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

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

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

Ans)

**Total No Of Possible OutComes=8(HHH,TTT,HHT,HTH,THH,TTH,THT,HTT)**

**No Of Possibilities For Getting 2 Head And 1tails=3(HHT,HTH,THH)**

**Probability = 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

No of possible outcomes =36

a)**No Possible For Getting Sum Equal 1=0**

**Hence Answer Is Zero**

B) **No Of Possibilities For Getting Less Than Or Equal To 4 =6**

**6/36=1/6**

**C) No Of Possibilities For Getting Sum Is Divisible By 2 And 3=6**

**6/36=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?

Sol**) total no of outcomes =9**

**No of possibilities that none of the balls drawn is blue=4**

**4/9=0.4444**

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

**Sol) Expected Number Of Candies For A Randomly Selected Child**

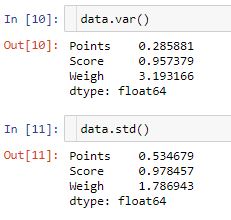
**1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09**

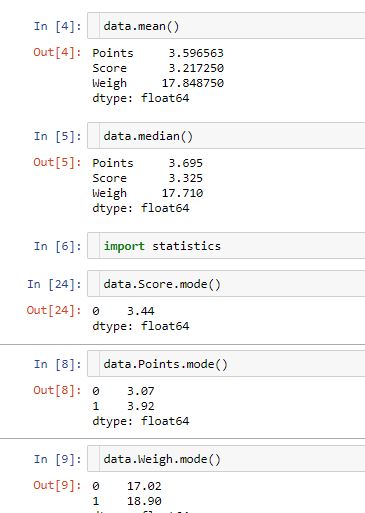
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

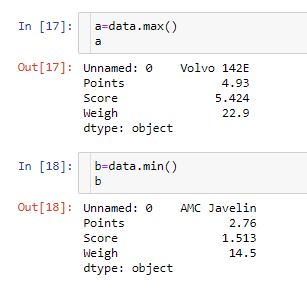
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Sol) From The Above Set Of The Data We Can Conclude That The Mean,Mode And Median Are Equal And The Data Set Follows The Normal Distriburion**





**Range=maxvalue-min value**



**Range for points=4.93-2.76**

**Range for score=5.424-1.513**

**Range for weigh=22.9-14.5**

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?

**Sol) 108+110+123+134+135+145+167+187+199=1308**

**Probability =1308/9**

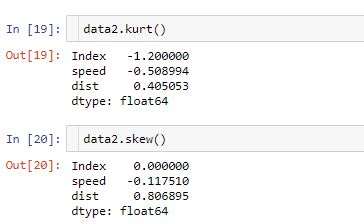
**145.34**

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

**Cars speed and distance**

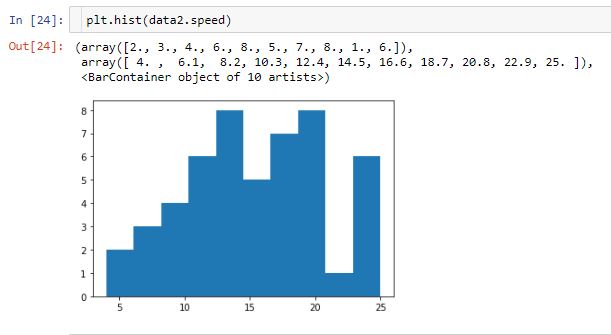
**Use Q9\_a.csv**

**Sol)**



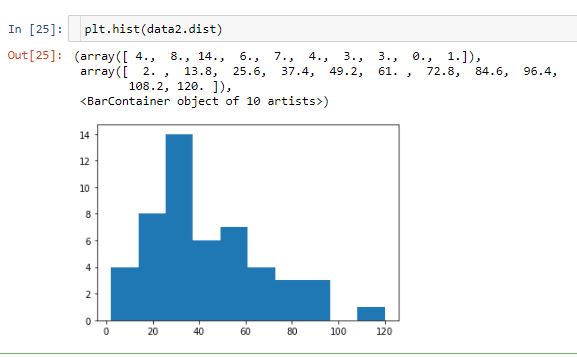
**Regarding to speed:**

**By observing the histogram of the speed we can conclude that is the negative skew as the most of the data is concentrated on the left side of the graph**



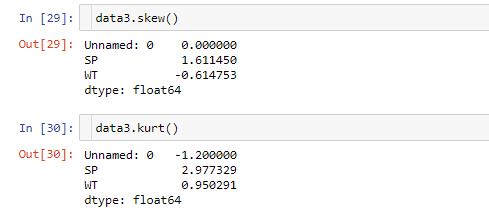
**Regarding to distance :**

**By observing the histogram of the Distance we can conclude that is the positive skew**

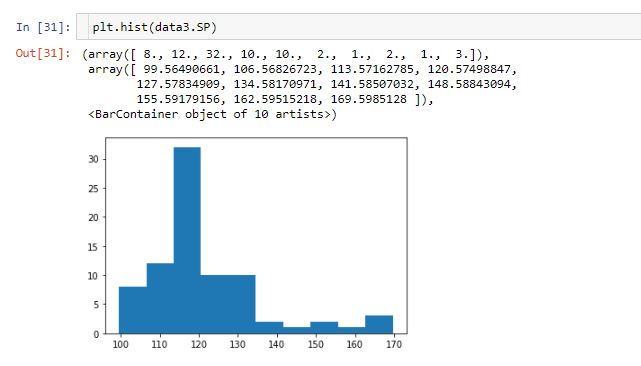


**SP and Weight(WT)**

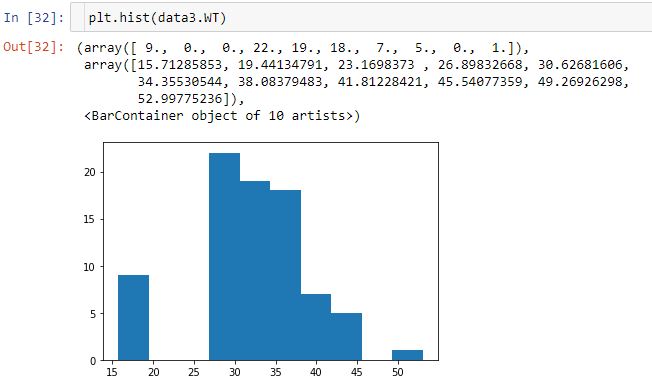
**Sol)skewness and kurtosis**



**Regarding to sp: by obseving the histogram of the sp we come to conclusion that sp has postive skewness because the max values on concentrated on the right hand side .**



**Regarding to the WT: by observing it is a positive skewed**



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



**Sol) By Observing The Histogram The Value Ranges From 0 To 400 And The Maximum Value Is Appering From 50 To 100 .It Is A Positive Skewed Plot As The Most Of The Value Is Concentrated On The Right Side .The Maximum Value Of The Data Is In The Range Of 170-180.It Doesn’t Have The Shape Of The Normal Distribution Curve.**



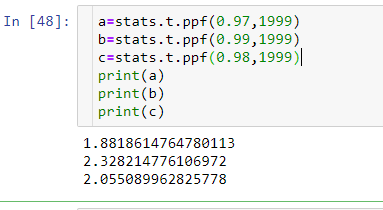
**Sol) By Observing The Box Plot Given Above We Can Say That It Is A Positive Skewed Curve As The Median Lies On The Side Of The First Quartile .The Range Between The Q1 And The Lower Outlier Is Less When Compare To The Q3 And The High Outlier.**

**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) **Sample Weight=200 Population Mean=?**

**Sample Deviation =30 Degrees Of Freedom=1999**

**T value=(sample mean-population mean)/sample std/sqrt(df)**



For 94% of confidence interval population mean =197.19

For 98% of confidence interval population mean = 196.54

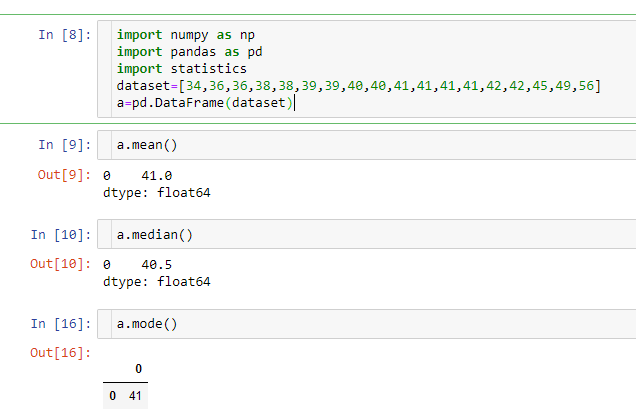
For 96% of confidence interval population mean =196.94

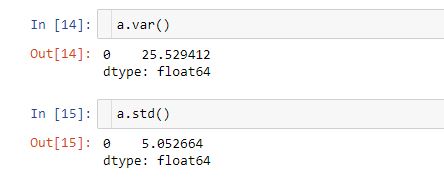
**Q12)** Below are the scores obtained by a student in tests

**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)** **From The Student Marks We Can Say That The Max Value Was 56 And The Min Value Is 34.The Range Of The Student Marks Is 22. The No Of Observation Is 18.The Mode Of The Data Is 41.**





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

Sol) **when the mean,median and mode are equal the curve is perfect bell shape curve and symmetric in nature .the nature of the skewness is “0”**

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

**Sol)The nature of skewness of the data when the mean >median is positive and it is right skewed data.**

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

**Sol)The nature of skewness of the data when the median >mean is negative and it is left skewed data.**

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

**Sol)The kurtosis is the measure of the peakdness of the data. A positive kurtosis indicates that the data is having a sharp peak and thick tails.**

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

**Sol)a negative kurtosis indicates that the data is distributed with out sharp peaks and distributed uniformly having thin tails**

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



What can we say about the distribution of the data?

**Sol) from the following data we can conclude that mean <median and the distribution of the data is concentrated on the left hand side of the graph.**

What is nature of skewness of the data?

**Sol) the nature of skewness of the data is negative and it is left skewed data.**

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

**Sol) The IQR of the data is given by 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.

**Sol) By looking to the distribution of data we can say that the median,mode and of the two box plot is equal and the both of them have symmetric or normal distribution.the boxplot 2 represent the population parameters where as the boxplot 1 represent the sample of the population.**

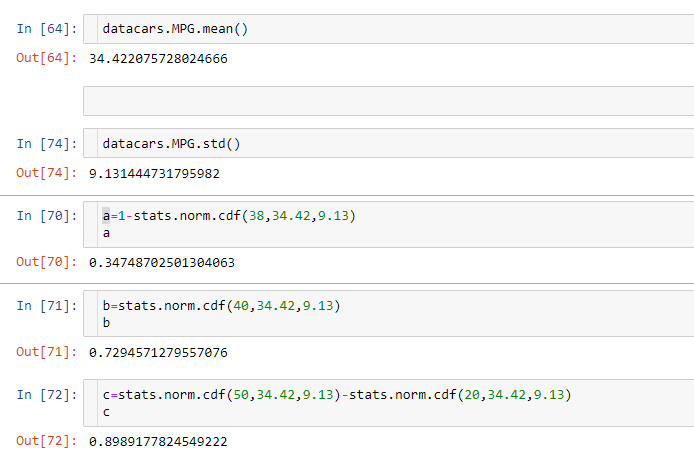
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

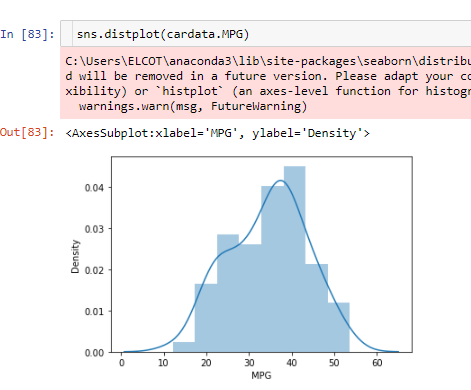
MPG <- Cars$MPG

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

Sol) 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

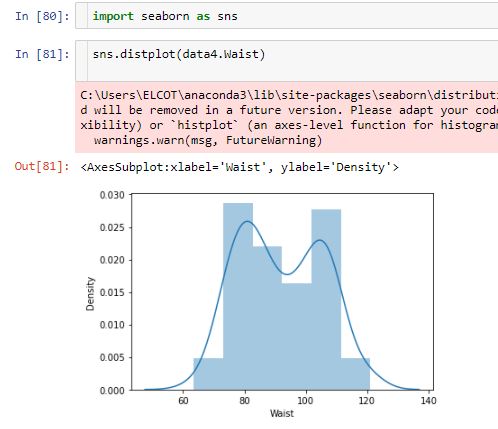


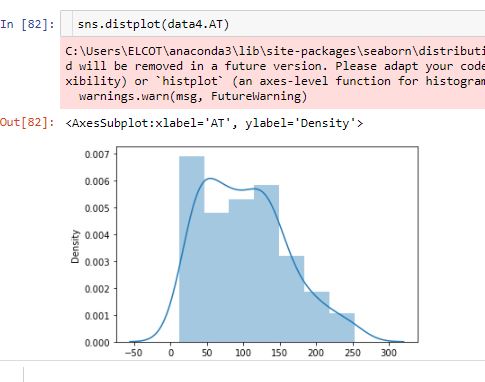
**By Observing The Distribution plot it Doesn’t Not Follow The Perfect Normal Distribution .It Has A Negative Skewness As The Most Of The Data Is Concentrated On The Left Hand Side .The Median Of The Data Is Less Than The Mean.**

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

Dataset: wc-at.csv

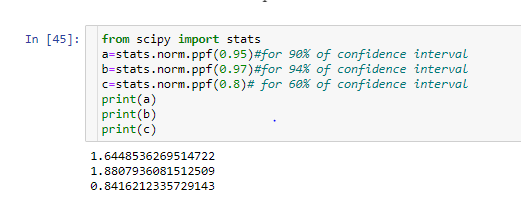
**Sol)By obseving the data distribution plot of the both WAIST and AT we can say that the data is not distributed normaly and have a positive skewness and the most of the data is concentrated on the right hand side**





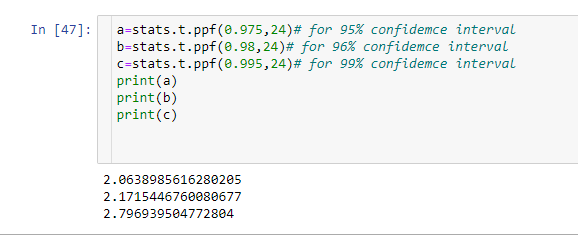
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Sol)**



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

**Sol)**



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) given data**

**Population mean =270**

**No of samples = 18**

**Sample mean = 260**

**Standard deviation= 90**

**For find the t value =(260-270)/90/(sqtr18)**

**=-0.4714**

**the probability that 18 randomly selected bulbs would have an average life of no more than 260 days=32.16%**

