|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

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

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

**Ans;**

When three coins are tossed together,

The total number of favorable outcomes = 8

{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

**Probability = number of favorable outcomes / total number of outcomes**

Numbers of outcomes that gives two heads and one tail = 3

{HHT, HTH, THH}

Probability = 3/8 (or) 0.375 =3.75%.

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;**

When two dices are rolled n(s) = 6\*6 =36

**Probability = number of favorable outcomes / total number of outcomes**

**a)** The sum of equal to 1 is 0

There cannot be any probability of 1 outcome

I.e., =0

**b)** the sum is equal to 4

B= {(1,3), (2,2), (3,1)}

n(B) =n(B)/n(s) =3/36 = 0.0833 =8.33%

**C)**sum is divisible by both 2 and 3

Favorable outcomes C= {(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

n(C) = n(C)/n(s) = 6/36 =0.166 = 1.66%

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;**

**Probability = number of favorable outcomes / total number of outcomes**

Probability **=** (2R,3G,2B) = (2+3+2) = 7

Total number of outcomes = 7c2

7c2 = (7x6)/(2x1) = 21

Number of favorable outcomes = 5c2

5c2 = (5x4)/(2x1) = 10

Probability = 7c2/5c2

i.e., Probability =10/21= 0.476 = 47.6%

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

**ANS;**

Expected random values = ΣX \* P(X)

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

Expected number of candies for randomly selected child = 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**

**ANS;**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard deviation | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |

Inferences;

* Mean value is close for both ‘Points’ & ‘Score’
* Mean = Median =Mode.

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 = P(X) \* E(X)

Total 9 patients, the probability of each patient P(X)= 1/9

E(X) = 108, 110, 123, 134, 135, 145, 167, 187, 199

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

= (1/9) (1308)

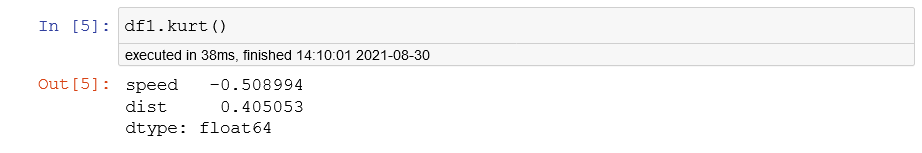
= 145.33

Expected value of the weight of the patient = 145.33(pounds)

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**ANS;** 

|  |  |  |
| --- | --- | --- |
|  | speed | dist. |
| Skewness | -0.11751 | 0.806895 |
| Kurtosis | -0.50899 | 0.405053 |

Skewness;

* The skewness value for speed is (Negative skewness), so it is left skewed.
* And for distance, is right skewed (positive skewness).

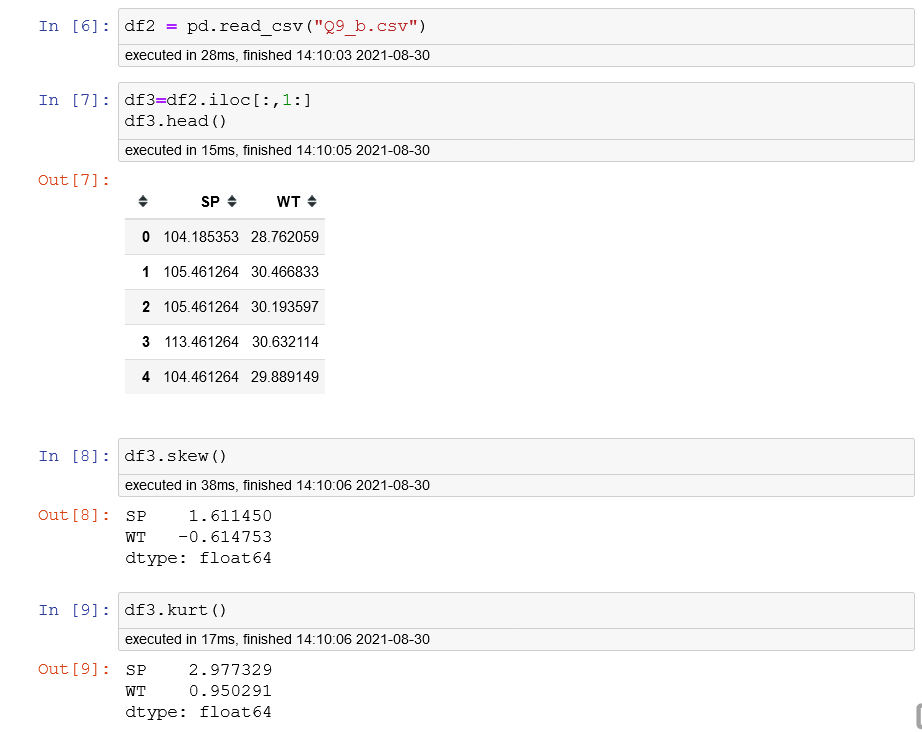
Kurtosis;

* Speed is negative kurtosis, (flatter than normal distribution)
* distance is positive kurtosis (peaked than normal distribution)

**SP and Weight (WT)**

**Use Q9\_b.csv**

**ANS;**



|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.61145 | -0.61475 |
| Kurtosis | 2.977329 | 0.950291 |

Skewness;

* Sp is right skewed (positive skewness)
* Wt is left skewed (negative skewness)

Kurtosis;

* Peaked than normal distribution

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



**Ans;**

* The most of the data points are concerted in the range 50-100 with in frequency 200.
* And least range of weight is 400 somewhere around 0-10.
* Skewness – we can notice a long tail towards right so it is heavily right skewed.

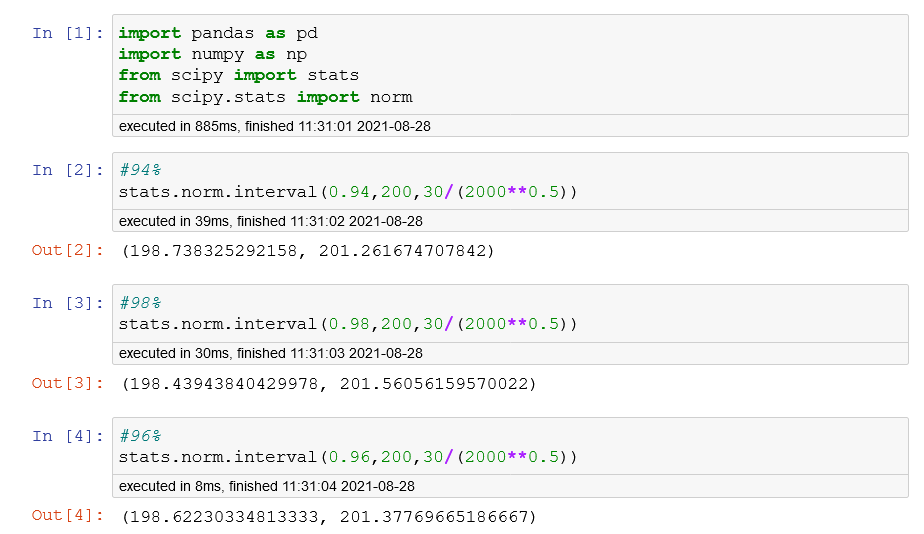


**Ans;**

* Median is less than the mean right skewed.
* we have outlier on the upper side of the box plot.
* There are less data points between q1 and bottom points.

**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?

**Ans;**

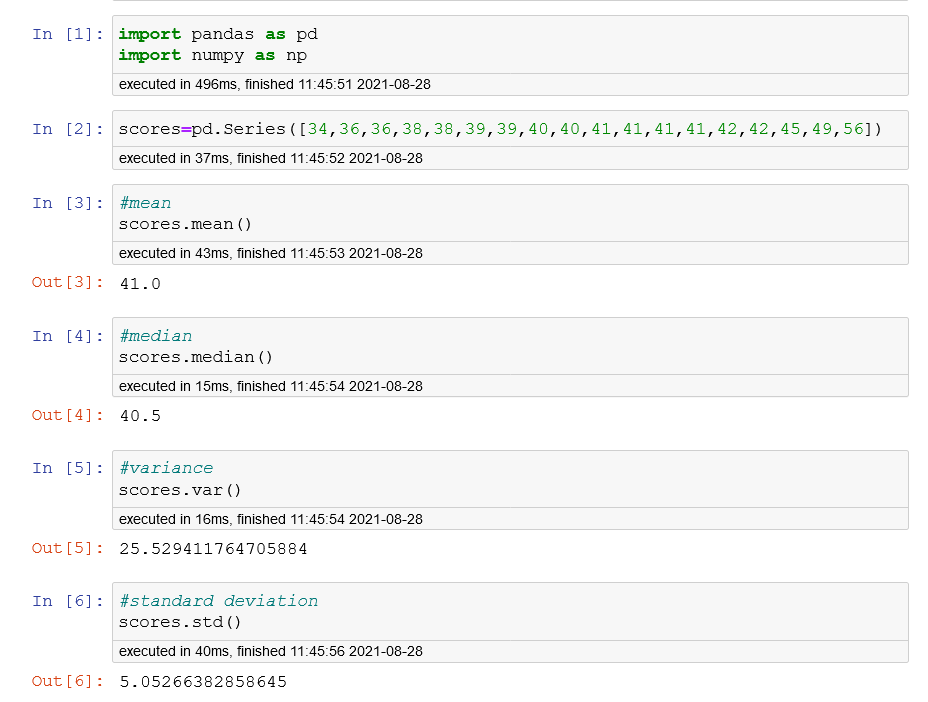


**Q12)** Below are the scores obtained by a student in tests

**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.

**Ans;**



2)What can we say about the student marks?

**Ans;**

* There are 2 outliers in the student’s marks 49 & 56.
* The mean is approximately equal to the median.

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

**Ans;**

* Mean=median=mode
* perfect skewness (**Normally distributed)**

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

**ANS:**

* mean > median
* positively skewed data (Right skewed)

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

**ANS:**

* mean < median
* negatively skewed data (Left skewness)

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

**ANS:**

* The data is normally distributed and kurtosis value is 0.
* Bell curve structure.

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

**ANS:**

* The distribution of the data has lighter tails and a flatter peak than the normal distribution.

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



1. What can we say about the distribution of the data?

**ANS;**

* Most of the data lies between 10-18
* Quartile
  + Q1 = 10
  + Q2 = 15(MEDIAN)
    - (Most of the values lies below the median.)
  + Q3 = 18
* median is greater than mean.

1. What is nature of skewness of the data?

**ANS;** Negative skewed data (outliers is present)

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

**ANS;**

INTER QUARTILE RANGE (IQR) = Q3-Q1 = 18-10

Approximately (IQR) = -8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**ANS;**

|  |  |
| --- | --- |
| **Boxplot 1** | **Boxplot 2** |
| Range b/w 240 - 280 | Range b/w 190 - 340 |
| Mean = Median = Mode = 260 | Mean = Median = Mode = 260 |
| Quartile(Q1) = 255 | Quartile(Q1) = 220 |
| Quartile(Q2) = 260 | Quartile(Q2) = 260 |
| Quartile(Q3) = 280 | Quartile(Q3) = 310 |
| Inter Quartile Range (IQR) = 25 | Inter Quartile Range (IQR) = 90 |

* By observing both the plots whisker’s level is high in boxplot 2.
* Mean=median=mode
* perfect skewness (**Normally distributed)**

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

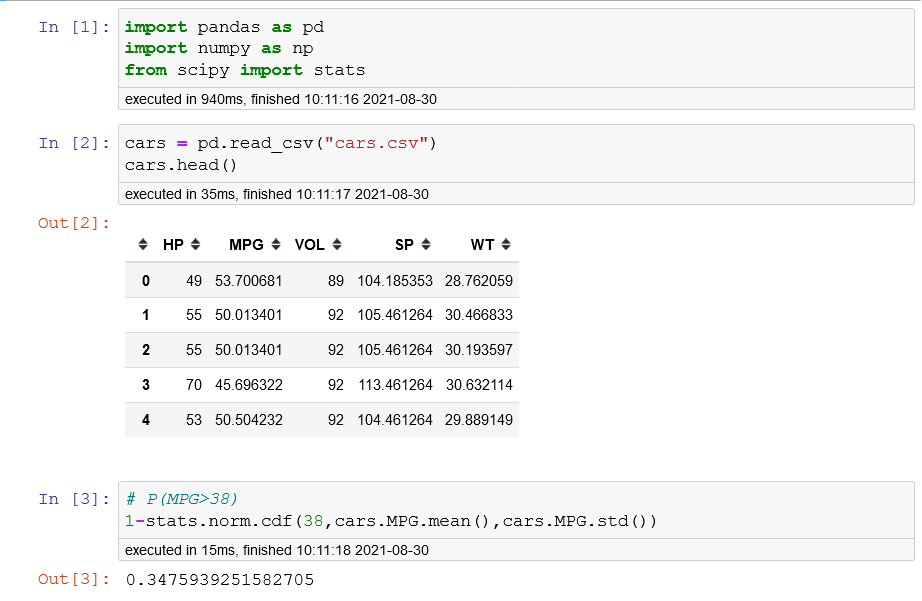
Data \_set: Cars.csv

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;**

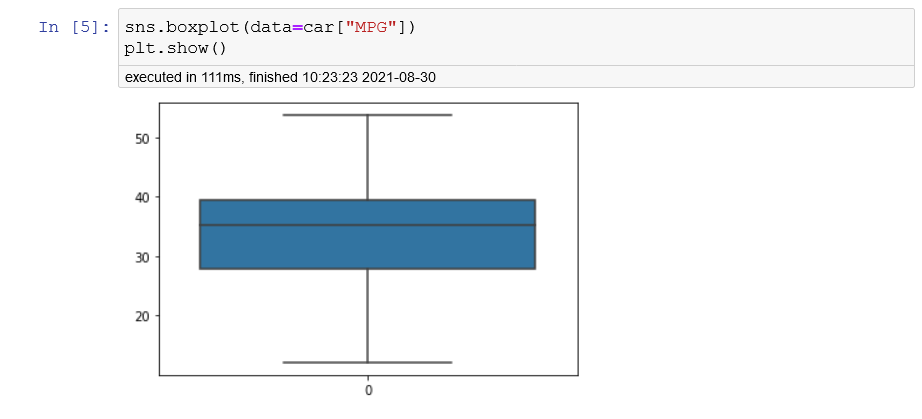
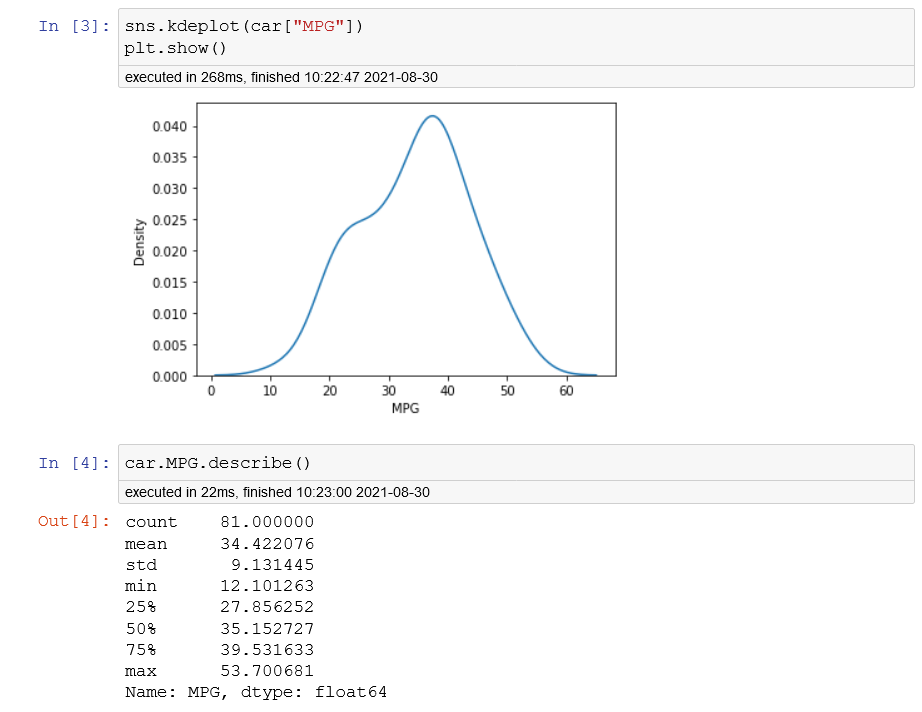
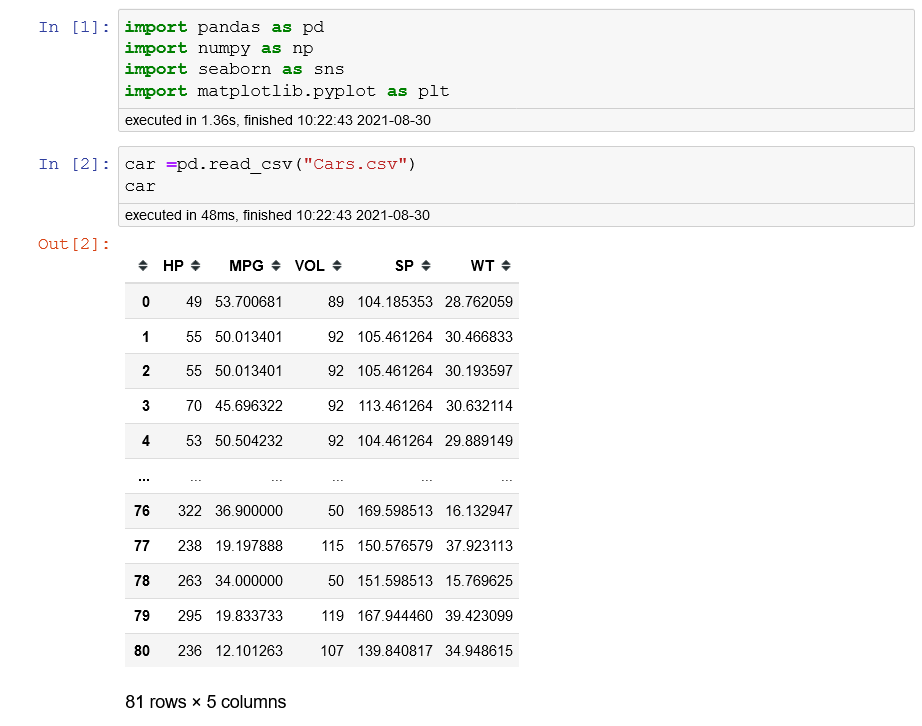


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS;**



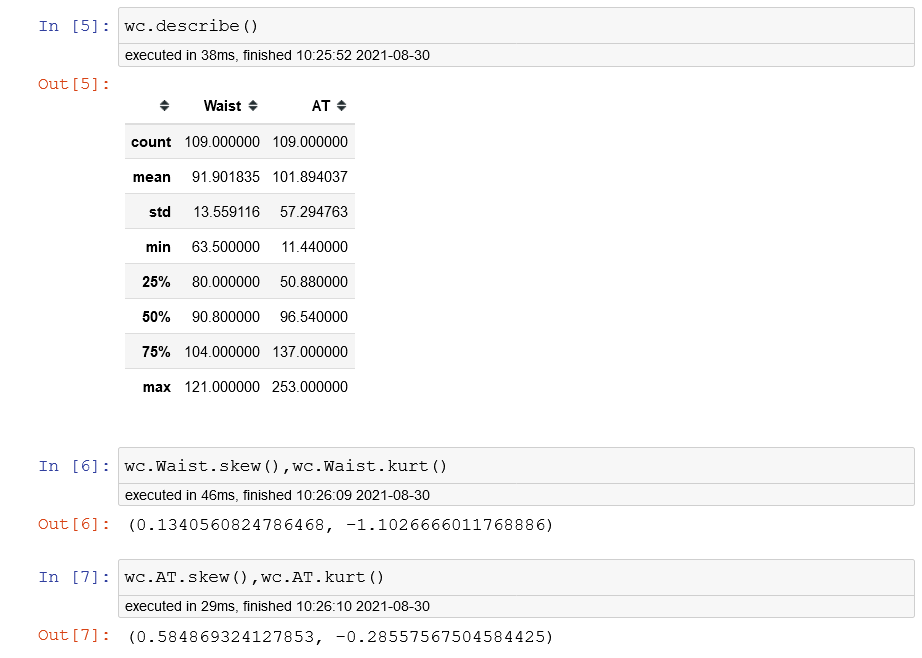
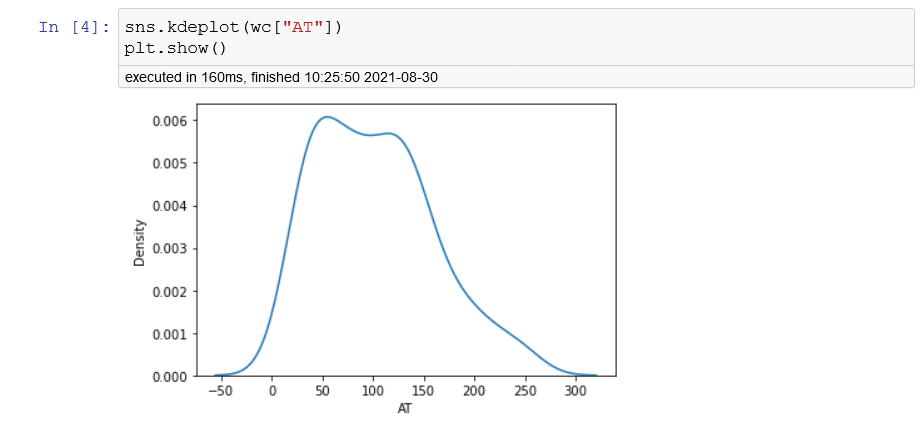
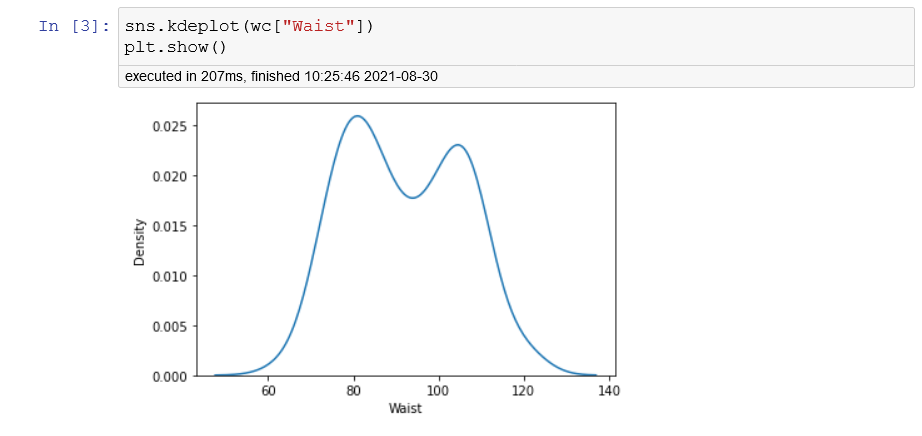
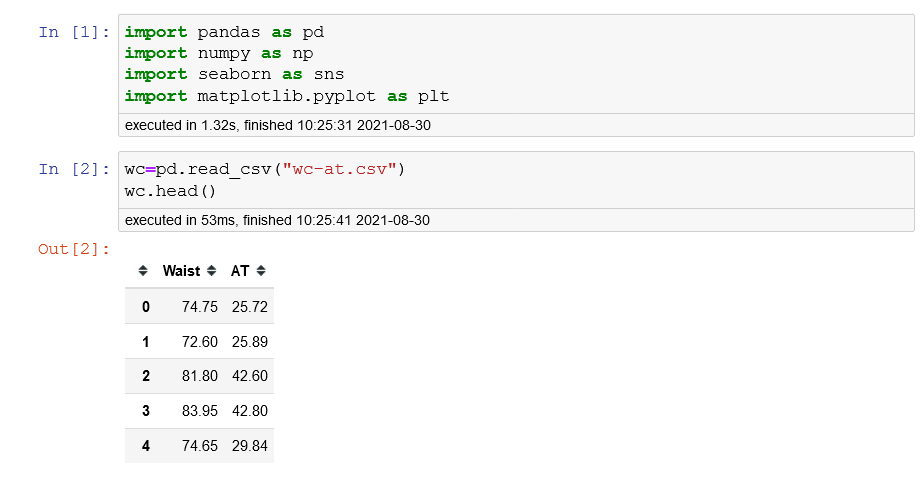
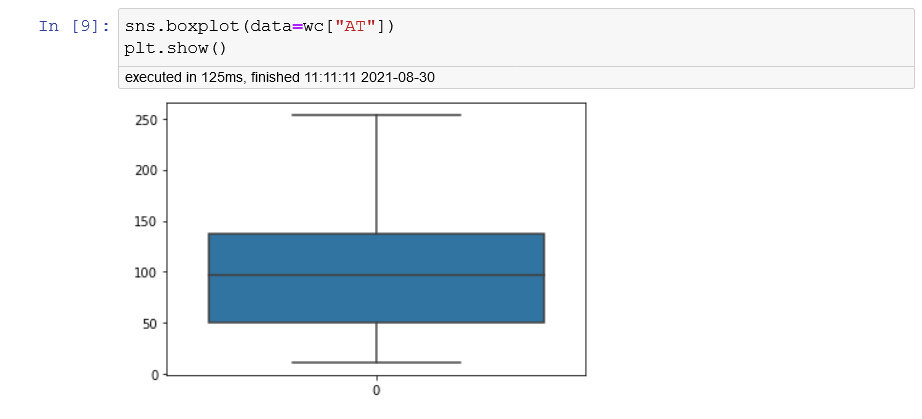
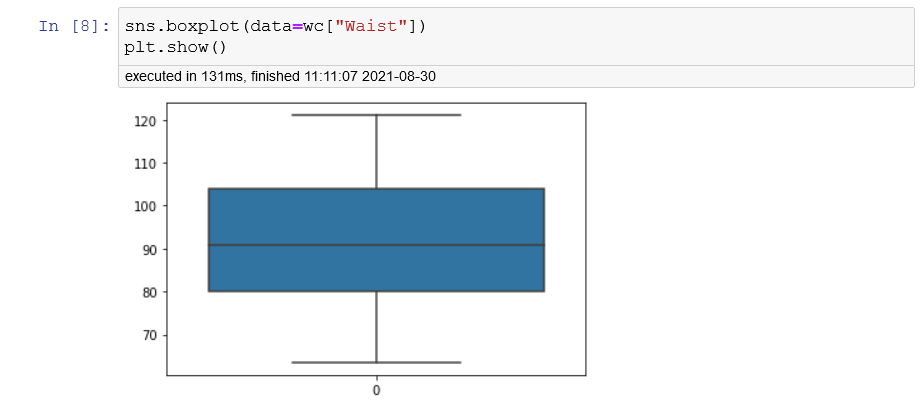
**Inference**;

* In boxplot Q2 is not accurate center, whisker is less negative side.
* Median (Q2) is nearer to Median(Q3), but not equal to it.
* Bell curve slightly skewed towards negative.
* MPG of Cars can follow normal distribution approximately (as mean and median are approximately same)

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

Dataset: wc-at.csv

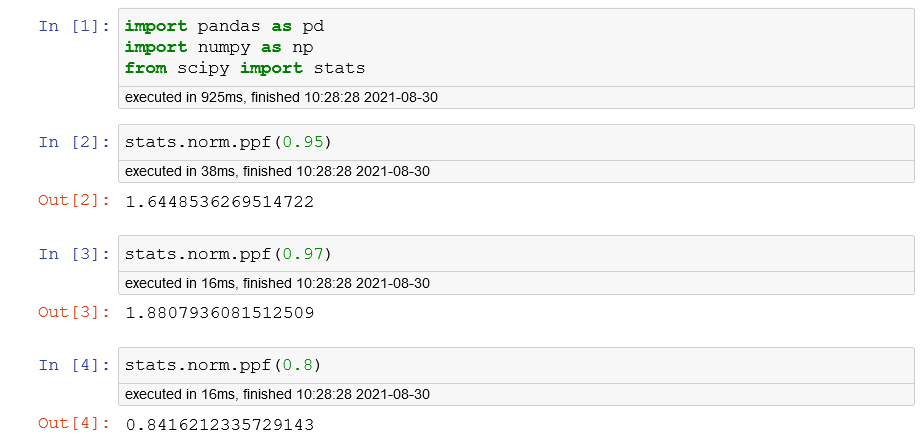
**ANS;**

 **Inference**;

* Both the (AT) and (Waist) data set are approximately equal to each other.
* mean > median (Right skewed)
* slightly positively skewed data.

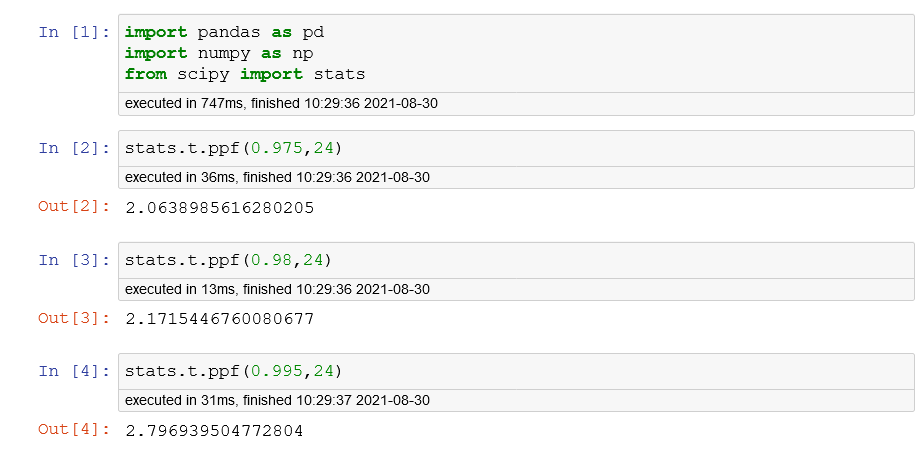
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**ANS;**



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

**ANS;**



Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 buslbs 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;**

Average light bulb (μ)=270

Sample bulb (n)=18

Average Sample (x)=260

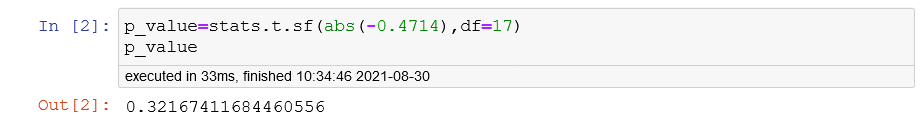
Standard deviation (S)=90

**T = (X –** **μ) / [ s/√(n)]**

T = (260-270)/ (90/18\*\*0.5)

T = - 0.4714

Pt= - 0.4714, df=17



Probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32.17%.