|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| 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 | Ordinal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Ans:** There are total 3 coins with 2 sides (head and tail each). Total number of possible sides = 23 = 8 sides

|  |  |  |
| --- | --- | --- |
| Coin 1 | Coin 2 | Coin 3 |
| Head | Head | Head |
| Head | Head | Tail |
| Head | Tail | Head |
| Tail | Head | Head |
| Tail | Head | Tail |
| Tail | Tail | Tail |

Probability of coming two heads and one tail = **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:**

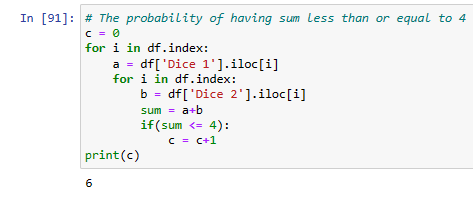
There are two dice with 6 sides each which will have numbers from 1 to 6.

Total number of possible results = 62 = 36

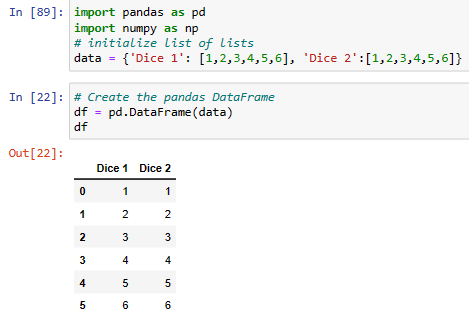
1. The probability of having occurrences which equals to 1 = 0 as minimum sum would be 2 in any combination in this scenario.
2. Probability of an event, P(A) = (Number of ways it can occur) ⁄ (Total number of outcomes)

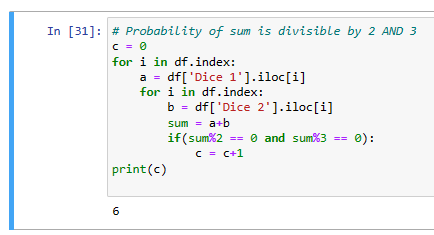
The probability of having sum less than or equal to 4 is 6/36 = 0.1666

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



1. Probability of sum is divisible by 2 and 3 = 6/36 = 0.1666





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 ball = 7

Number of balls which are non-blue = 5

Total combination from 7 balls with 2 balls each time = = = 7\*6\*5\*4\*3\*2\*1/(2\*1)\*(5\*4\*3\*2\*1)

= 21

Total combination from 5 balls (reduce 2 blue from total) with 2 balls each time = = = 5\*4\*3\*2\*1/(2\*1)\*(3\*2\*1)

= 10

Probability that the ball drawn are not blue is 10/21 = 0.476

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: Probability of expected number of candies for a randomly selected child is **3.09**

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

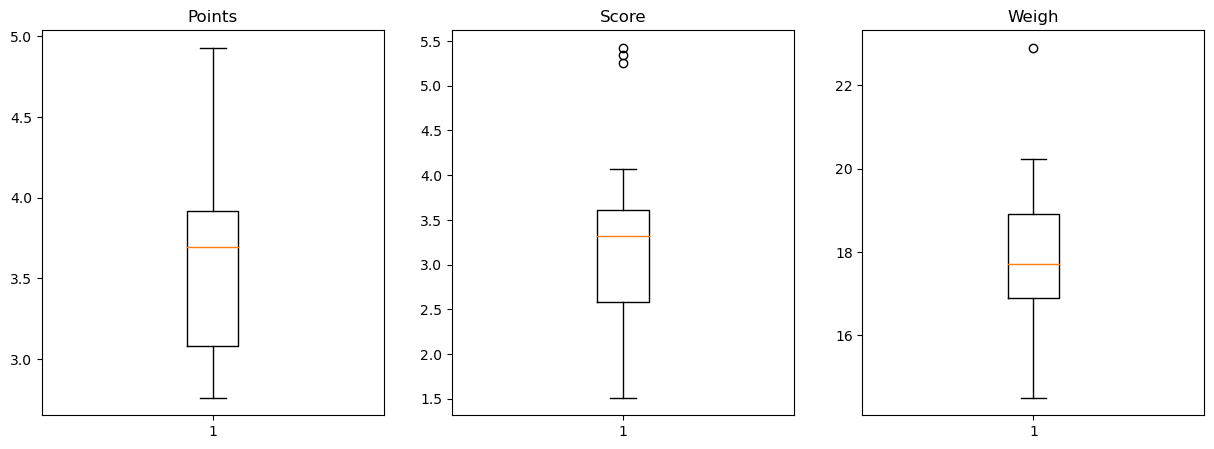
* For Points,Score,Weigh>

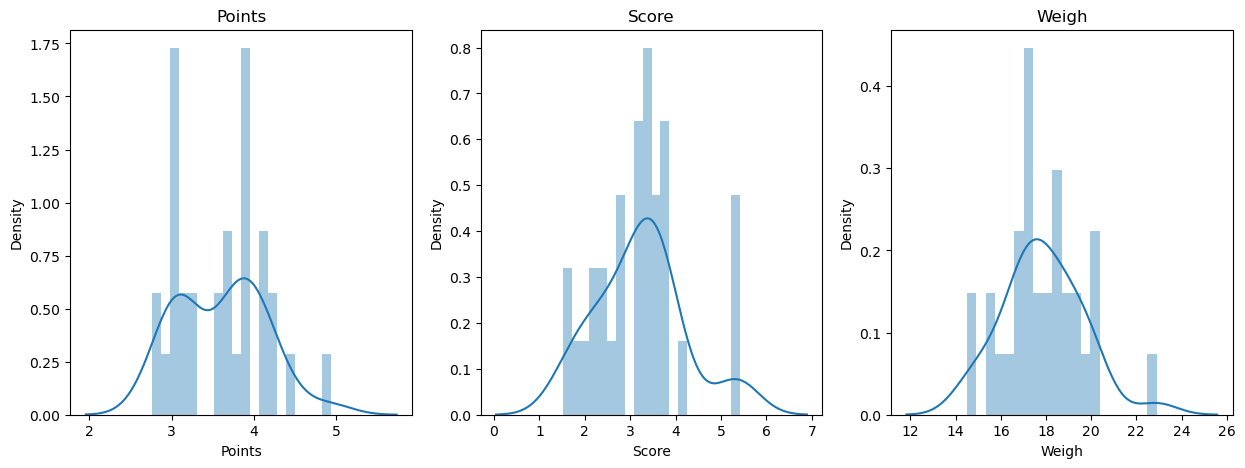
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

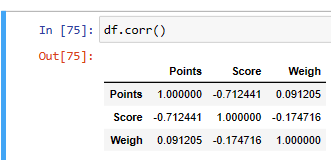
**Use Q7.csv file**

**ANS:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **mean** | 3.596563 | 3.21725 | 17.84875 |
| **median** | 3.695 | 3.325 | 17.71 |
| **mode** | 3.07,3.92 | 3.44 | 17.02,18.90 |
| **Variance** | 0.285881 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.534679 | 0.978457 | 1.786943 |
| **range** | 2.17 | 3.91 | 8.4 |

****

****

****

**Points**: More data concentrated below the mean and could see not a normal distribution data.

**Score**: More data concentrated below the mean and could see few outliers which resulted in right skewed distribution.

**Weigh**: More data concentrated above the mean and could see outlier which resulted in right skewed distribution.

There is no positive correlation between three observations, could see negative correlation between Points and Score.

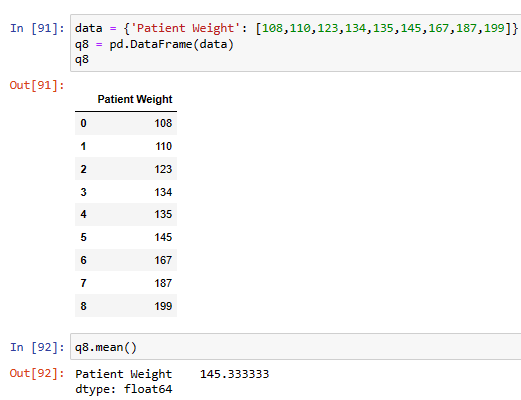
**===========================**

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**: 145.333333 pounds

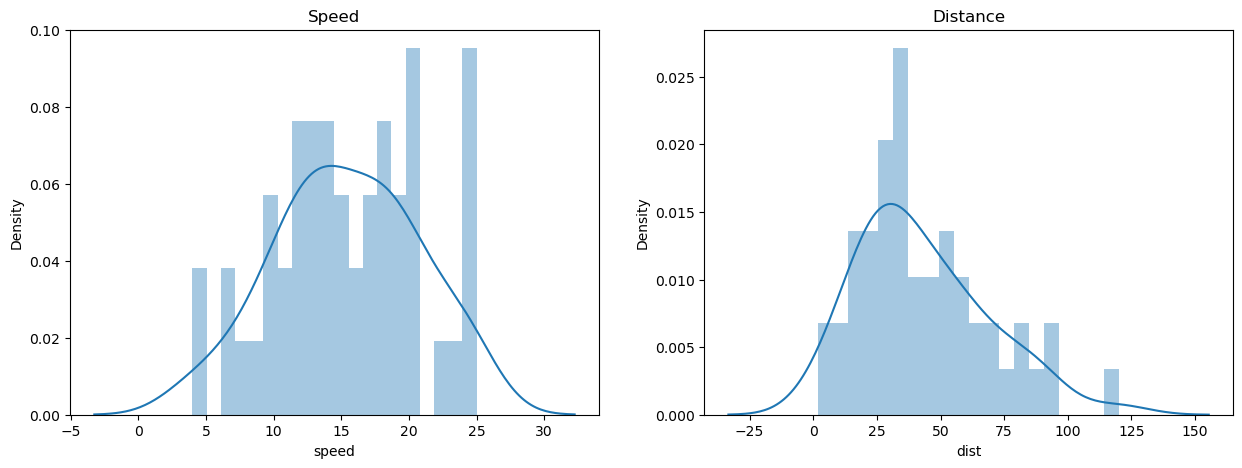
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:**

|  |  |  |
| --- | --- | --- |
|  | **speed** | **dist** |
| **Skewness** | -0.11751 | 0.806895 |
| **Kurtosis** | -0.508994 | 0.405053 |

****

**Speed**: Here mean, median and mode fall at different points and could see it is having negetive skewness.

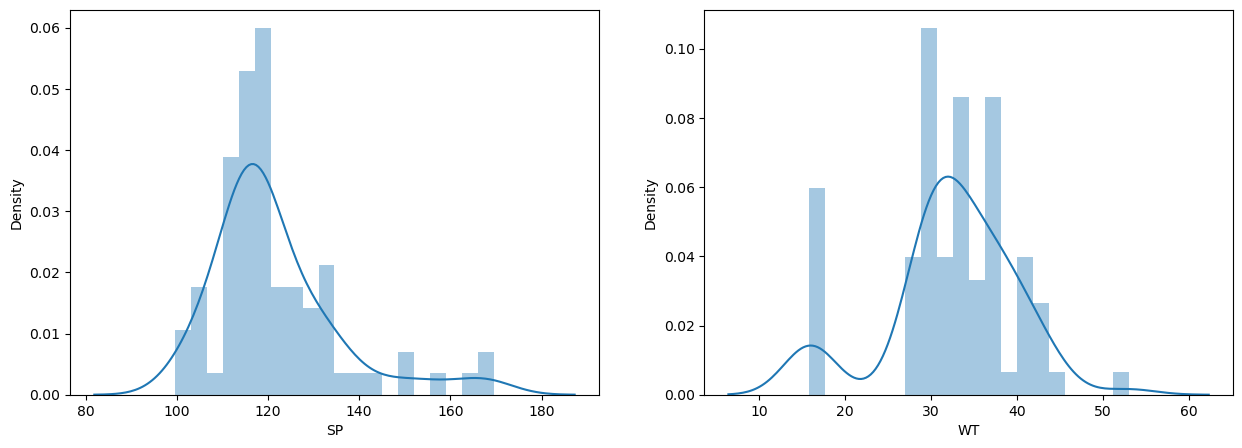
**dist**: Distance also having different mean, median and mode values but having positive skewness.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

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

****

**SP**: Here mean, median and mode fall at different points and could see it is having positive skewness.

Having negative kurtosis, distribution has flat peak (platykurtic)

**WT**: WT also having different mean, median and mode values but having negative skewness.

Having negative kurtosis, distribution has flat peak (platykurtic)

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



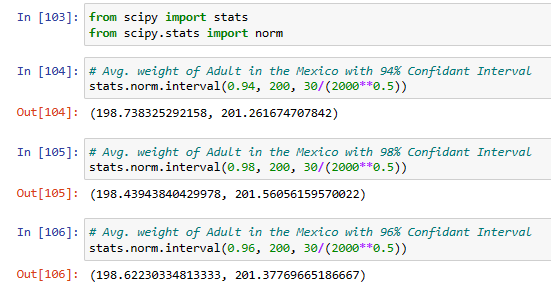
**Ans**: It is not a normal distribution data and data is having positive skewness. Most of the data concentrated between 50 to 150 and must be having some outliers. Mean>Median.



**Ans**: Outliers on the maximum side and data concentrated mostly above to the mean value.

**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**:

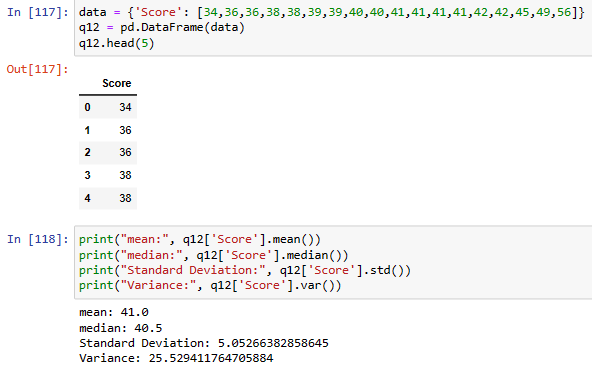


**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

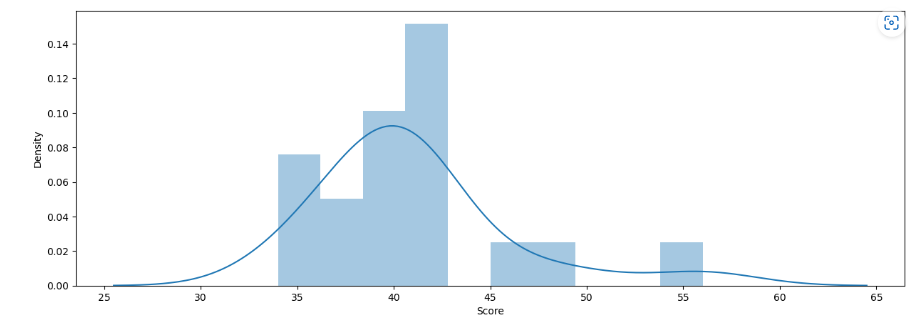
1. Find mean, median, variance, standard deviation.

**ANS**:



1. What can we say about the student marks?

**ANS**:



Most of the score data concentrated between 38 and 42 and it is positive skewness data

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

**ANS**: When mean == median, data is having normal distribution with no skewness.

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

**ANS**: When mean > median, data is right skewed where data concentrated to right side.

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

**ANS**: When mean < median, data is left skewed where data concentrated to left side.

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

**ANS**: Positive values of kurtosis indicate that distribution is thinner peaked and with thick tails.

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

**ANS**: Negative values of kurtosis indicate that distribution is thicker peaked and with thinner tails.

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



What can we say about the distribution of the data?

**ANS**: It is not normally distributed data with left skewed.

What is nature of skewness of the data?

**ANS**: It is having negative skewness where more data concentrated in left side. The whisker range of minimum value is greater than maximum one.

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

**ANS**: IQR = Q3-Q1

=18-10

= 8 approx.

Q19) Comment on the below Boxplot visualizations?



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

**ANS**: There are no outliers in the dataset. Secondly both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

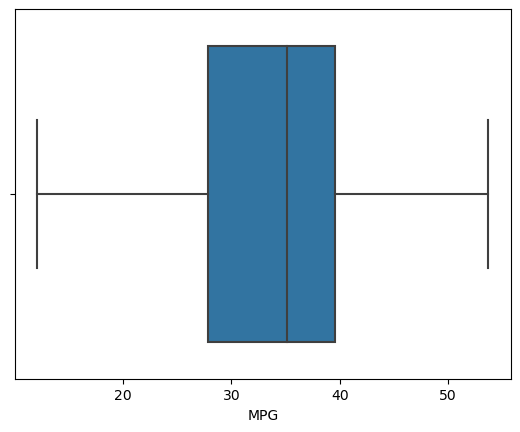
Data \_set: Cars.csv

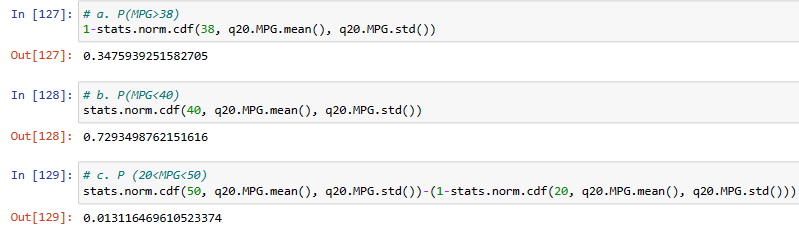
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

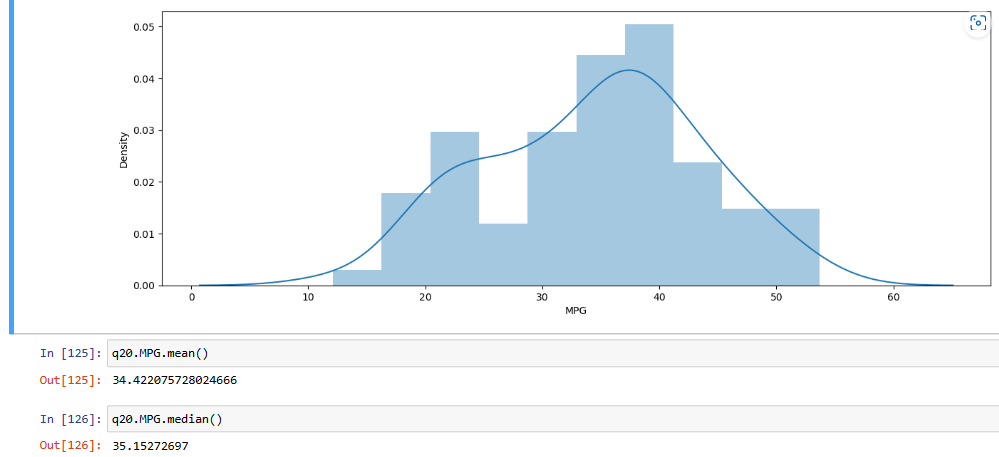




Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

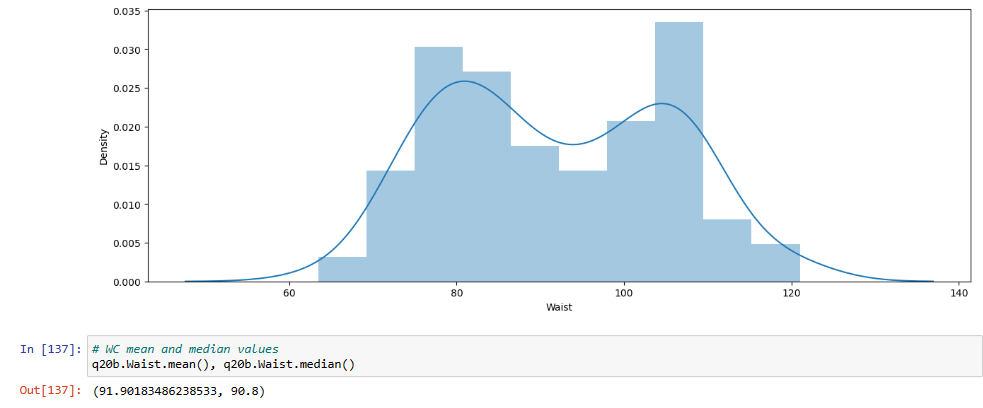
Dataset: Cars.csv

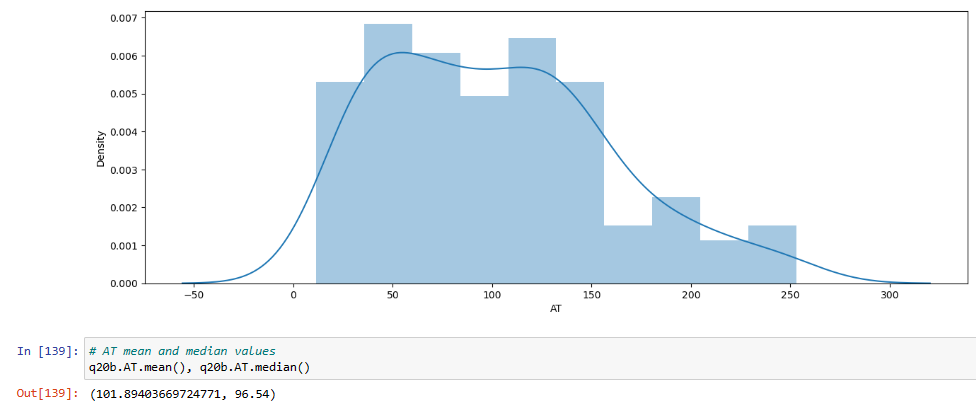
**ANS**: Data not normally distributed as we can see mean != median and data skewed left side.

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

Dataset: wc-at.csv

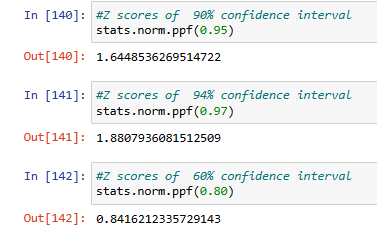
**ANS**: Data not normally distributed for both WC and AT as we can see mean!= median.





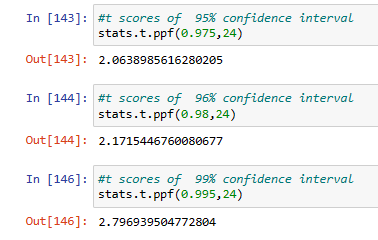
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**ANS**: Z-score values as below



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

**ANS:** t score values as below



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**: p value is 0.32, 32% probability is there that selected 18 bulbs would have an average life of more than 260 days.