|  |  |
| --- | --- |
| 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 |

Name : Uday Rao

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 | Interval |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans-> 3/8

HHT

HTH

THH

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. 0/36
2. 6/36 (1,1)(1,2)(1,3)(2,1)(2,2)(3,1)
3. 6/36 (1,5)(2,4)(3,3)(4,2)(5,1)(6,6)

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?

Solution ->

RR GGG BB

RR RR

RG RG

RG RG

RG RG

RB RB

RB RB

GR GR GR

GR GR GR

GG GG GG

GG GG GG

GG GB GB

GB GB GB

BR BR

BR BR

BG BG

BG BG

BG BG

BB BB

None of Ball is Blue 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 = ∑xP(x) = 1\*0.015+ 4\*0.20 +3\*0.65 +5\*0.005 + 6\*0.01 + 2\*0.120

= 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>

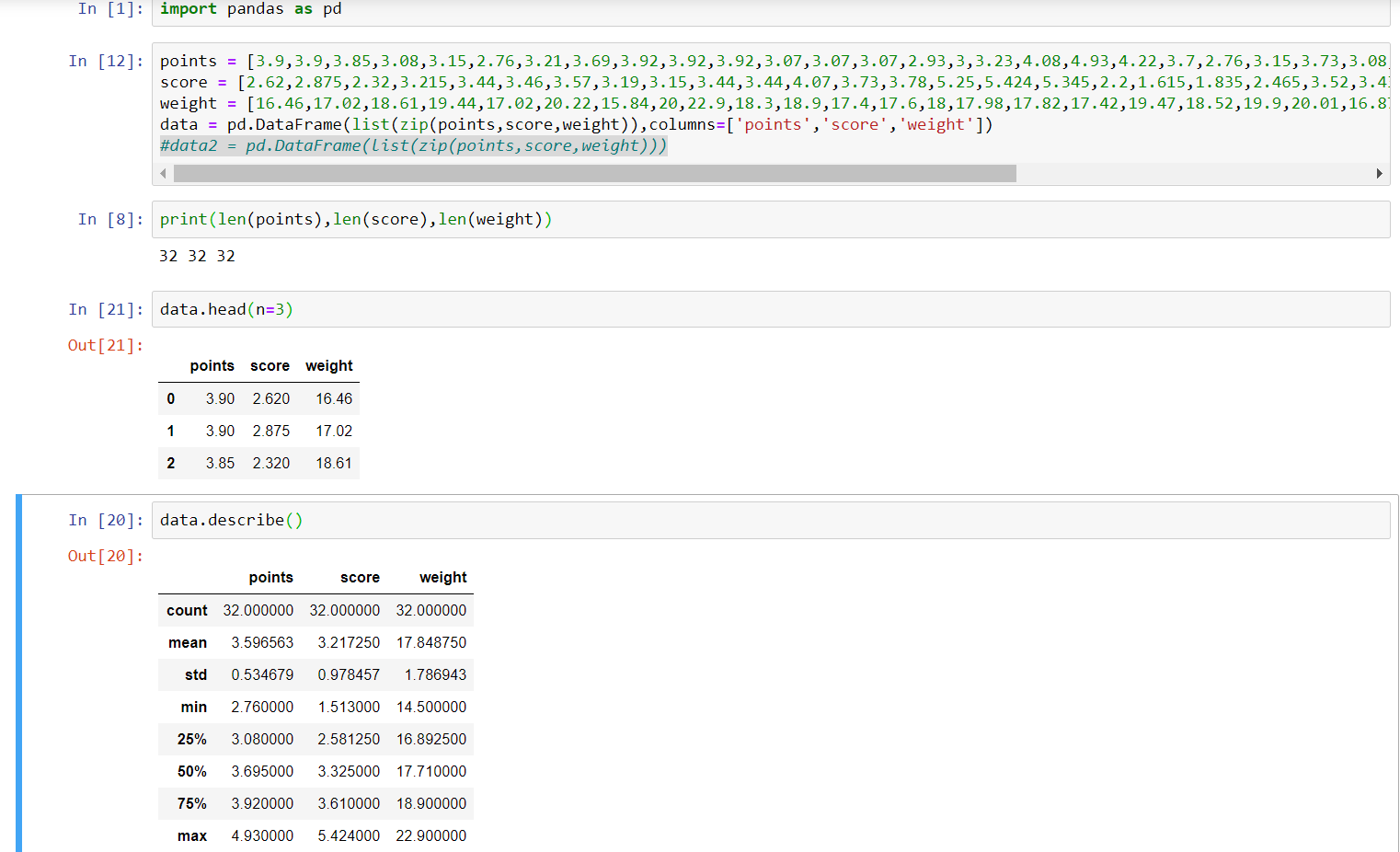
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.



Solution :

Inserting the above data into list and then coverting in DataFrame object.

Pandas while calculate the above values from the dataframe objects.



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?

Solution:

Expected value is nothing but the mean weight of patients.

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

= 145.33

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

**Cars speed and distance**



**SP and Weight(WT)**



**Solution :**

**Skewness = 3(Mean-Median) / Standard Deviation**

[1] -0.1139548 ## left skewness

kurtosis(speed)

[1] 2.422853 ## negative Kurtosis

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



Sol.

Following are the inferences can be drawn:

* **Shape :** The shape of the data is not **symmetric**.
* **It is Right-skewed data.**
* **Hence mean is not a good option to measure the central tendency**
* **As we can go for median.**
* **Because it is right-skewed data, median will be less then the mean.**

**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.

94% = (1+0.94)/2=1.88

98% = (1+0.98)/2=2.33

96% = (1+0.96)/2=2.06

X +- Z s/root n

X +- Z (0.6709)

94% 200 +- 1.2613

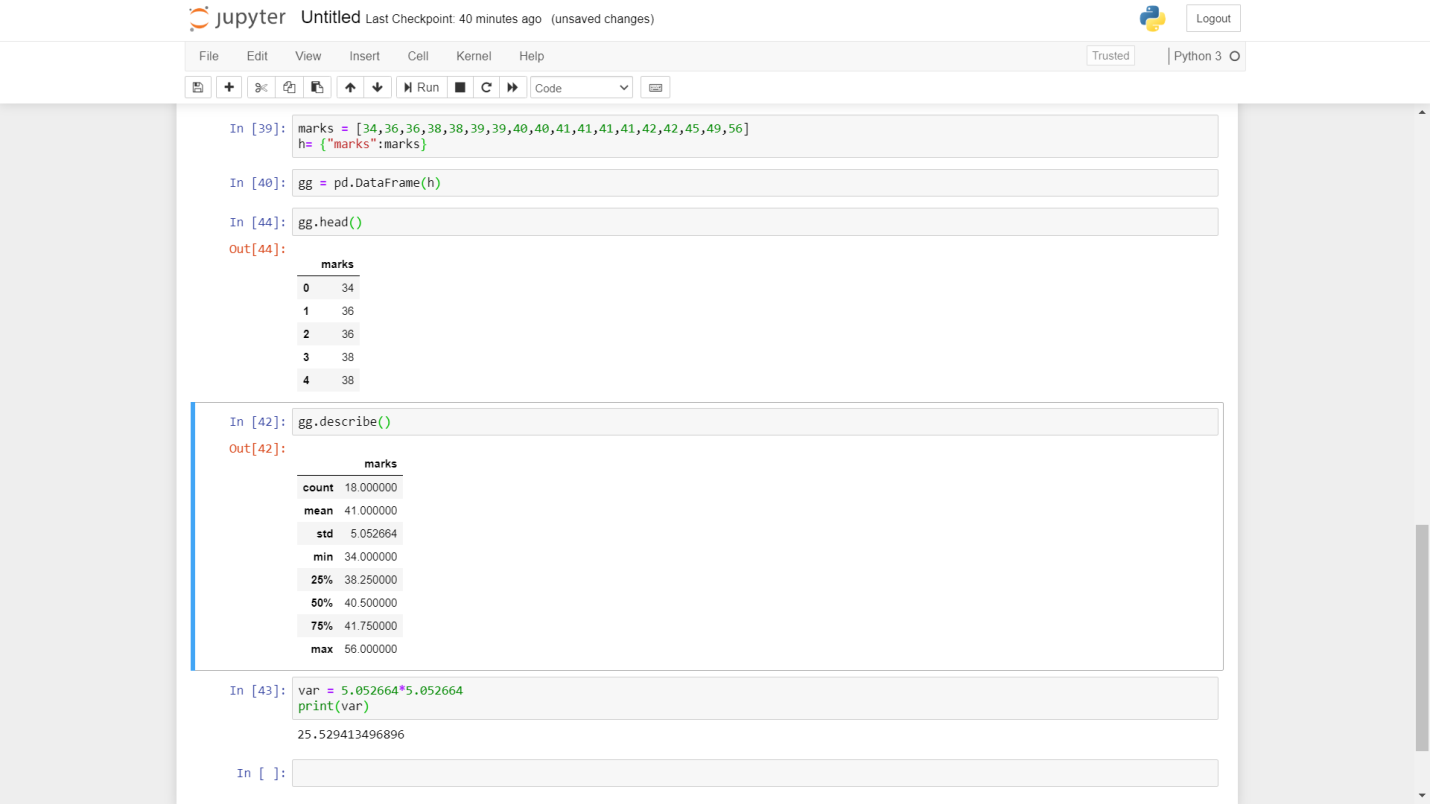
98% 200 +- 1.563197

95% 200 +- 1.382054

**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?

Sol. 

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

Sol.

Zero Skewness

Distribution will be symmetric.

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

Sol. Right-skewedData

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

Sol.Left-skewed data

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

Sol. Sharp Peak, Thick Tails

Positive (excess) kurtosis means that the outlier character of your data is more extreme that expected had the data come from a normal distribution.

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

Sol. Broad Peak, Thin Tails

Negative (excess) kurtosis means that the outlier character of your data is less extreme that expected had the data come from a normal distribution

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



What can we say about the distribution of the data?

Sol. Median = 15

What is nature of skewness of the data?

Sol. Left-skewed as the median divide into left more parts.

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

Sol. IQR = 3rd-1st

IQR = 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.

Sol.

Box1 has min value 235 where Box2 has min value 150

Box has same median value.

First Quartile and Third quartile are different.

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

Data \_set: Cars.csv

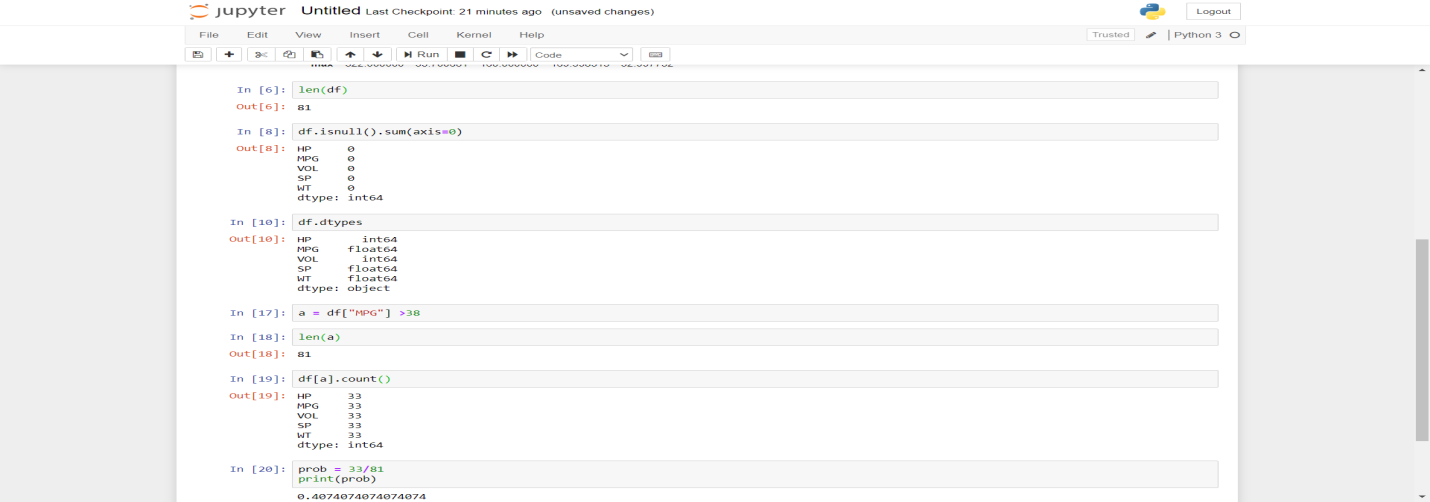
Calculate the probability of MPG of Cars for the below cases.

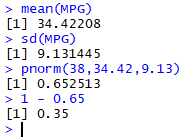
MPG <- Cars$MPG

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

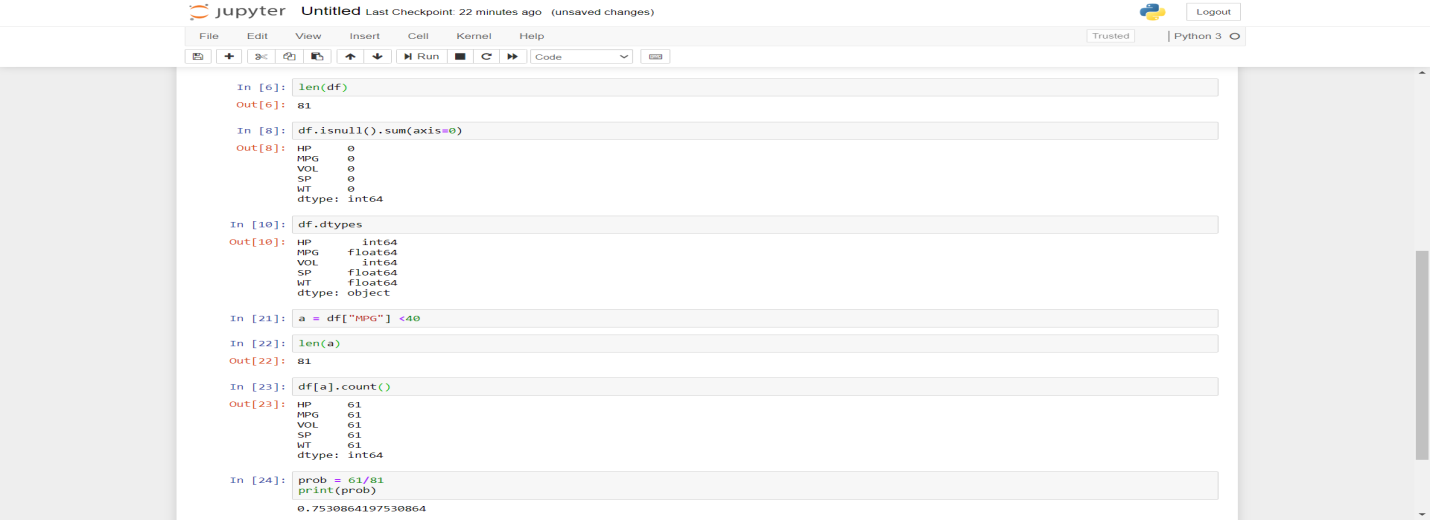
c. P (20<MPG<50)

Sol. A



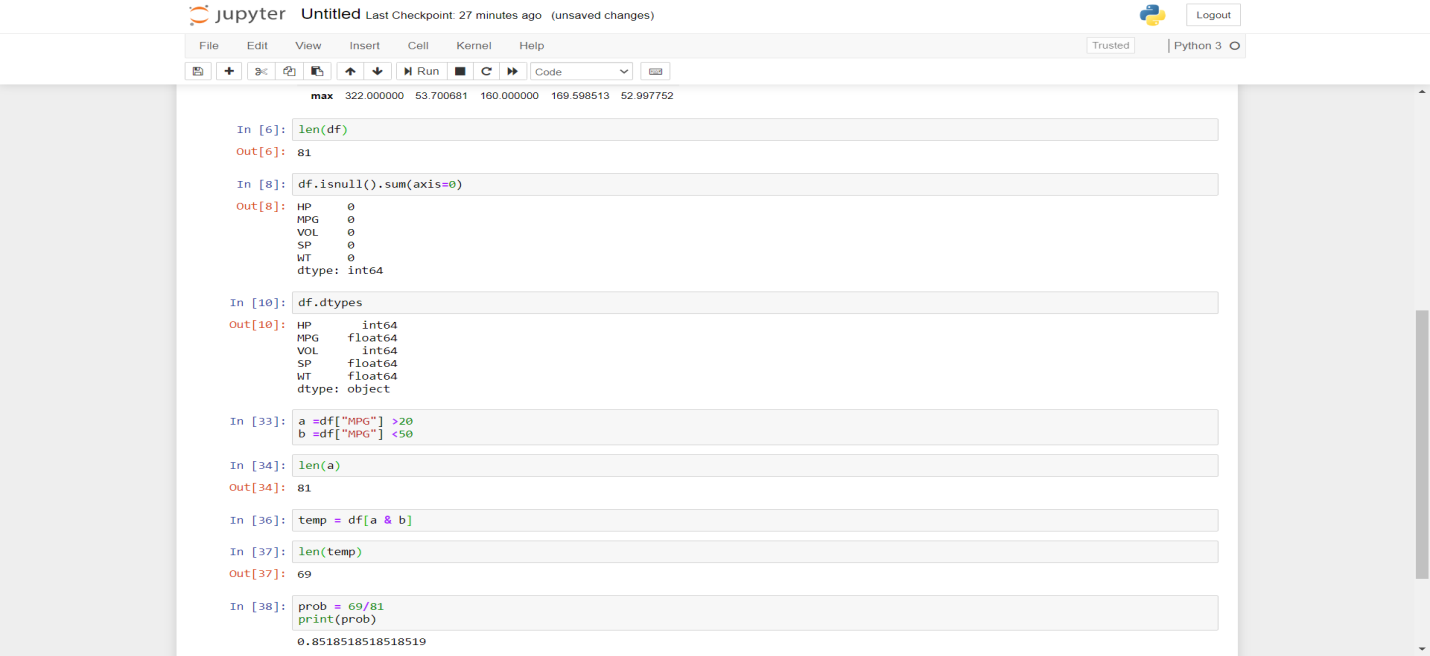


Sol. B





Sol. C

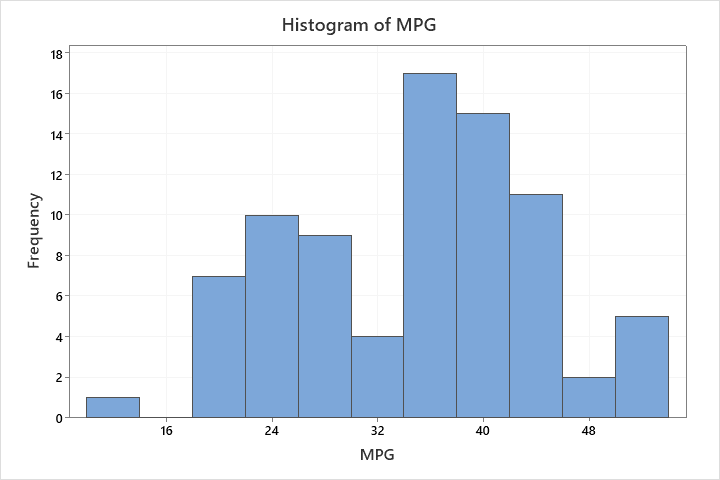
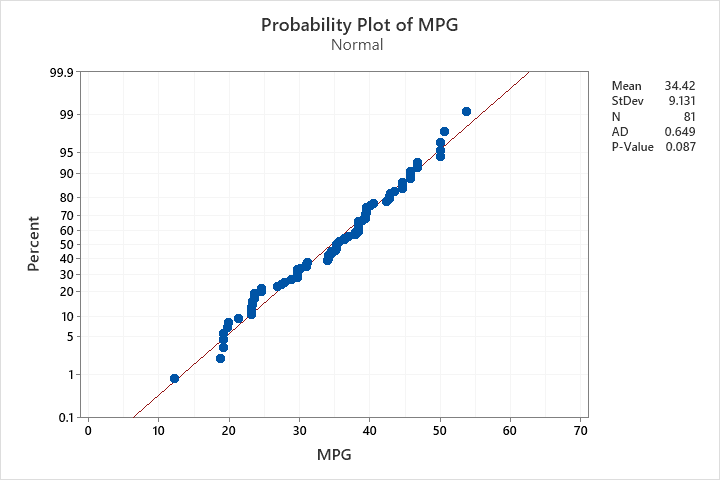




Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol. 

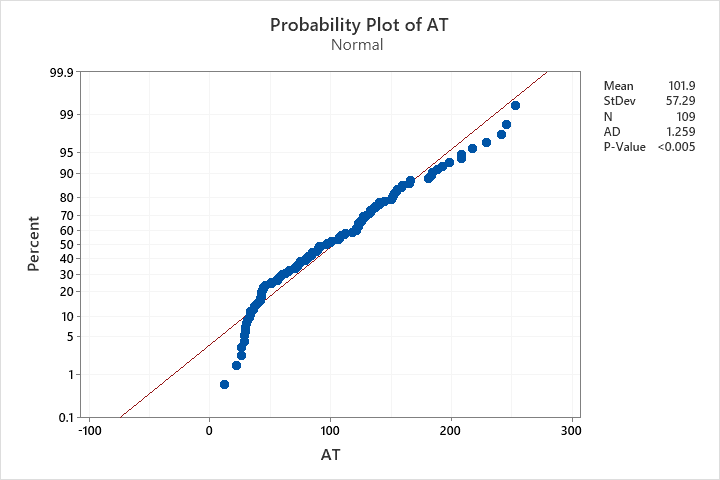
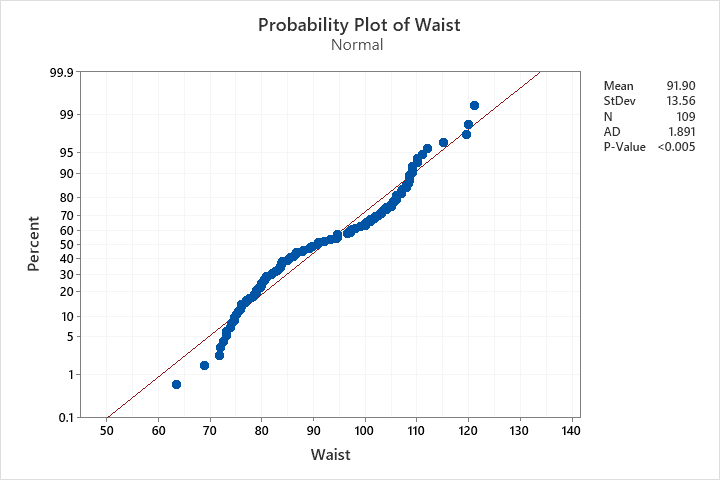
Follows normal distribution.

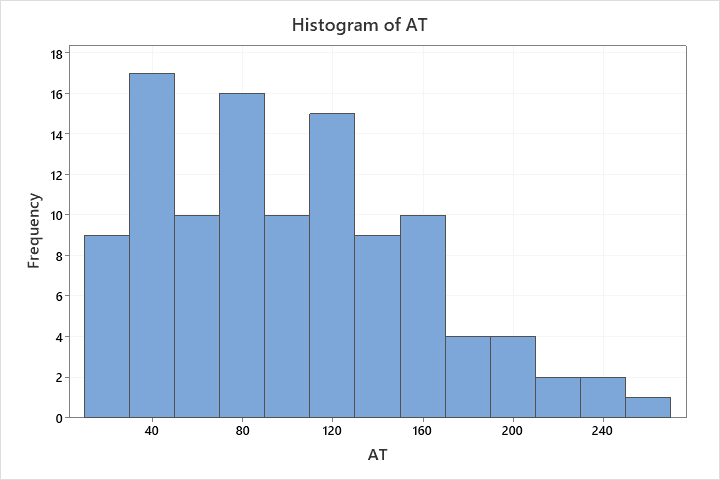
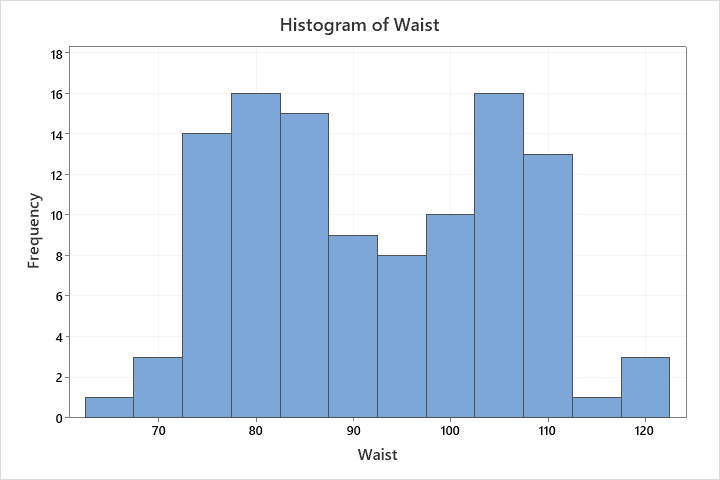
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Sol.

Both are not normally distributed.





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

Sol. 90% Z=(1+0.9)/2=0.95 1.65

94% Z=(1+0.94)/2=0.97 1.88

60% Z=(1+60)/2=0.8 0.84

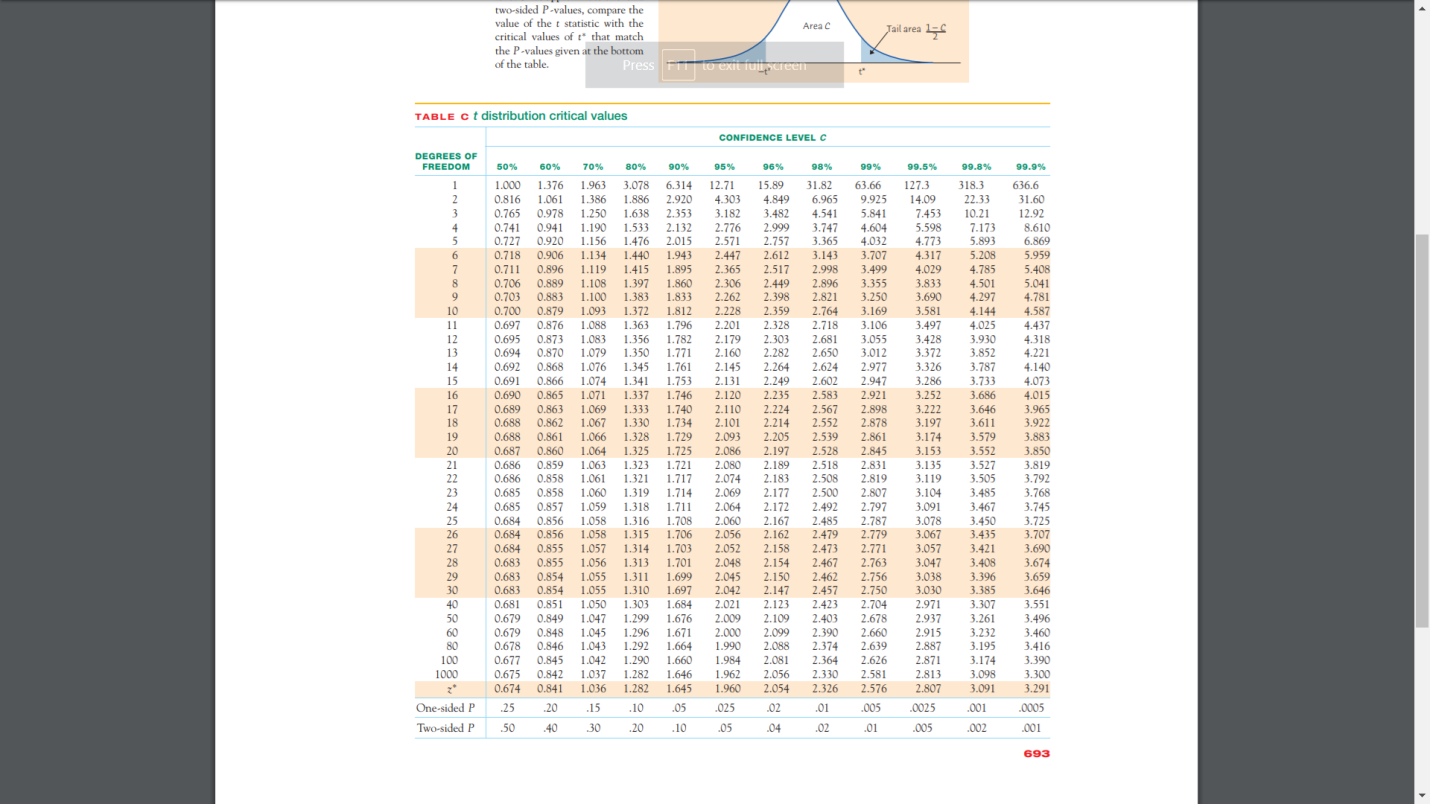
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Sol. Degree of freedom = 24

95 2.064

96 2.172

99 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

Sol.

0.3218