|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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?

Total number of events (S)= {HHH, HHT, HTT, TTT, TTH, THH, HTH, THT}

n(S)=8

E= {HHT, THH, HTH}

n(E)=3

P(E)=n(E)/n(S)

P(E)=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

Ans: Total Numbers of outcome =6\*6=36

S=(1, 1)(1, 2)(1, 3)(1, 4)(1, 5)(1, 6) (2, 1)(2, 2)(2, 3)(2, 4)(2, 5)(2, 6) (3, 1)

(3, 2)(3,3)(3, 4)(3, 5)(3, 6) (4, 1)(4, 2)(4, 3)(4, 4)(4, 5)(4, 6)(5, 1)(5, 2)(5, 3)

(5, 4)(5, 5)(5, 6)(6, 1)(6, 2)(6, 3)(6, 4)(6, 5)(6, 6)

N(S)=36

1. Equal to 1= 0% probability
2. Less than or equal to 4= 6/36 = 1/6
3. Sum is divisible by 2 and 3=

={2 3 4 5 6 7 3 4 5 6 7 8 4 5 6 7 8 9 5 6 7 8 9 10 6 7 8 9 10 11 7 8 9 10 11 12}

=5/36

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?

Ans=

Total no. of balls= 2 red +3 green +2 blue= 7

S= two balls are drawn randomly

N(S)= = = 7 \* 6 /2 \* 1 = 21

A= none of the ball drawn is blue

N(A)= = 5 \*4 / 2 \*1= 10

P(A)= n(A)/n(S) =10/21 = 0.47

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

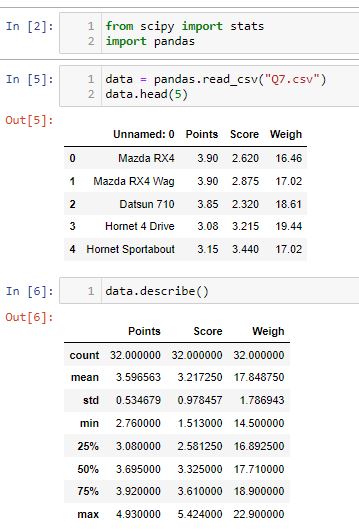
Ans= Expected number = E(x) ==1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= **3.09**

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.

ANS=Using Python





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:

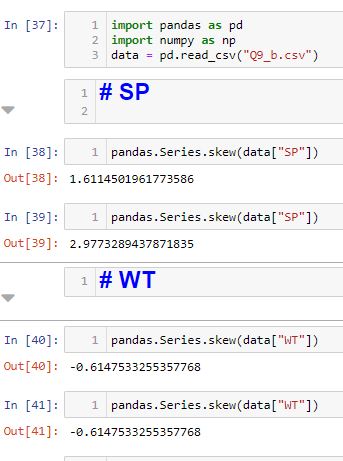
EV=Σx/n = = 145.33

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

**Cars speed and distance**

**ANS= Using python**

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

****

Inferences:

#### Speed:

1. Skewness = -0.117
   * data is slightly Negatively Skewed or Left Skewed data (Mass of data is on right side of median),
   * means data spared is More on left side of the Median
2. Kurtosis = -0.508
   * Data has platykurtic distribution& has thin tails compared to Normal dist.,
   * The distribution is flat as compared to Normal distribution.

#### Distance:

1. Skewness = 0.806
   * data is skewed Positively or Right skewed data (Mass of data is on left side of median),
   * Means data spared is more on right side of the Median
2. Kurtosis = 0.405
   * Data has Leptokurtic distribution & has thick tails as compared to normal dist.,The distribution is peak as compared to Normal Distribution.

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



**Inferance :**

* Positivelly Skewed data(Right Skewed data)
* Spared of the data on right side of the distribution is More & Mass of data is on left side of Median

Frequency of the data between 50 – 100 is more



**Interance:**

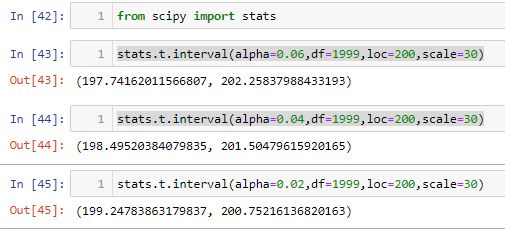
* Positivelly Skewed data or Right skewed data (Whisker is More on right side on median)
* Spared of the data on right side of the distribution is More & Mass of data is on left side of Median
* Positive Outliers are there on Right side of the

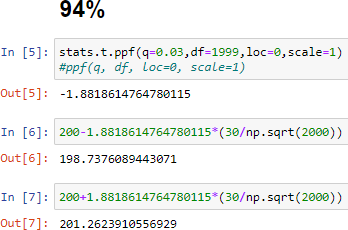
distribution

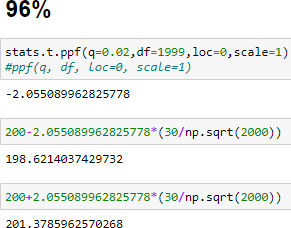
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

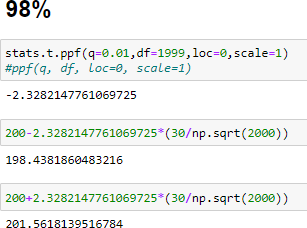
ANS:

**Using Python**

****







**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?

1)Ans =

Mean=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=738/18= 41

=AVERAGE(K3:K20)=41

Median= =MEDIAN(K3:K20)

=40.5

Varience= =VAR.S(K3:K20)

=25.529

Standard Deviation= =stdev.s(K3:K20)

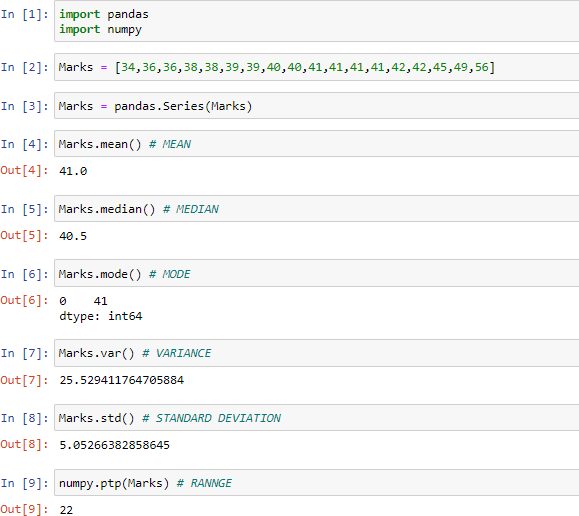
=5.052

2) Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.

|  |
| --- |
| **Mean = 41**   * Most of students’ marks are nearer to 41 |
| **Median = 40.5** ≅ **Mean**   * There is no too high (like 98,76) & too low marks (like 0,2) (Outliers) present |
| **Standard deviation = 5.05**  As mean is approximately equal to median follows Normal distribution,   * 1𝜎 = (41-5 =36, 41+5 = 47) * 68% of students are scored between 36 to 47 * 2 𝜎 = (41-10=31,41+10=51) * 95% of students are scored between 41 to 51 * 3SD= (41-15=26,41+16=57) * All most all (99.7%) students are scored between 26 to 57 |

ANS=

USING PYTHON



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

Ans= **Mean = Median,** we can say data is Normally Distributed.

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

Ans) **Mean > Median,** we can say Positively Skewed data (Right Skewed data).

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

Ans) **Mean < Median**, we can say Negatively Skewed data (left Skewed data).

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

### Ans) Positive Kurtosis (Excess Kurtosis) indicates that,

### • Distribution is Leptokurtic (peak of bell curve is more as compared to Normal distribution)

### • Spared There are more values around mean.

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

Ans) Negative Kurtosis indicates that,

• Distribution is Platykurtic (peak of bell curve is less as compared to Normal distribution)

• Spread of the data is More (There are more far values from mean).

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



What can we say about the distribution of the data?

• Most of the data lies between 10 t0 18.

• Q1 = Quartile 1 = 10

• Q2 = Quartile 2 = 15 = MEDIAN = 50th Percentile

• Q3 = Quartile 3 = 18

What is nature of skewness of the data?

Negatively skewed data: There are negative outliers present in the data

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

IQR = Q3 – Q1 = 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)



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

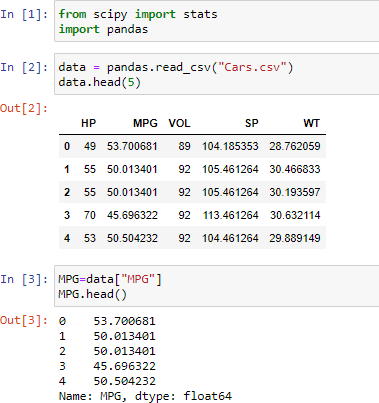
Data \_set: Cars.csv

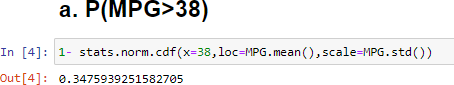
Calculate the probability of MPG of Cars for the below cases.

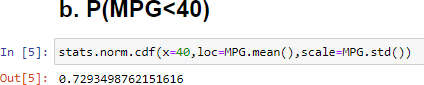
MPG <- Cars$MPG

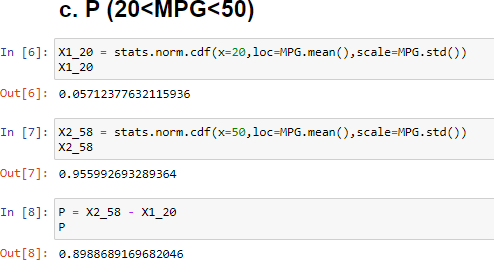
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans= Using Python





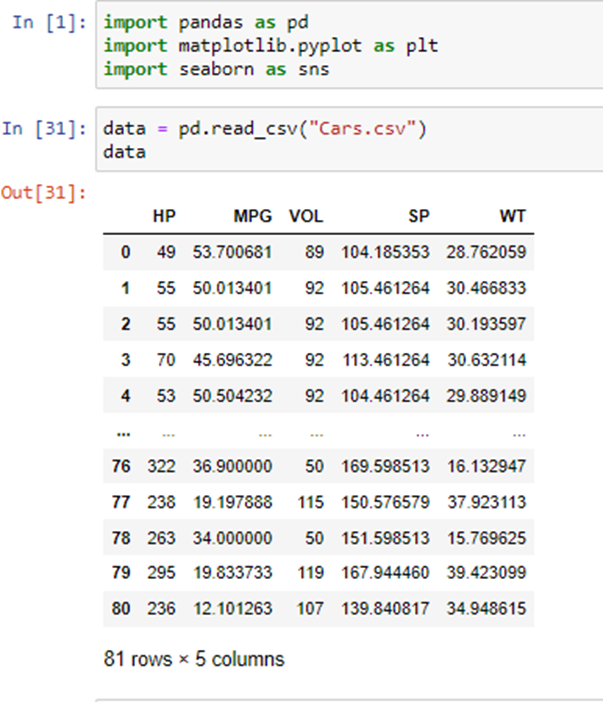


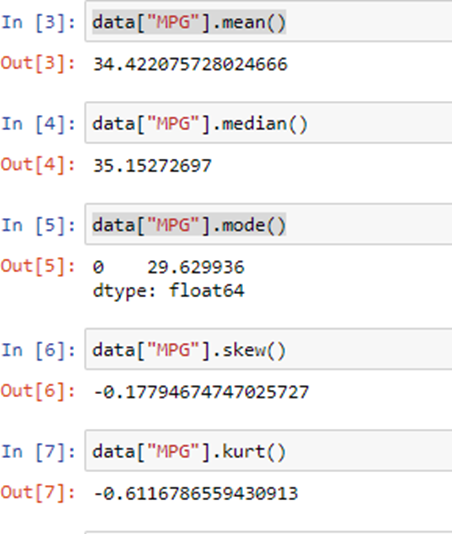


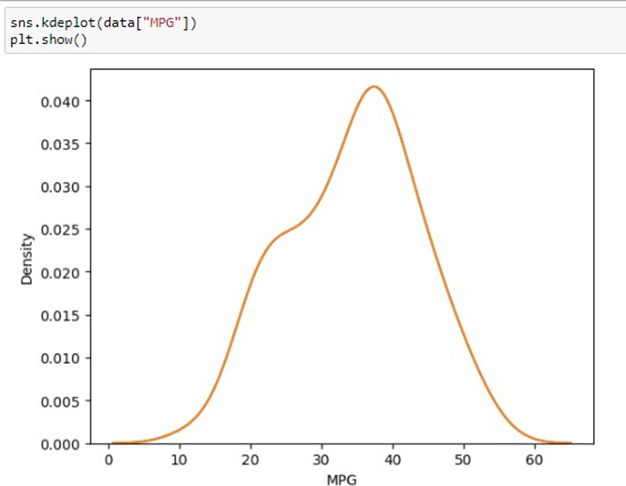
Q 21) Check whether the data follows normal distribution

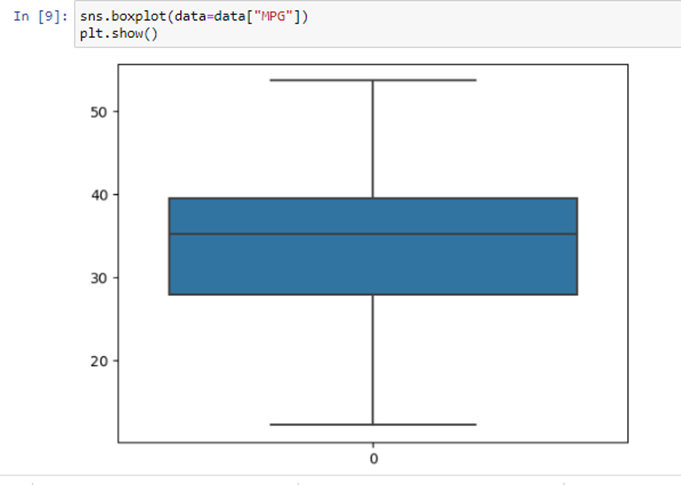
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv









1)MEAN ≠ MEDIAN,

2)Skewness = 0.177

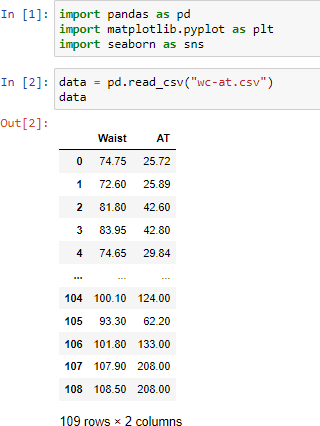
3)Kurtosis = 0.6116

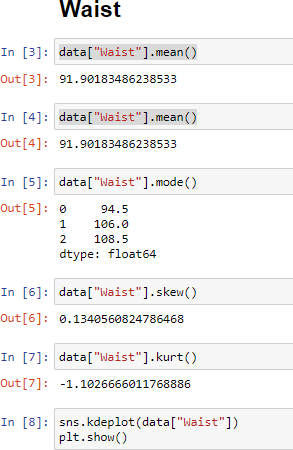
4)IN Box plot Q2 is not at center, whisker is more negative side , Midian(Q2) is nearer to Q3 and in bell curve skewed towards negative numbers

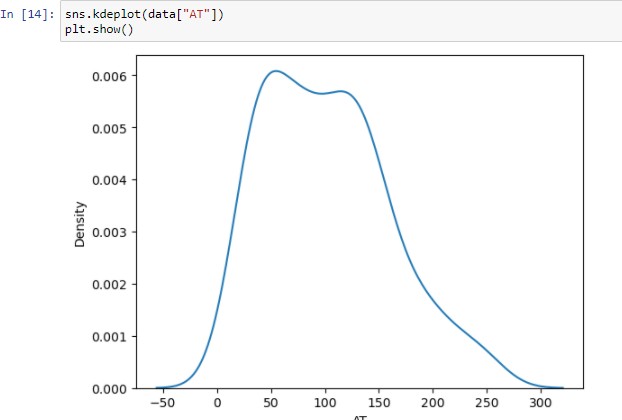
We can Say That the “MPG” data is Sightlly Right skewed or Negatively Skewed data.

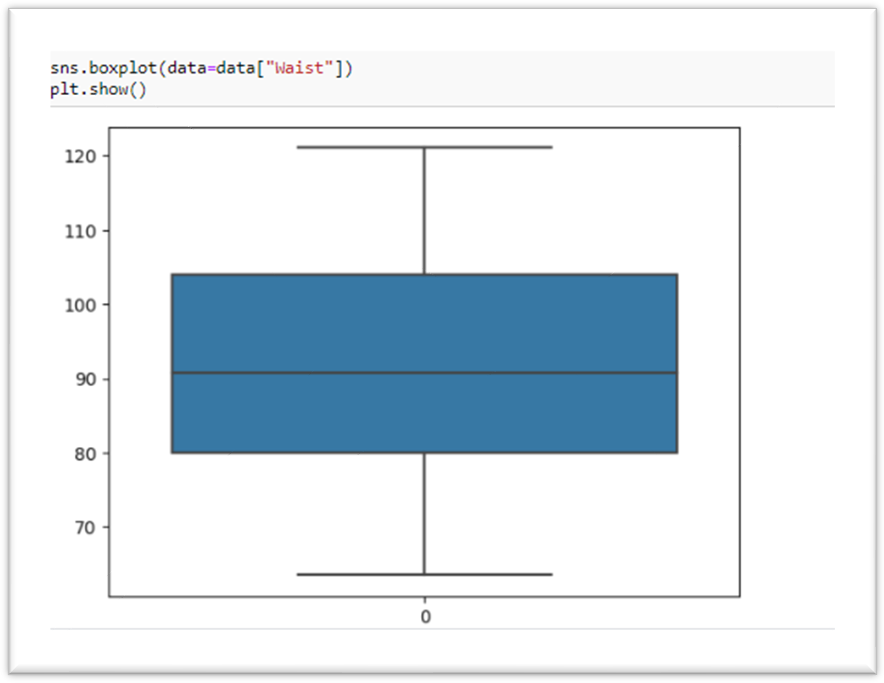
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv









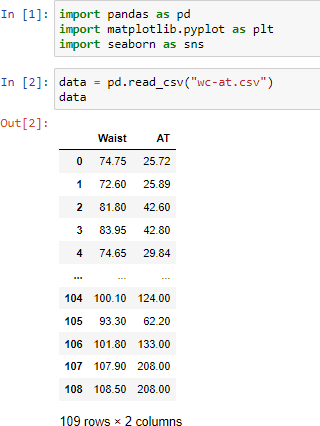
1)MEAN = MEDIAN =91.9018,

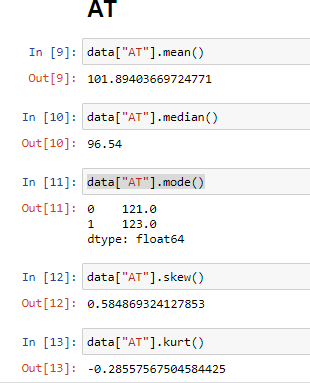
2)Skewness = 0.134 ≅ 0

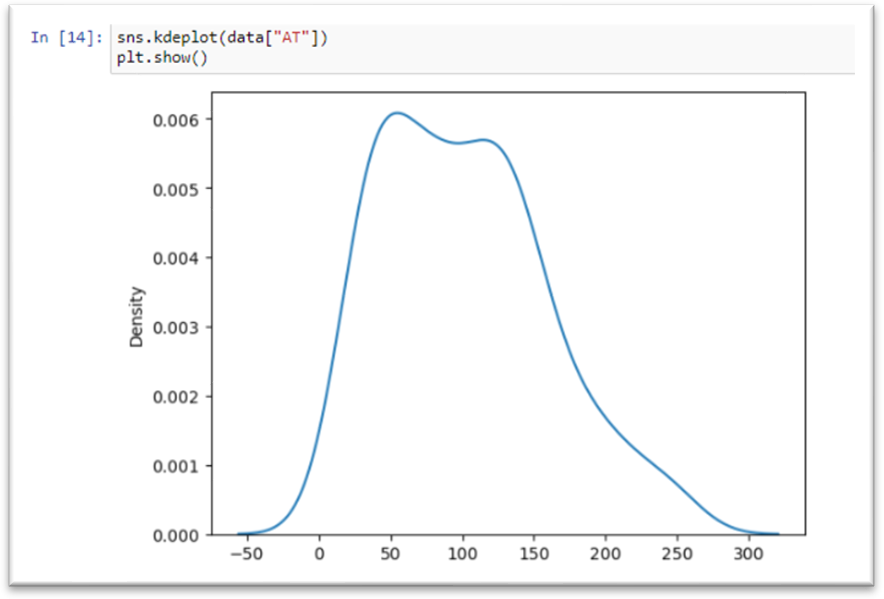
3)Kurtosis = -1.01

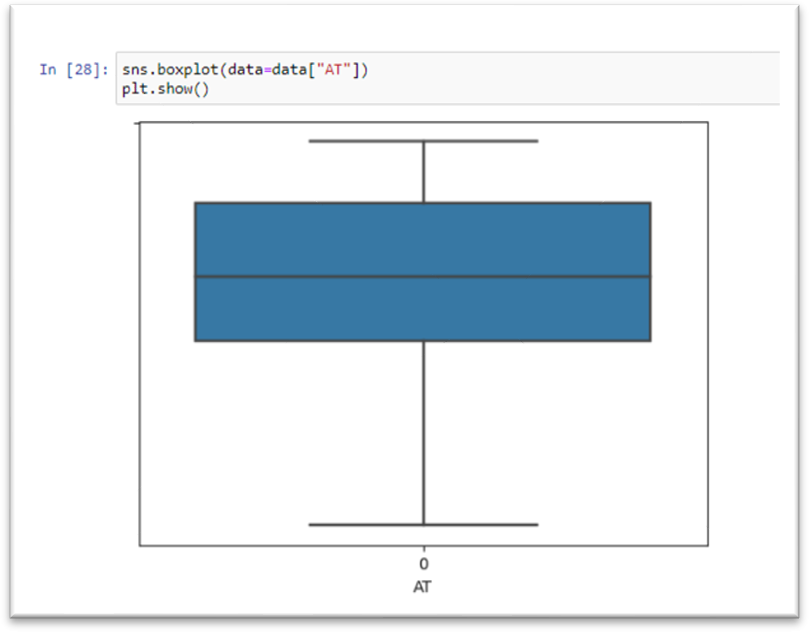
4)IN Box plot Q2 is approximately at center

We can Say That the “Waist” data is Normally Distributed









1) MEAN ≠ MEDIAN

2) Skewness, is not nearer zero 3)Kurtosis = -2.855 is not nearer to zero

4)IN Box plot Q2 is not at center and whisker is More in Positive side

We can Say That the “AT” data is Moderately Positively Skewed Data.

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

ANS=

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Alpha(α)**  **=(1-CL)/2** | **Z score (Z table)** |
| 90% | 0.10/2=0.05 | ±1.64 |
| 94% | 0.06/2=0.03 | ±1.88 |
| 60% | 0.40/2=0.20 | ±0.84 |

90% 🡪> qnorm(0.95)

[1] 1.644854

94% 🡪> qnorm(0.97)

[1] 1.880794

60% 🡪> qnorm(0.8)

[1] 0.8416212

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

ANS=

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Df** | **T score (t table)** |
| 95% | 25 | 2.060 |
| 96% | 2.060 |
| 99% | 2.787 |

95% 🡪> qt(0.975,24)

[1] 2.063899

96% 🡪> qt(0.98,24)

[1] 2.171545

99% 🡪 qt(0.995,24)

[1] 2.79694

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=

µ=270

=260

SD=90

n=18

df=n-1=18-1= 17

tscore= = = -10/21.23= -0.47

> pt(-0.47,17)

[1] 0.3221639

Required probability = 0.32=32%

Ans= Using Python

