|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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) | Ratio |
| Sales Figures | Interval |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| 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?

Ans: Total cases for coin tossed = = 8 and two heads and one tail = 3{HTH,HHT,THH} so P(one tailed)/p(total case) = 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

Ans: a) 0 or Impossible Event

b) 6/36 or 1/6 (by counting)

c) P(N(sum is divisible by 2 & 3)) = 6 such cases

total case = 36 so the answer is 6/36 or 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?

Ans: since total ball presents are 2+3+2 = 7 and P(not blue) = 5/7

but for 2nd case it will be (7-1)6 and not blue will be (5-1)4 so the final answer is

(5/7)\*(4/6) = 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: The case of Total Probability :

Expected no. of candies for a randomly selected child will be:

candies count (a) \* probability (a) + candies count (b) \* probability (b) + ……

candies count (f) \* probability (f) = 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 + 0.6 + 0.24 = 3.09

Expected number of candies for a random child will be ~= 3

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:

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

Median:

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

Mode:

Points Score Weigh

3.07 3.44 17.02

13.92 NaN 18.90

Variance:

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

Standard Deviation:

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

Range:

Points 2.170

Score 3.911

Weigh 8.400

dtype: float64



Values from all mean median mode we know it round around from 3.2 to 3.9

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:

The expected value (E) can be calculated using the formula:

E = (X1 + X2 + X3 + ... + Xn) / n

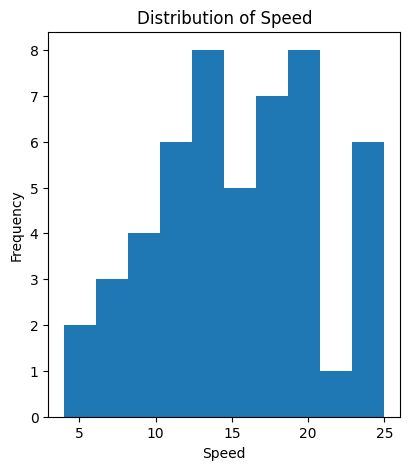
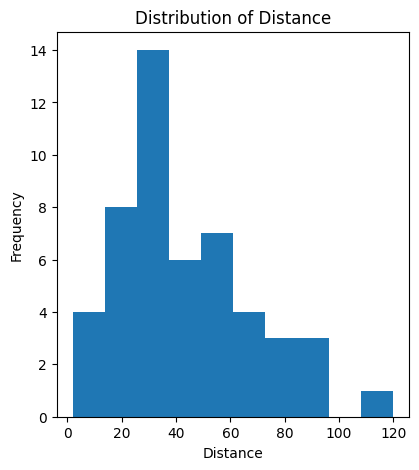
so the following will be :

* (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199) / 9
* 1308 / 9 = 145.33…
* ~= 145

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

**Cars speed and distance**

**Use Q9\_a.csv**

** Ans:**

**Skewness:**

Cars -1.200000

speed -0.508994

distance 0.405053

dtype: float64

**Kurtosis:**

Cars 0.000000

speed -0.117510

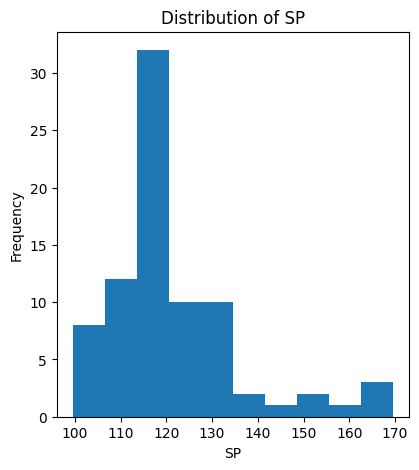
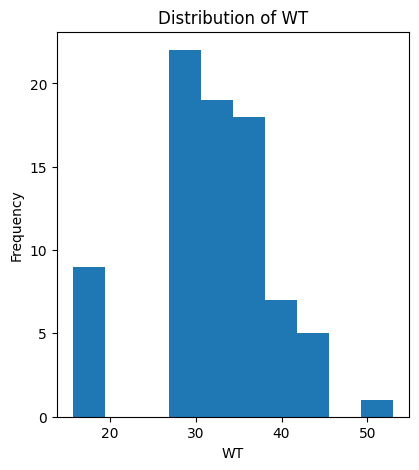
distance 0.806895

dtype: float64

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:**

****

**Skewness**:

Index 0.000000

SP 1.611450

WT -0.614753

dtype: float64

**Kurtosis:**

Index -1.200000

SP 2.977329

WT 0.950291

dtype: float64

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



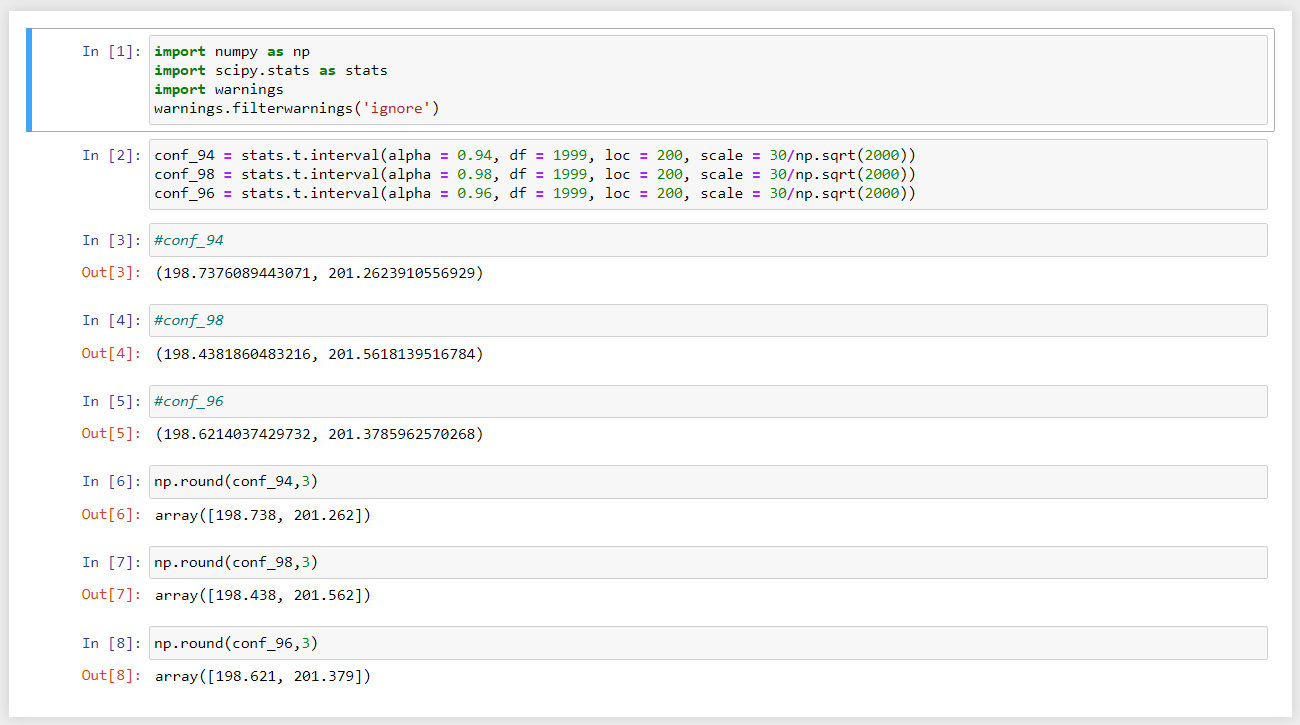
Answer: The historgram show above shows the right skewness and tail is on right which means Mean > Median and the outliners are on the high side.



Answer: The boxplot show above have outliners on the one side with maximum or minimum values.

**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:



**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: 1. Mean = (34 + 36 + 36 + 38 + 38 + 39 + 39 + 40 + 40 + 41 + 41 + 41 + 41 + 42 + 42 + 45 + 49 + 56) / 18

Mean = 738 / 18 => Mean = 41

2. Median, since it is even:

Median = (40 + 41) / 2 => Median = 81 / 2 Median = 40.5

3. Mode is 41

4. Applying formula of std: std = sqrt(Σ(xi - x̄)² / (n - 1))

=> s ~= 6.45

1. What can we say about the student marks?

Average students have 41 marks and graph is little right skewed.

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

Answer: Graph is perfect symmetry or normal distribution

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

Answer: Right Skewed

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

Answer: Left Skewed

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

Answer: A positive kurtosis value indicates that the data has heavier tails and a more peaked distribution compared to a normal distribution, Outliner are more.

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

Answer: A Negative kurtosis value indicates that the data has lighter tails and a less peaked distribution compared to a normal distribution, little or no outliner.

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



What can we say about the distribution of the data?

Answer: The boxplot show above is not normal distribution with median around 15.4

What is nature of skewness of the data?

Answer: left skewed

What will be the IQR of the data (approximately)?   
Answer: The Inter Quantile Range: Q3 – Q1  {Upper Quantile – Lower Quantile}= 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.

Answer: The median of the box boxplots should be in range between 275 and 250

and both the graphs are normally distributed with no Outliners

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

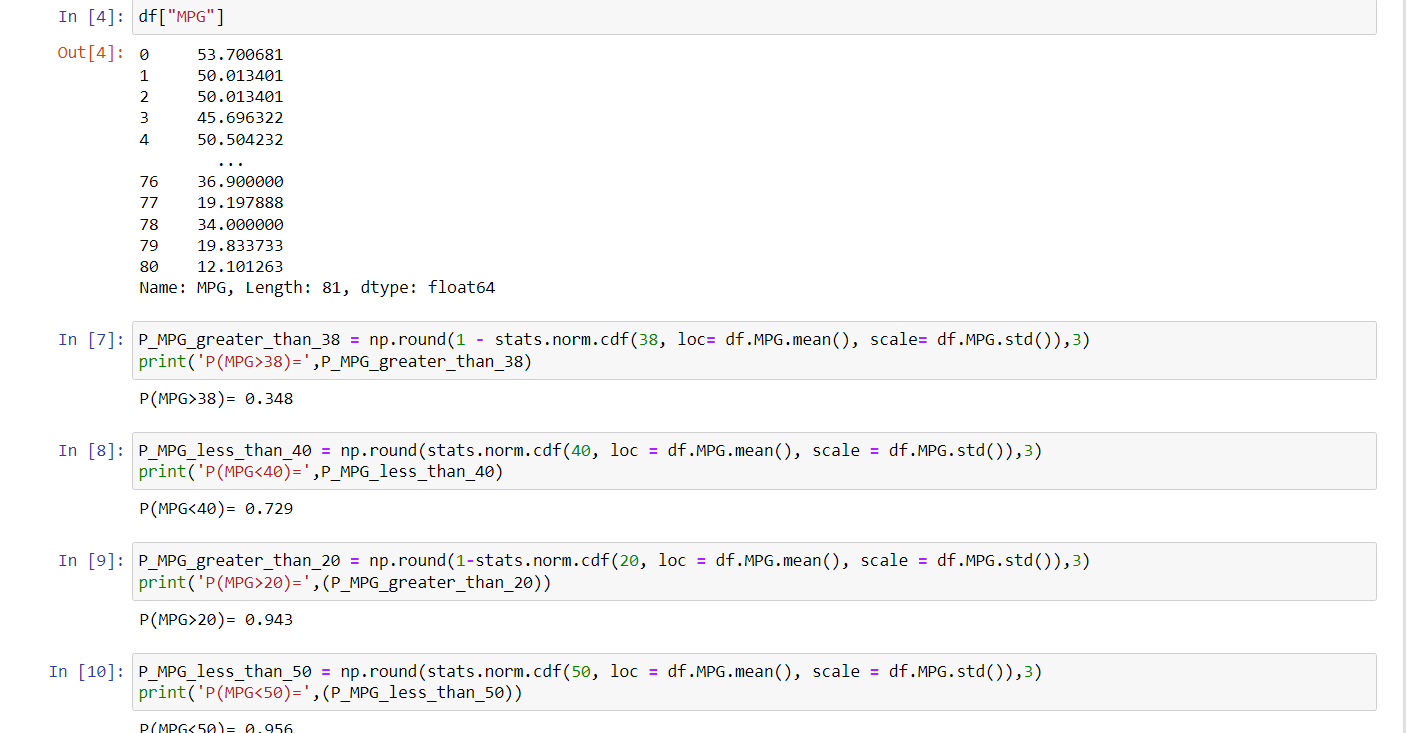
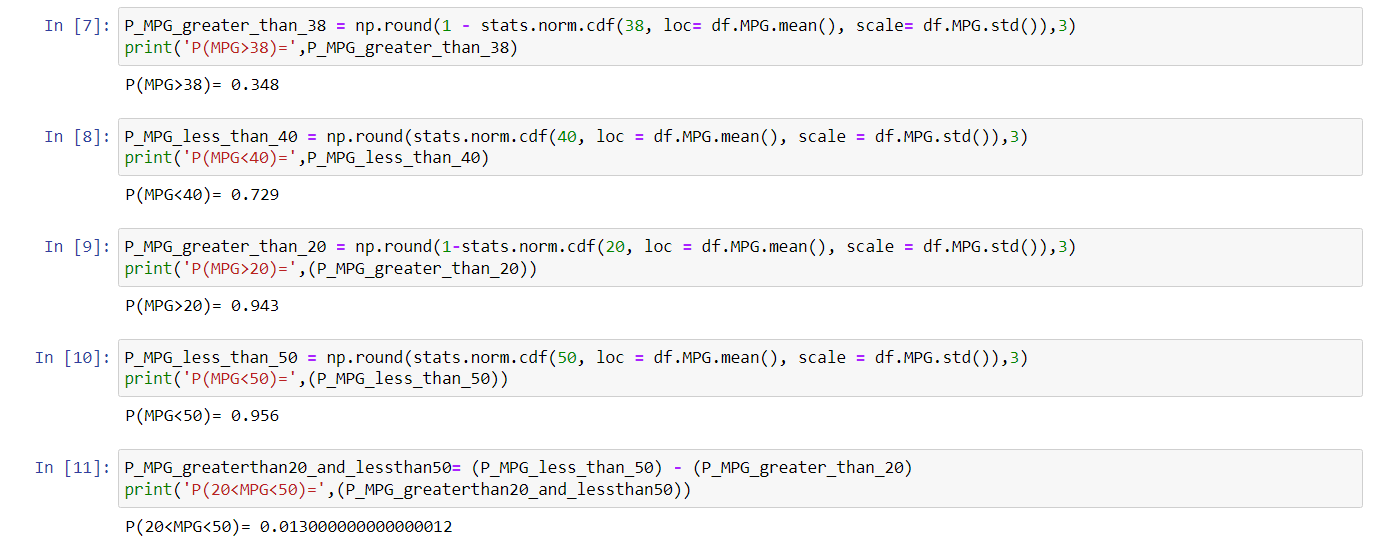
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

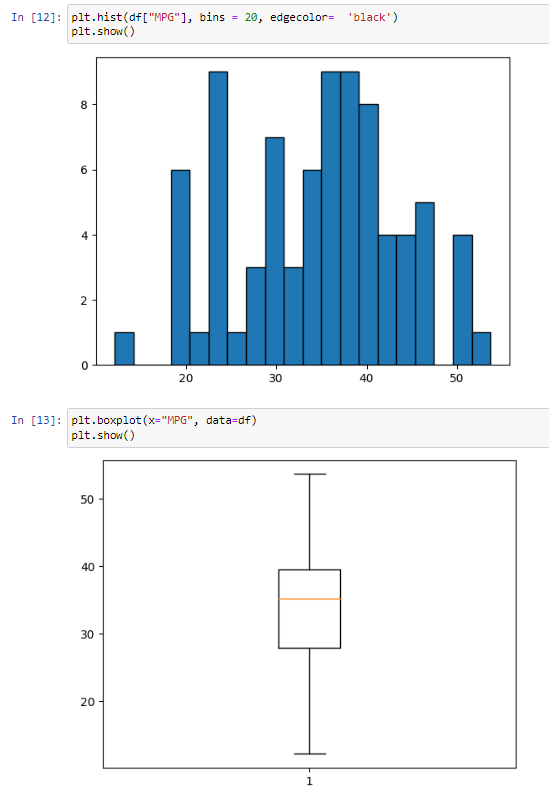
c. P (20<MPG<50)

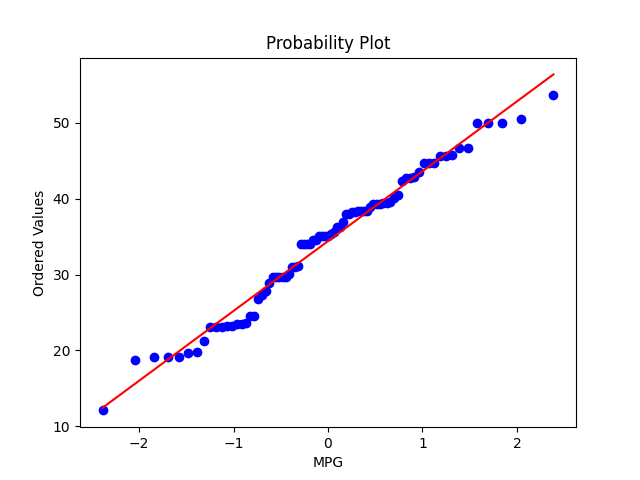
  

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

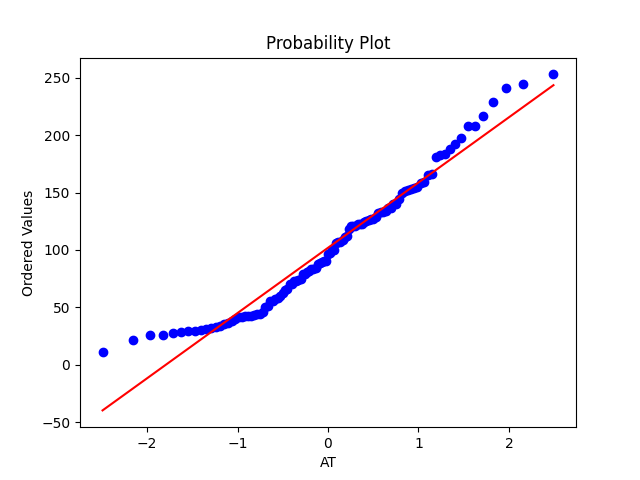
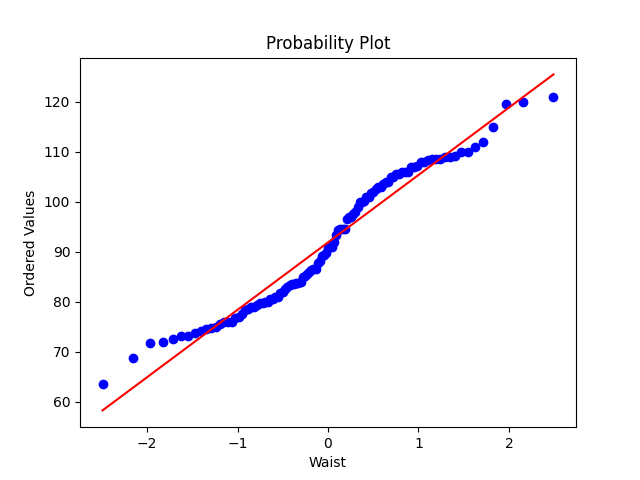
Answer: 

Yes [MPG] of the given dataset (cars) follows normal distribution 

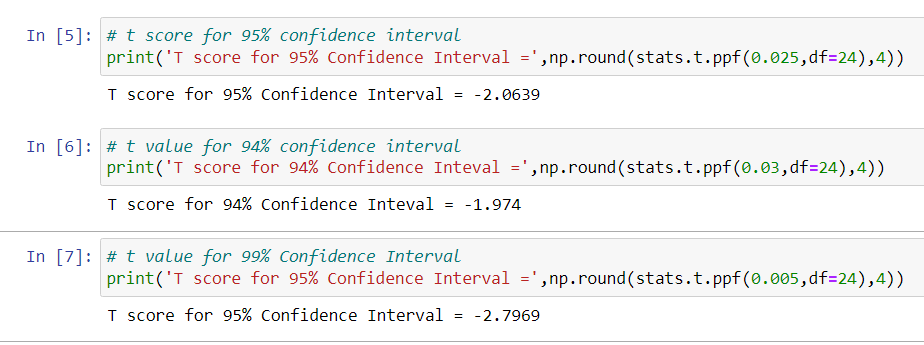
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer: No Adipose Tissue(AT) and Waist Circumference(waist) from wc-at data set does not follows Normal Distribution



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

