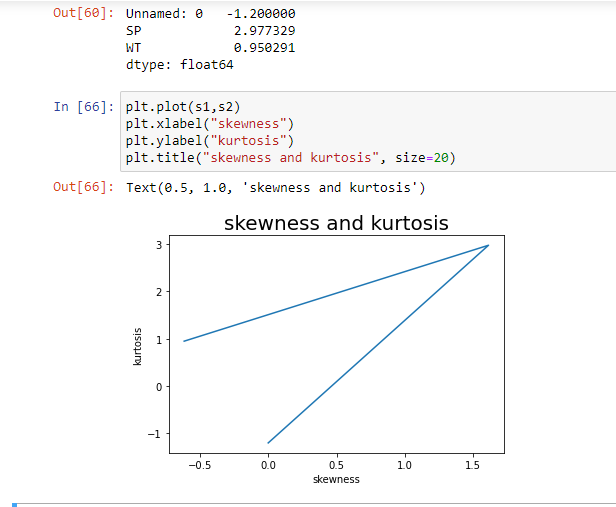


**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**





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



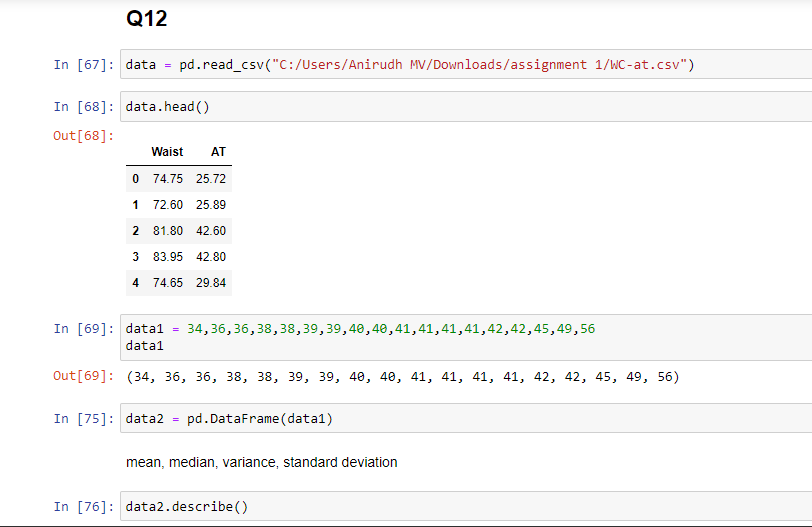
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

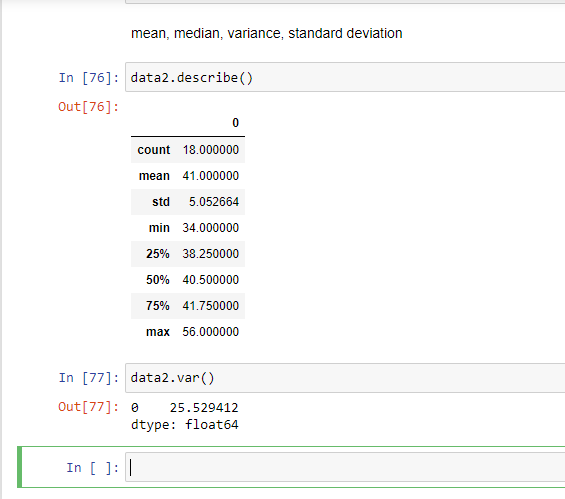
**Q12)** Below are the scores obtained by a student in tests

**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?

Ans:





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

Ans:Nature of skewness can be positive, negative or zero. when the value of a mean, median and mode are equal then there is no skewness.

Skewness is a measure of the asymmetry of probability distribution of real values of mean.

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

Ans:If the mean is greater than median, then the distribution is positively skewed.

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

Ans:If the mean is less than median, then the distribution is negatively skewed.

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

Ans:A distribution with a positive kurtosis value indicates that the distribution has heavier tail and sharper peak than the normal distribution.

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

Ans:A distribution with a negative kurtosis value indicates that the distribution has lighter tail and flatter peak than the normal distribution.

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



What can we say about the distribution of the data?

The distribution of data is mostly in between 10 to 18 and the median in somewhere near to 15.5.

What is nature of skewness of the data?

The nature of skewness is positively skewed.

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

IQR = Q^3 – Q^1

Approximately= 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.

The median of both the plot looks same and they don’t have outliers.

The quartile 1 for plot one is nearly 280 and for the plot two it is nearly 315 which shows plenty of difference between them.

The quartile 3 for plot one is approximately 260 and 218 for the boxplot 2 which therefore gives lot of difference.

Approximate IQR for Boxplot one = Q3 - Q1 = 260-280 = 20 and for boxplot two = Q3 - Q1 = 315-218 = 97.

Conclusion, The date in the boxplot one has lower variability than the data in boxplot two. 100% data of the boxplot one is less than 50% data of the boxplot two.

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

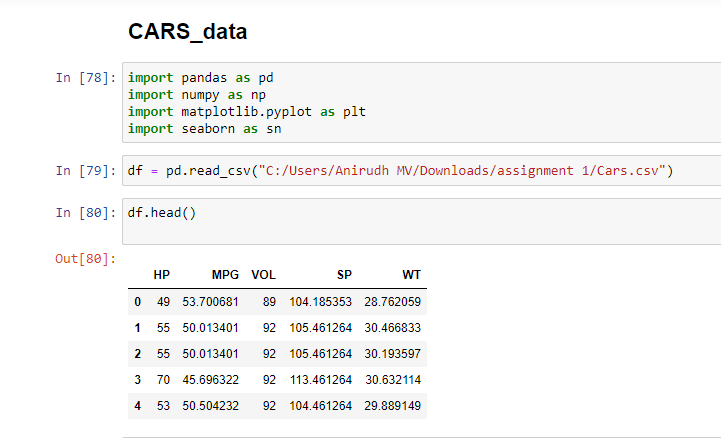
Data \_set: Cars.csv

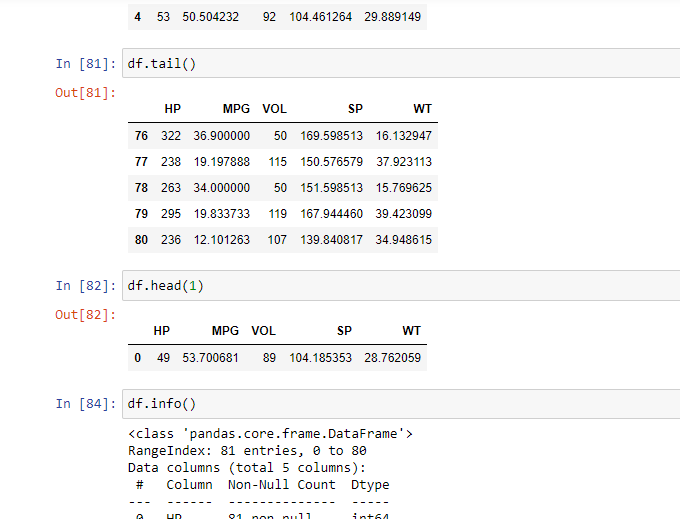
Calculate the probability of MPG of Cars for the below cases.

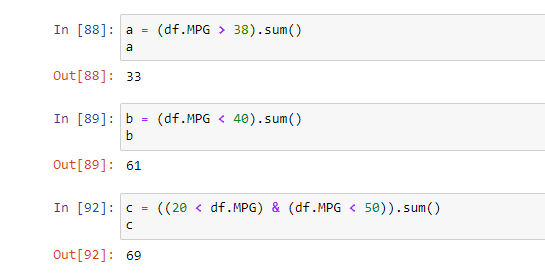
MPG <- Cars$MPG

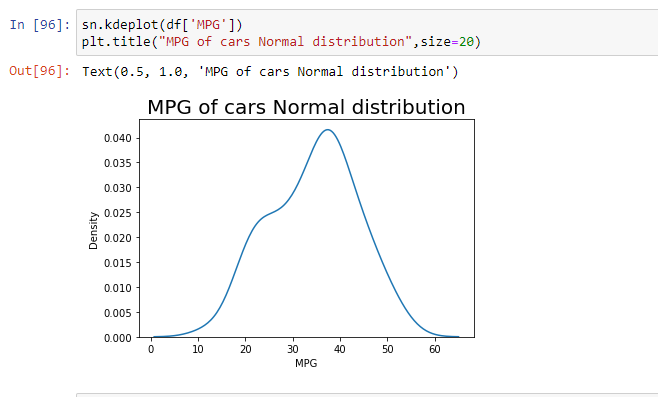
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans:





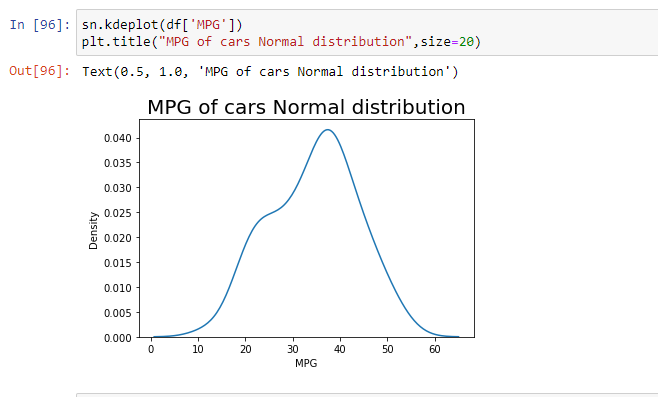




Q 21) Check whether the data follows normal distribution

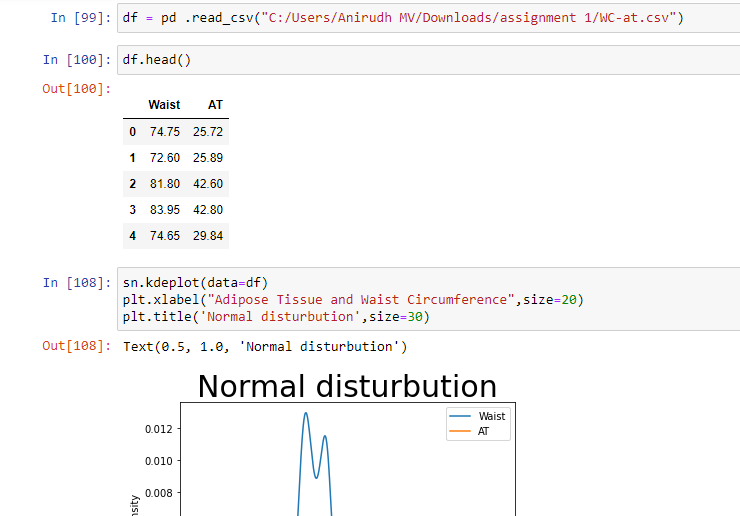
1. Check whether the MPG of Cars follows Normal Distribution

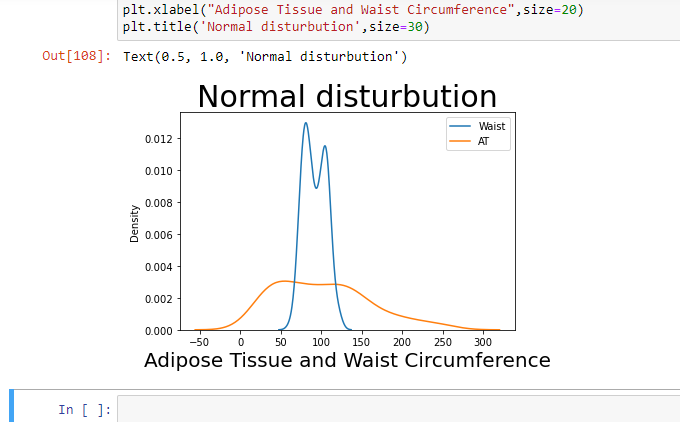
Dataset: Cars.csv



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

Ans:T = n-1 = 25-1 = 24

1. CI 95% = 0.95

(1-CI)/2 = (1-0.95)/2 = 0.025

T(24),0.025 = 2.064

1. CI 96% = 0.96

(1-CI)/2 = (1-0.96)/2 = 0.02

T(24),0.02 = 2.492

1. CI 99% = 0.99

(1-CI)/2 = (1-0.99)/2 = 0.005

T(24),0.005 = 2.797

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:

Mean sample of bulbs =x = 260

Population of mean = u = 270

Sample Standard Deviation = s = 90

Number of items in sample = n = 18

Formula: (x-u)/[s/sqrt(n)]

(260-270)/[90/sqrt(18)]

In R Language:

pt((260-270)/(90/sqrt(18)),18)