|  |  |
| --- | --- |
| 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) | 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 | Ratio |

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

Ans: Total no of events=8

Favourable events= 3

P(E)=3/8

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 2and 3

Ans: Total no of outcomes=(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)=36

1. Probability of sum equal to one= 0%
2. Less than or equal to 4= 6/36=1/6
3. Sum is divisible by 2 and 3=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?

Ans: total no of outcomes= nCr=7C2=7!/(5!\*2!)=21

No of favourable outcomes=5C2=10

P(E)=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.

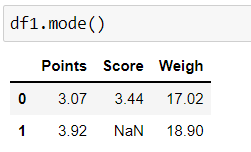
**Use Q7.csv file**

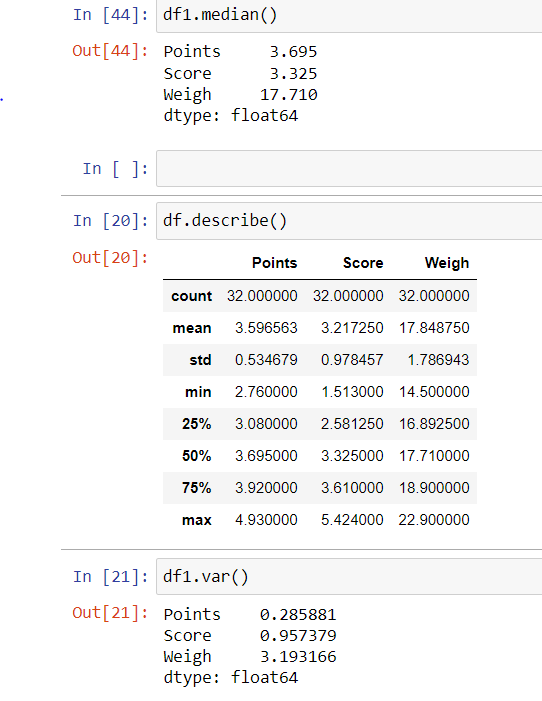
**Ans**: df.describe(): - This will give data set with values like mean,std,max,min etc.

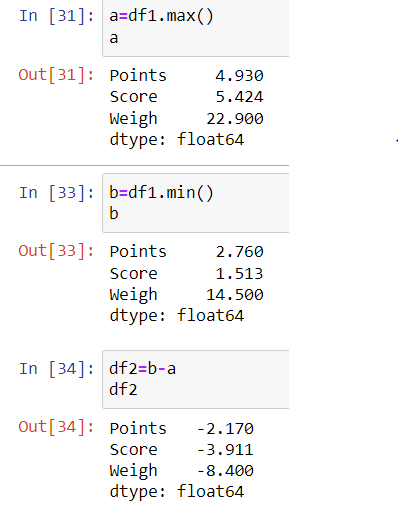
Df1.var(): this gives variance for given dataset for columns with numerical values.

Range= max-min

Df1.mode()= This will give mode for each columns



****



Here, range= df2=b-a i.e : For points=2.170

Score=3.9111

Weigh=8.400

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?

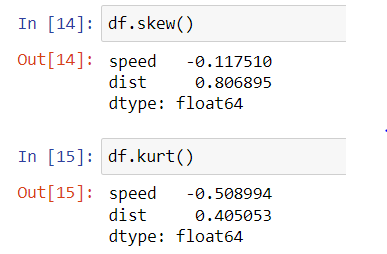
Ans: Expected Value=(108+110+123+134+135+145+167+187+199)/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: a)** Skewness and Kurtosis for speed and distance are below:

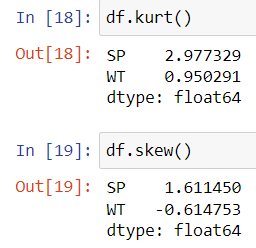
****

**Inferences:** As you see from above data there is a huge differences in the kurtosis value and this is because different equation used by the packages to find kurtosis.

**Use Q9\_b.csv**

**SP and Weight(WT)**

**Ans : b) Skewness and Kurtosis for SP and Weight are below :**

****

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



Ans: The above Boxplot suggests that the distribution has a lots of outliers toward upper extreme.

**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: N=3000000,n=2000

Mean=200,std.dev (S)=30

C=94%

Delta=1.96\*(30/sqrt2000)

=1.314

Interval=[(200-1.314)-(200+1.314)]

=[198.686,201.314]

C=98%

Delta=2.324\*(30/sqrt2000)

=1.560

Interval=[198.44,201.56]

C=96%

Delta=1.96\*(30/sqrt2000)

=1.31

Interval=[198.686,201.314]

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

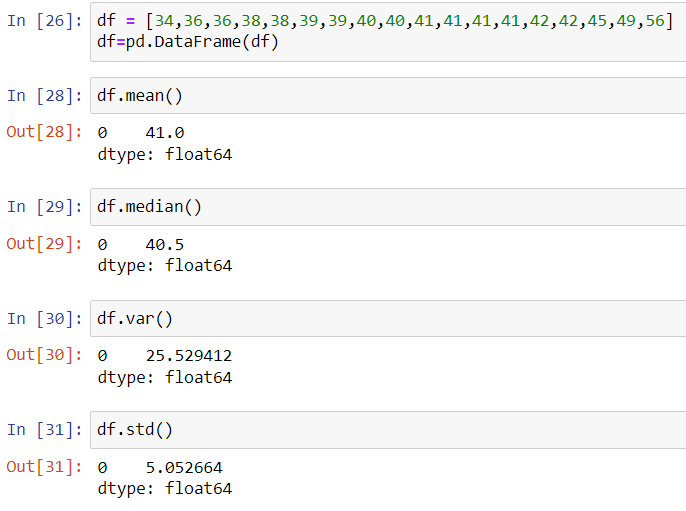
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean,median,variance,standard deviation.
2. What can we say about the student marks?

Ans:1)

df = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

Df.mean(),df,median(),df.var(),df.std() are below respectively



Ans 2): Mean>Median

This means that the distribution is slightly skewed towards right. No outliers are present.

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

Answer:- If the mean and median are equal then the skewness of the distribution is Zero. The distribution is symmertric.

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

Answer:- The distribution is Right skewed or positively skewed(Tail on the right side)

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

Ans:- If the median is greater than mean the distribution is negatively skewed(Tail on the left side)

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

Ans: It indicates that the distribution is peaked and possess thick tails. An extreme positive kurtosis indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean.

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

Ans: It indicates that the distribution is flat and has a thin 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 a normal distribution.

What is nature of skewness of the data?

Ans: It is left skewed.

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

Ans: IQR(inter quartile range) =Upper Quartile- Lower Quartile

= 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: The median of two boxplots are same that is 260 and outliers doesn’t exist in both the boxplots and they are neither positive skewed nor negative skewed.

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

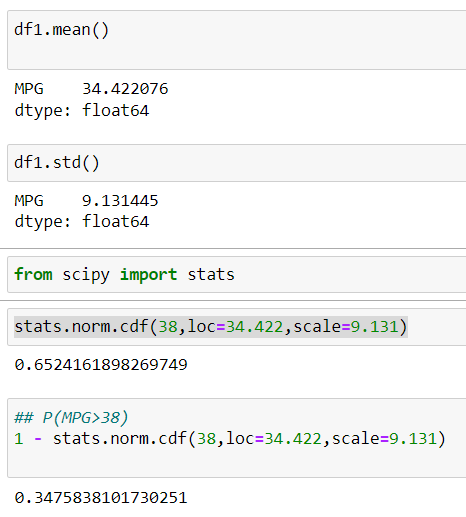
Data \_set: Cars.csv

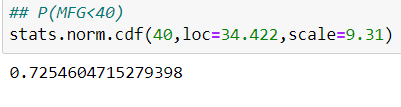
Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

Ans:a)



c)  

Q 21) Check whether the data follows normal distribution

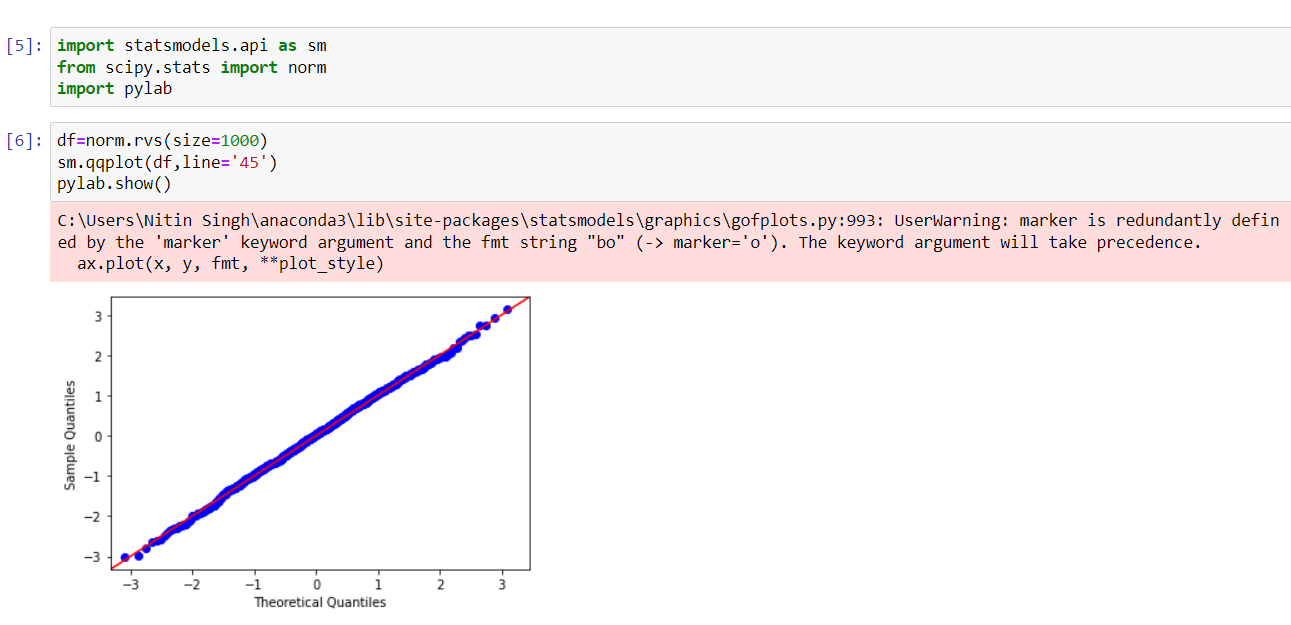
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

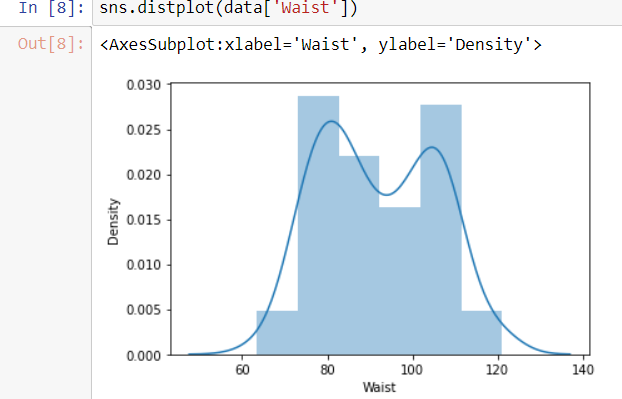
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

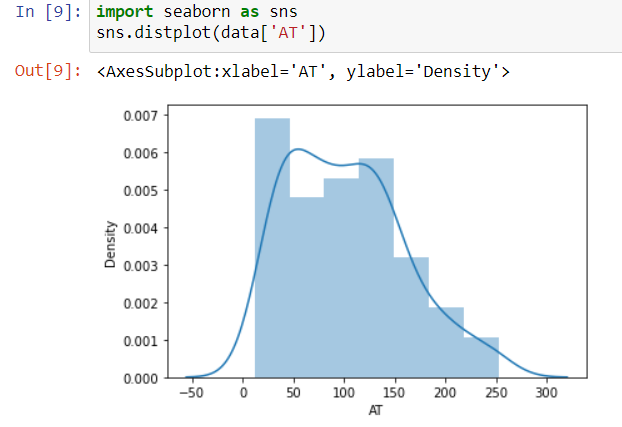
Dataset: wc-at.csv

Ans:a) This QQ plot of MPG of Cars shows that dataset follows the normal distribution

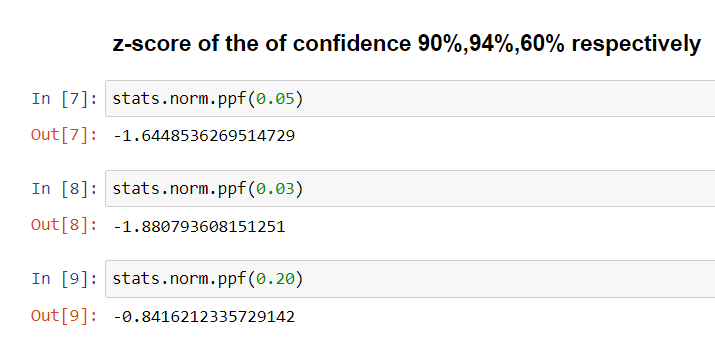


Ans: b) The graph below shows that dataset is not Normally distributed





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: At n=25

(n-1)=24

t-score at 95% of C=2.064

t-score at 96% of C=2.064

t-score at 99% of C=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

Ans: Here n is 18 which is less than 100 so, we go for t –distribution .To calculate t-statistics

Mean of the sample of bulb (x)=260

Population mean=270

Standard deviation=90

t =(260-270)/(90/sqrt18)

t = (-1\*sqrt2)/3

t=-0.471

For probability calculation no of degree of freedom is (n-1),so here you need the t-distribution with 17 degree of freedom .

The probability that t<-0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained with 17 degrees of freedom and a t-score of -0.471,the probability of the bulbs lasting less than 260 days on avg of 0.3218 assuming the mean life of the bulbs is 300 days.

OR (using python)

