|  |  |
| --- | --- |
| **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 | **Nominal** |
| Level of Agreement | **Ordinal** |
| IQ(Intelligence Scale) | **Ratio** |
| Sales Figures | **Ratio** |
| Blood Group | **Nominal** |
| Time Of Day | **Ratio** |
| Time on a Clock with Hands | **Ratio** |
| Number of Children | **Nominal** |
| Religious Preference | **Ordinal** |
| Barometer Pressure | **Ratio** |
| SAT Scores | **Interval** |
| Years of Education | **Nominal** |

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

**Sol:**

Three coins tossed:

Number of outcomes(**X**):

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

Probability of getting two heads and one tail:

**P(X) =NO.OF POSSIBLE OUTCOMES/TOTAL NO.OF OUTCOMES**

**HHT+HTH+THH=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

**Sol:**

Two dice are rolled possible values:

**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:**

n(A)=0

P(A)=n(A)/n(S)=0/36=0

**P(A)=0**

1. **Less than or equal to 4:**

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

P(B) = n(B)/n(S)

= 6/36=1/6=0.06

**P(B)=0.06**

1. **Sum is divisible by 2 and 3:**

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

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

= 24/36=1/3

**P(C)=0.03**

**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 number of balls (S) = 2+3+2=7

## Then Number of ways of drawing 2 balls n(S)= = 7c2=7\*6/2\*1=42/2

## n(S) = 21

## Two balls are drawn at random that none of the balls drawn is blue as E

## n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls. = 5c2=5\*4/2\*1=20/2 n(E) =10

## Therefore P(E) = n(E)/n(S)​ =10/21

**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 |

**Sol:**

Child A – probability of having 1 candy = 0.015

Child B – probability of having 4 candies = 0.20

**Excepted number of candies E(X) =**  **∑ x\* P ( x )**

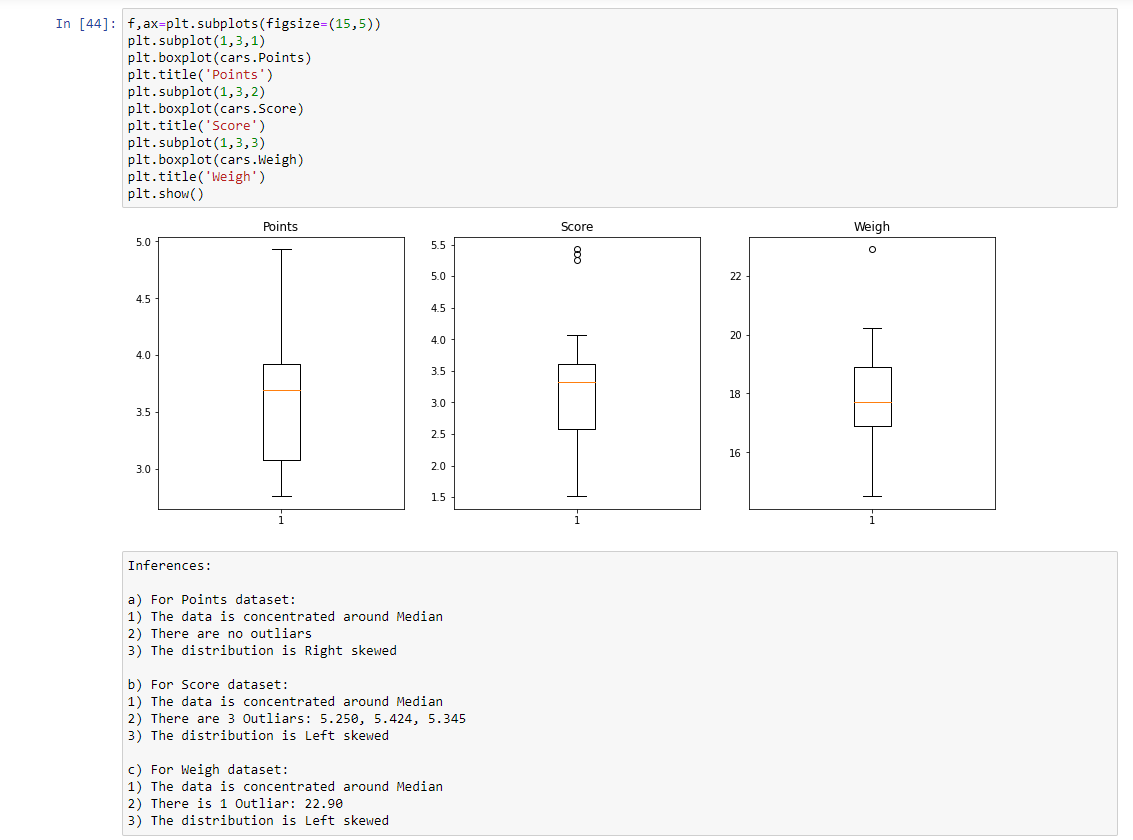
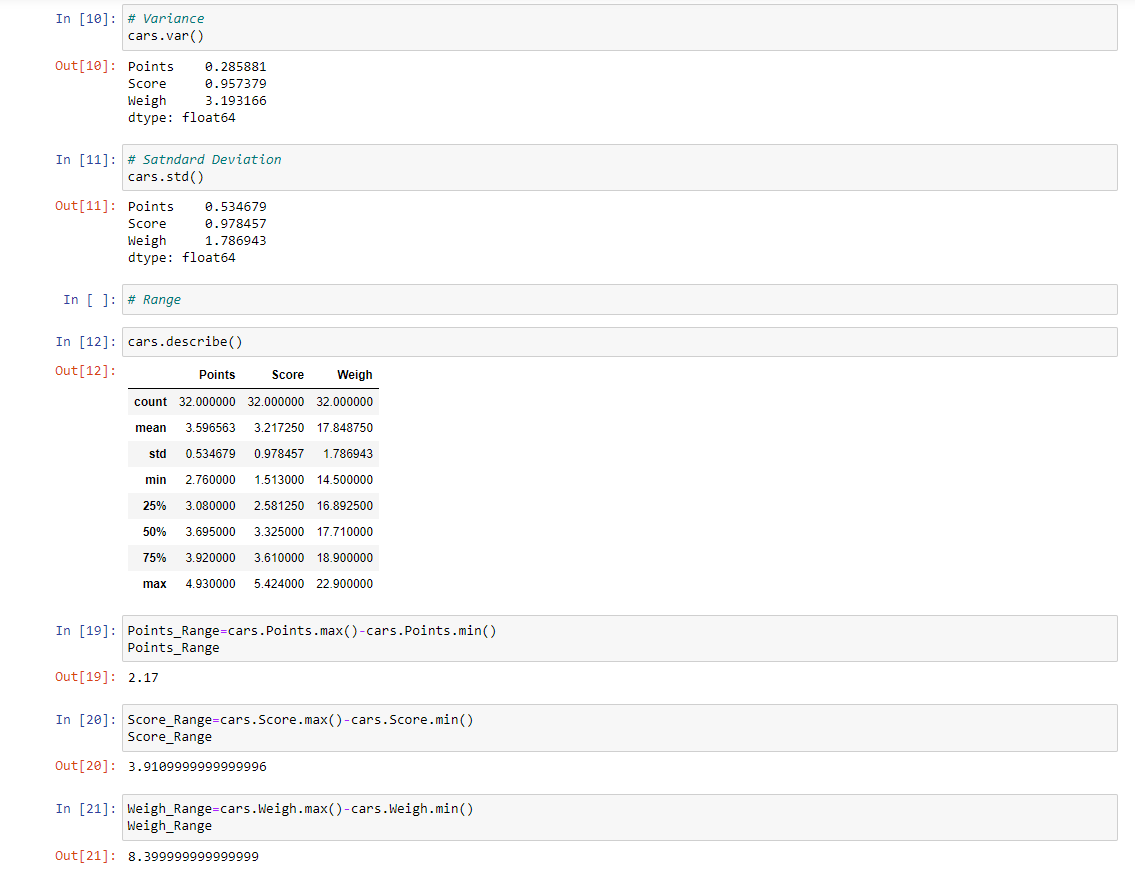
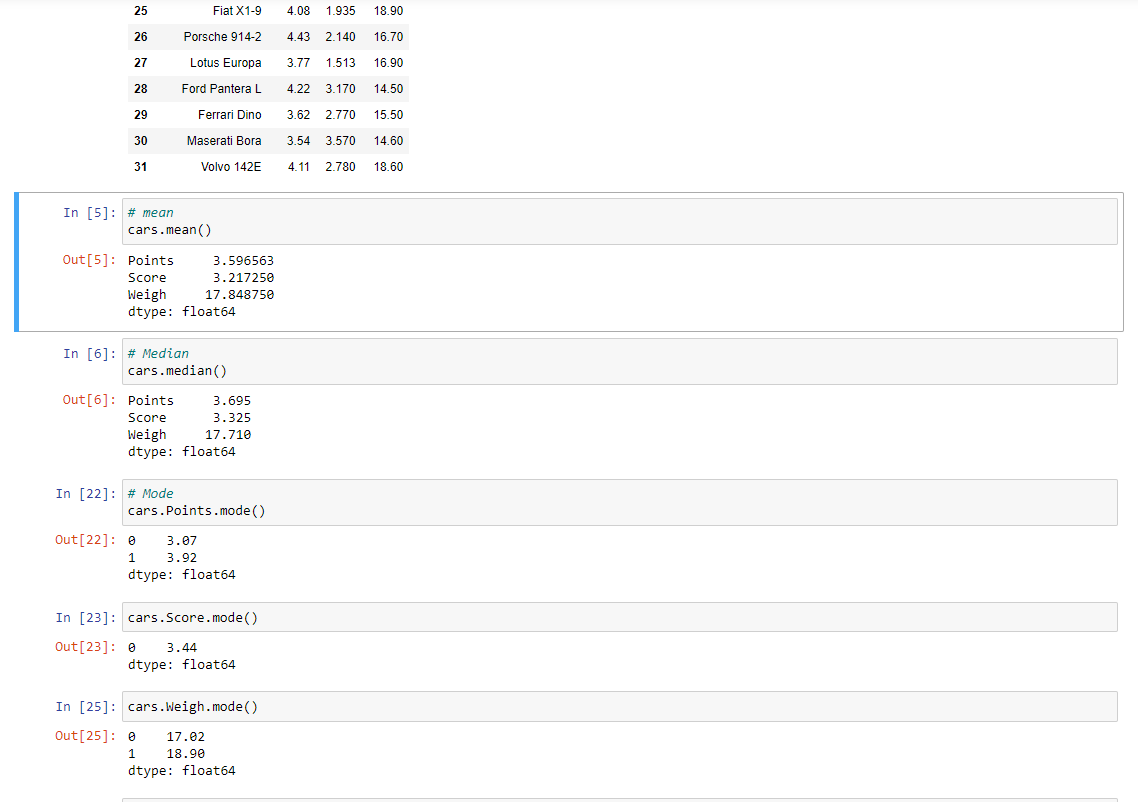
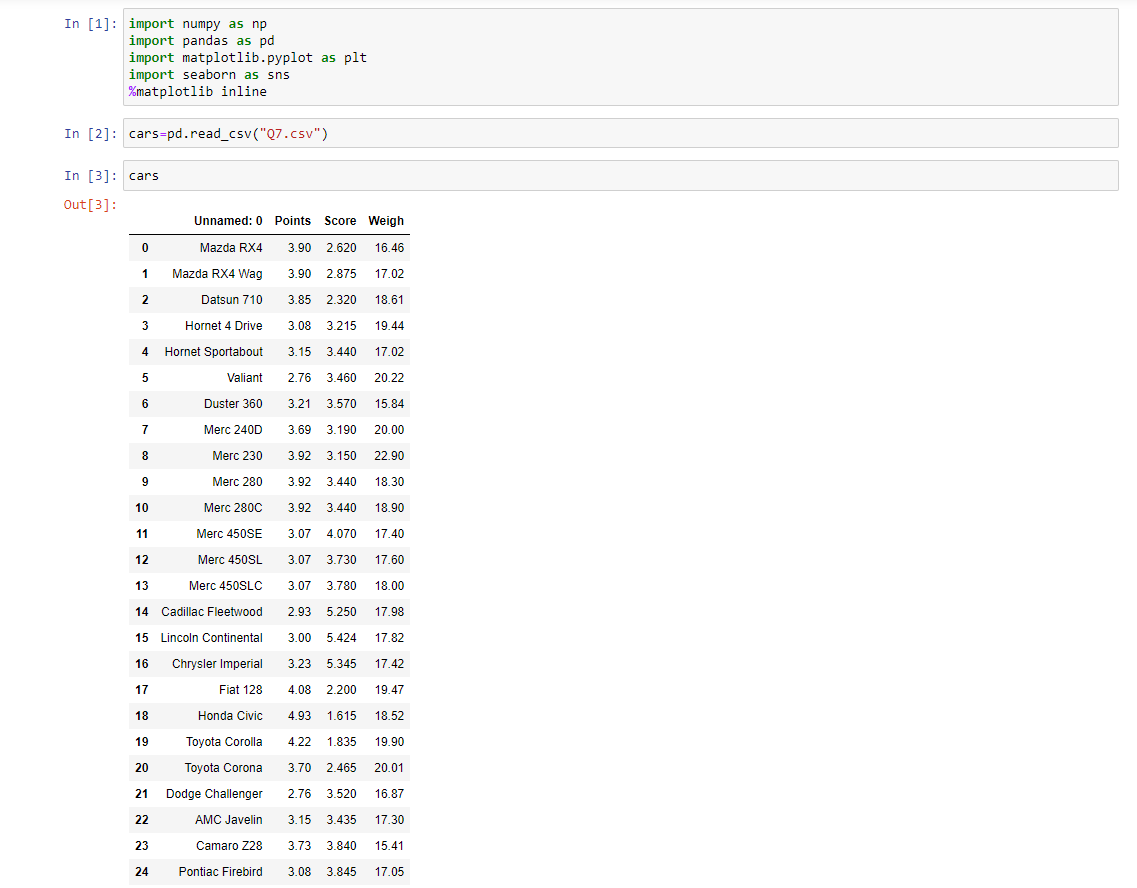
=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

**E(X)= 3.09**

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

* **For Points,Score,Weight**

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

**Use Q7.csv file **

**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?**

**Given:**

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

**Sol:**

There are 9 patients, when one of the person chosen at random **P(X) =1/9**

**Excepted number of patients E(X) =**  **∑ x\* P (x)**

=1/9\*108+1/9\*110+1/9\*123+1/9\*134+1/9\*135+1/9\*145+1/9\*167+1/9\*187+1/9\*199

=1/9[108+110+123+134+135+145+167+187+199]

=1/9[1308]

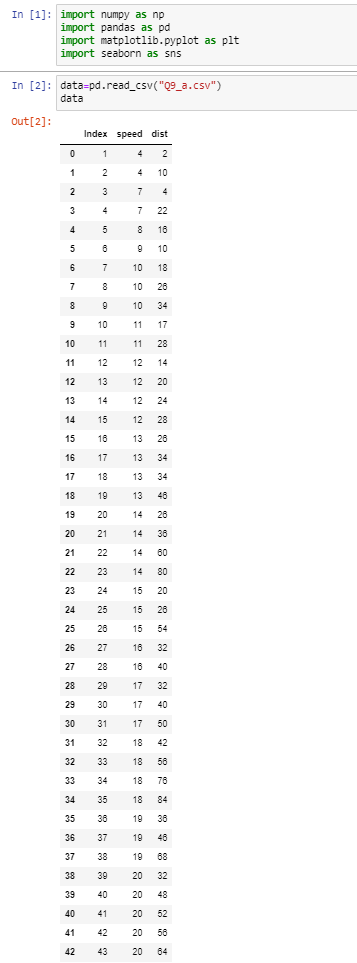
**Excepted Value of the Weight of that patient E(X) =145.33**

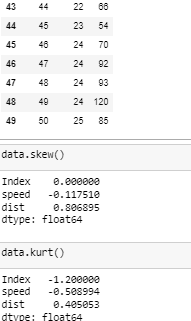
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Sol:**

**Q9\_a.csv**

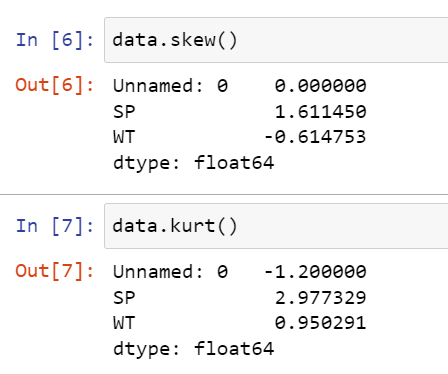




**SP and Weight(WT)**

**Sol:**

**Q9\_b.csv**



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



**Sol:**

**For above Histogram, most of the data points are concentrated in the frequency of (50-100) and the least value around frequency(0-10), hence it is observed that long tail towards right and hence it is a right skewed distribution where Skewness is positive.**



**Sol:**

**The above Boxplot shows that the distribution is right skewed 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?

**Given:**

Sample size n=2000

Population N=3000000

Sample mean https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%3D%20200

Standard deviation sigma=30

**Sol:**

**For 94%:**

**Confidence interval:** https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%5Cpm%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D

=1.96\*(30/sqrt 2000)

=1.314

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

**=[198.686,201.314]**

**For 98%:**

**Confidence interval** =2.324\*(30/sqrt2000)= 1.560

=[(200+1.560),(200-1.560)]

**=[198.44,201.56]**

**For 96%:**

**Confidence interval** =1.96\*(30/sqrt2000)=1.31

=[(200+1.31),(200-1.31)]

**=[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?

**Sol for 1:**

**Mean=41, Median40.5, Variance=25.52941, Standarddeviation=5.052664**

**Sol for 2:**

**Mean is greater than median, so it is a right skewed and has no outliers.**

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

When Mean=Median then it is **symmetrical distributed**

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

**Right Skewed, positively skewed**

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

**Left Skewed, Negatively skewed**

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

Indicates that the distribution is **peaked and has fat tails**

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

Indicates that the distribution is **flatten and has thin tails**

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



**What can we say about the distribution of the data?**

Median is not in the center of the plot and data is skewed, **thus distribution of data is not normally distributed.**

**What is nature of skewness of the data?**

Median is greater than Mean so, **negatively skewed**

**What will be the IQR of the data (approximately):** **IQR=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:**

**Mean and Median are equal in both boxplot1 and boxplot2 and hence it is symmetrically distributed and has no outliers.**

Q20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

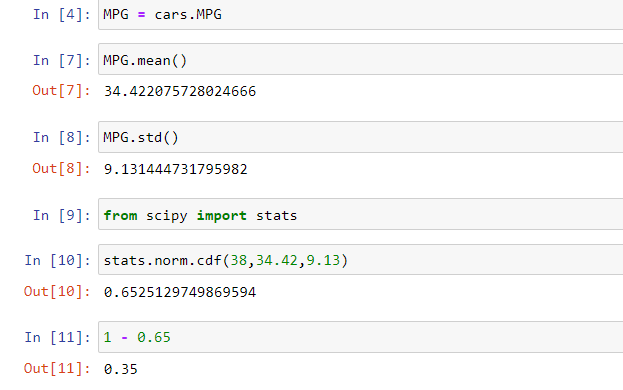
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

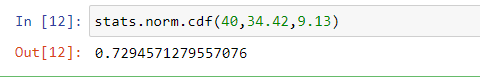
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**Sol: z =**

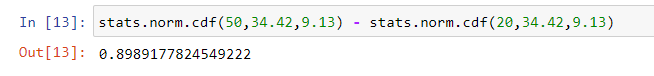
* 1. **P(MPG>38)**



**b. P(MPG<40)**



**c. P (20<MPG<50)**

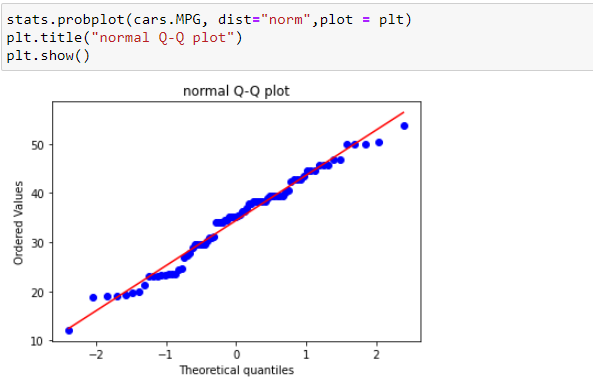


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

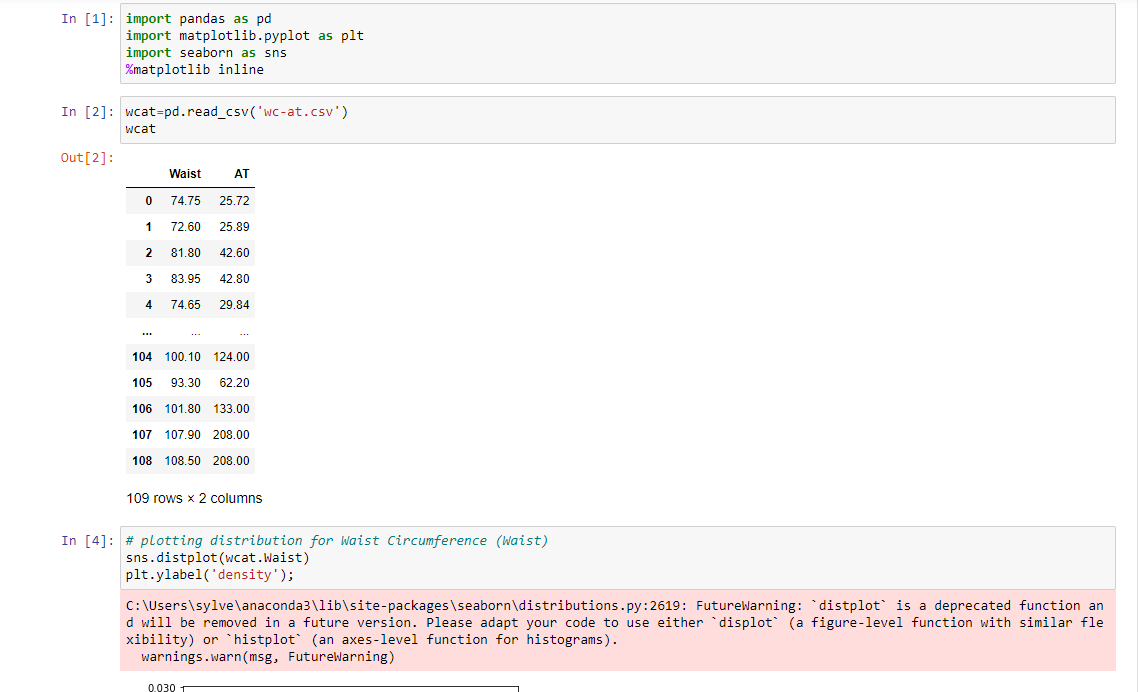
**Sol: Normal distribution indicates in Q-Q plot**

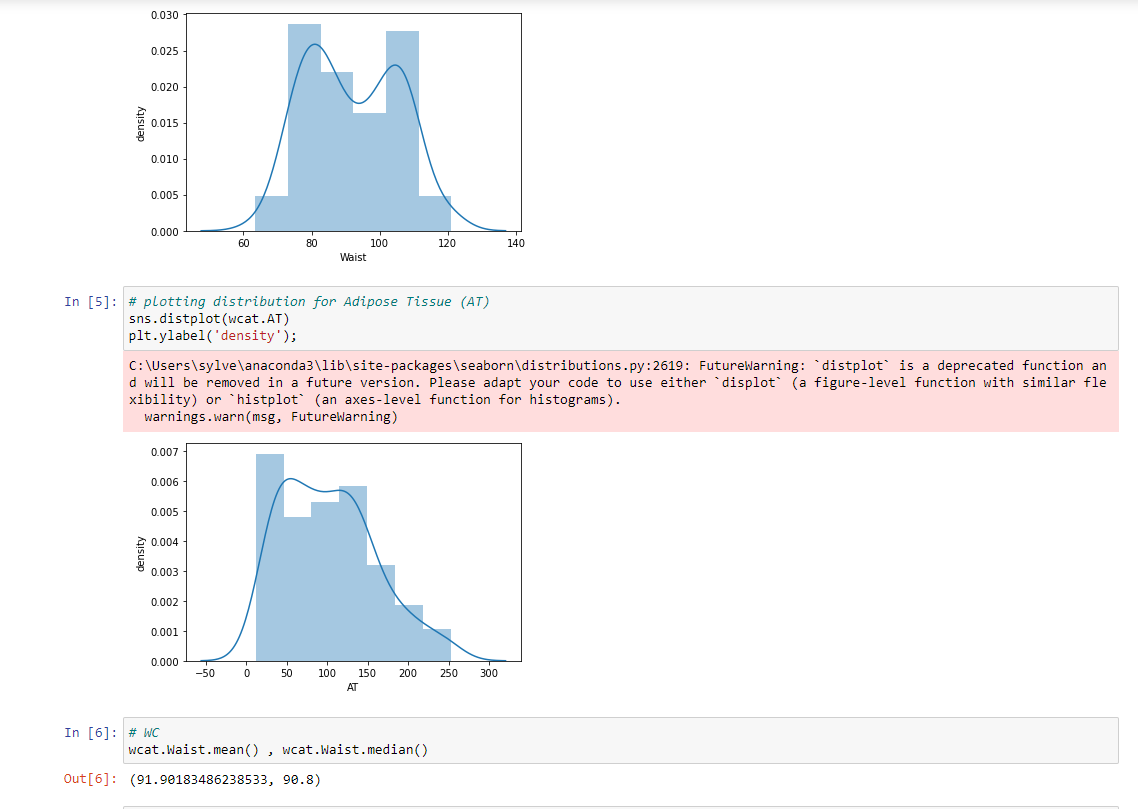


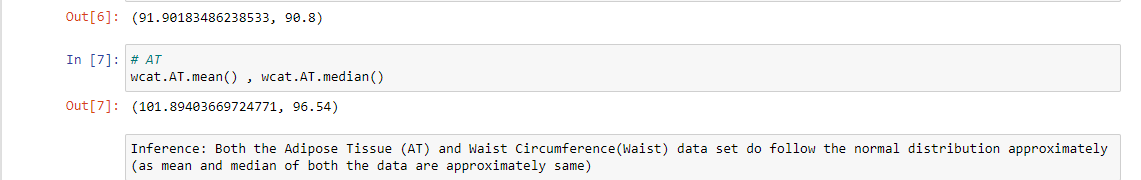
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Sol:**

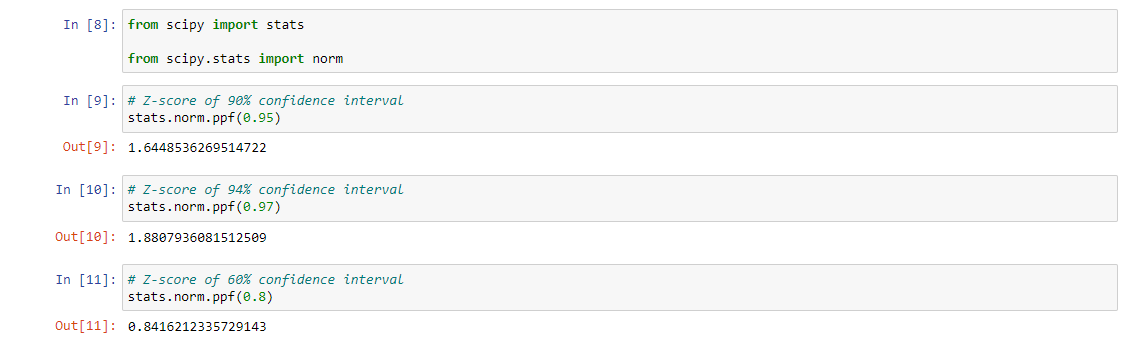






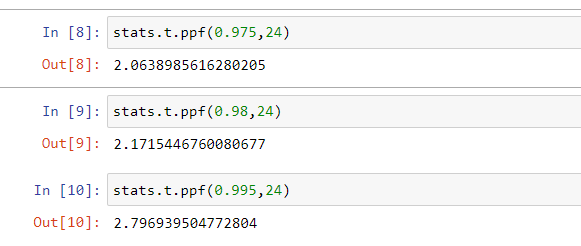
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Sol:Confidence interval for 90%,94%,60% are below**



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

**Sol: Confidence interval for 95%,96%,99% are 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

**Given:**

consider n is 18 which is less than 100 so, we go for t –distribution .

Mean of the sample of bulb (x)=260

Population mean(μ)=270

Standard deviation=90

**Sol:**

To calculate t-statistics:

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

t = (-1\*sqrt2)/3

**t=-0.471**

For probability calculation no of degree of freedom is (n-1),therefore degree of freedom is 17.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.