|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete(count) |
| Results of rolling a dice | Discrete(count) |
| 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(Categorical) |
| Number of kids | Discrete(count) |
| Number of tickets in Indian railways | Discrete(count) |
| Number of times married | Discrete(count) |
| Gender (Male or Female) | Nominal |

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 | Interval |
| Celsius Temperature | Ratio |
| Weight | Nominal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| 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 | Ratio |
| Religious Preference | Ordinal |
| 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?

Total no of desirable outcomes = 2^3 = 8

Two Head and One Tail = 3

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

Total no of desirable outcomes = 6^2 = 36

1. Equal to 1

As sum of two dice will never be equal to 1

**Probability is 0.**

1. Less than or equal to 4

Interested events = 5

So probability = 6/36 = 01.667

1. Sum is divisible by 2 and 3.

Interested events = 6

So probability = 6/36 = 1/6 = 1.667

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?

Total no of balls 7  
two balls drawn 2  
so 7C2 is 21  
No blue ball should drawn is 2  
except blue total no of ball is 5  
so 5C2is 10

Answer = 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 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

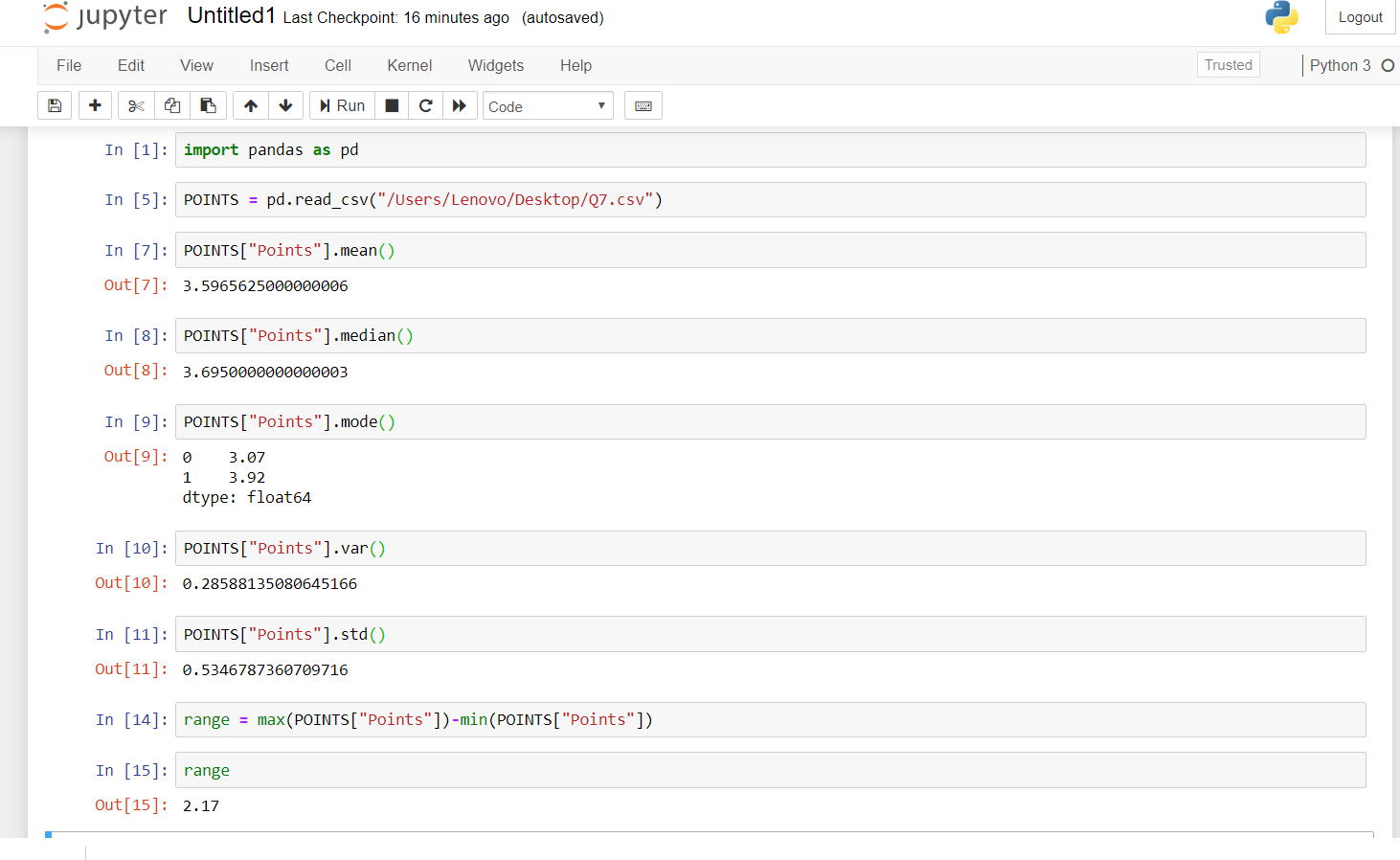
**= 3.09**

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

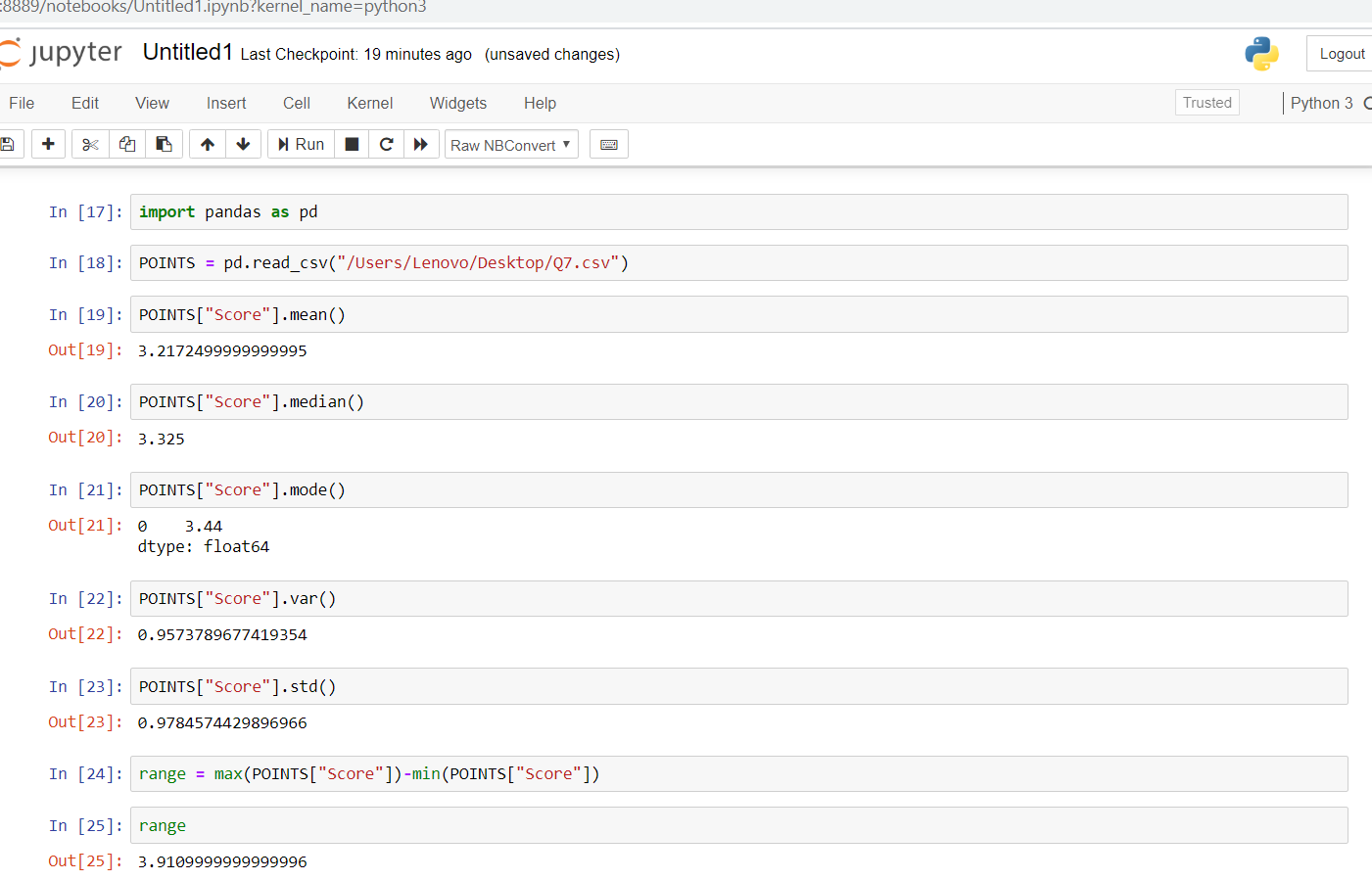
* For Points,Score,Weigh>

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

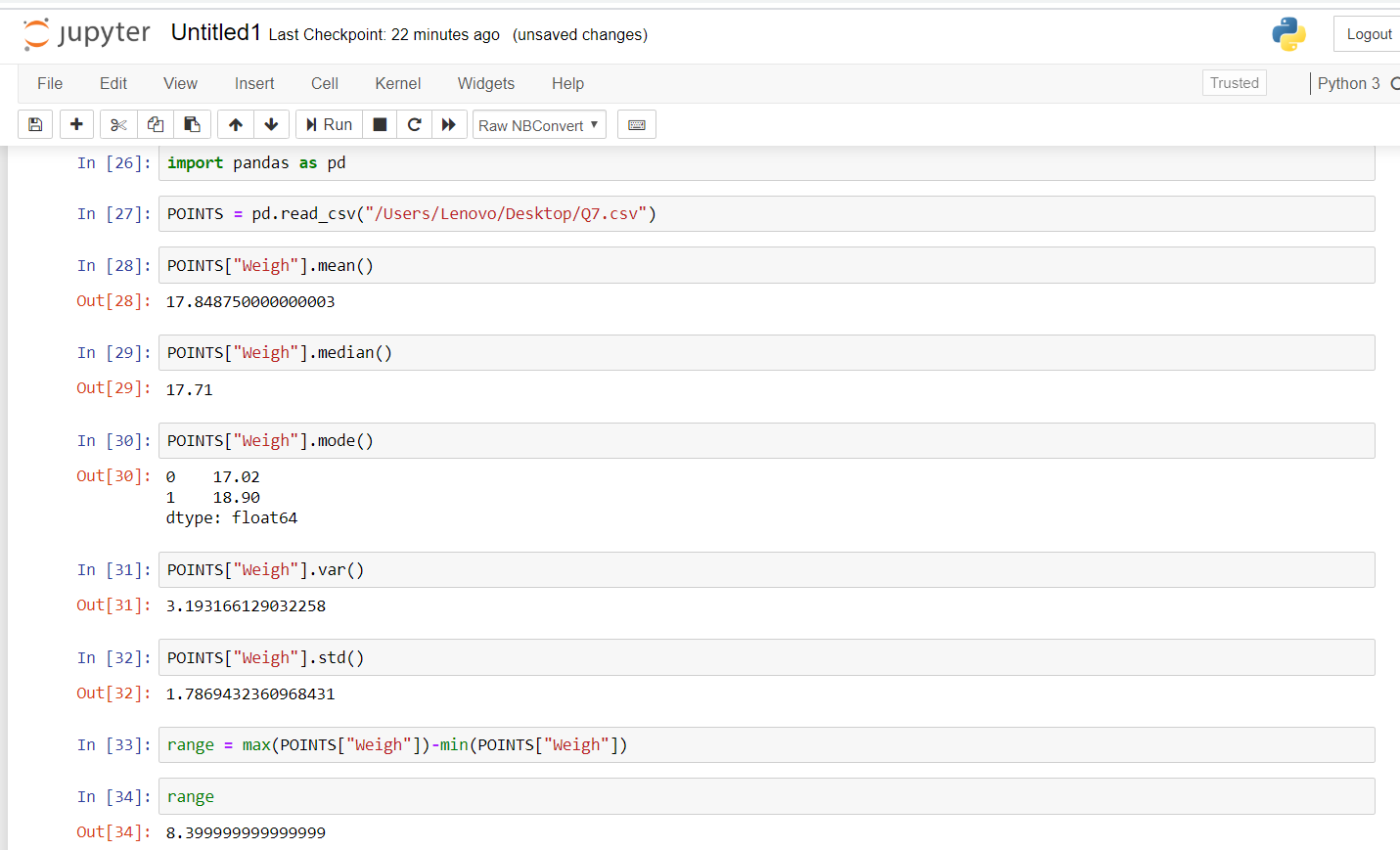
Ans) For Points-



For Score-



For Weigh-



Inferences-

As mean and median are close to each other, our data is normally distributed in

our graphs of frequency vs Points/Score/Weight

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

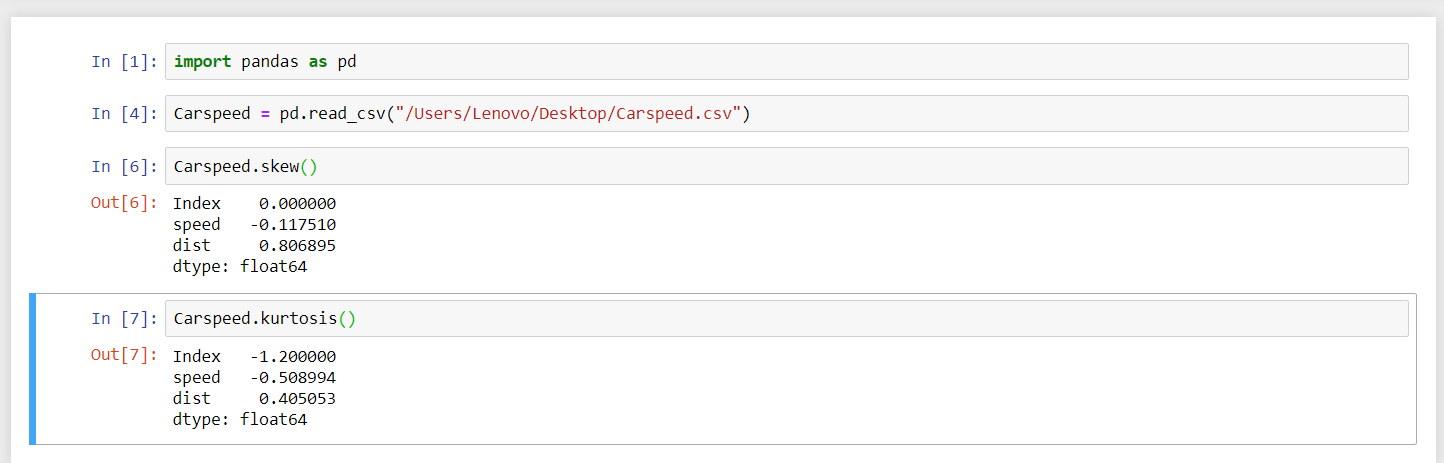
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Ans- 145.33

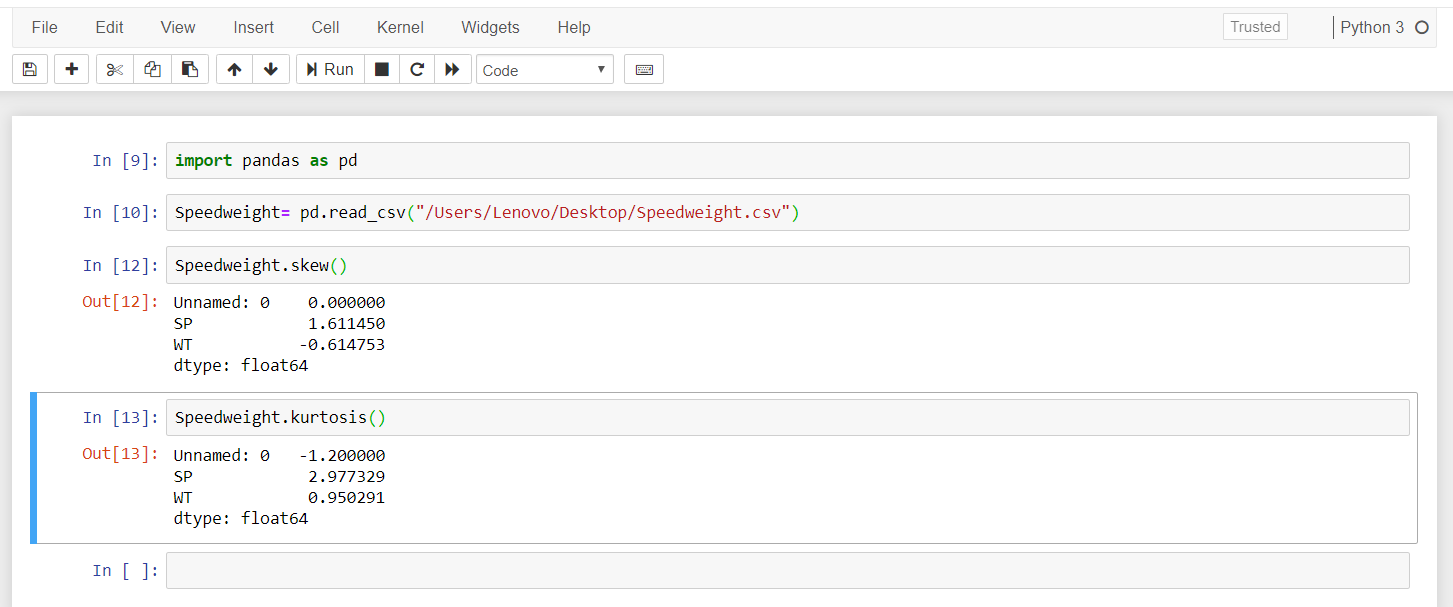
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Ans -Skewness & Kurtosis of car speed and distance is calculated in Python**

As we can observe, we have negative skewness for speed but the magnitude is low so it is nearly distributed, but skewness for distance is higher in magnitude which mean data set are not normally distributed and higher nodes are on left side

**SP and Weight (WT)**

****

As we can observe skewness are kurtosis are high and positive so high magnitude data points are available on the right side

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



Ans. We can conclude that mean and median lies in the range of 80-120. And the boxplot tell us that we don’t have many outlier in our graph, there are few but they can be neglected

**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 ?

> avg <- 200

> standdiv<- 30

> samp <- 2000

> error<- qz(confInt, df= samp-1 )

> se<- (standdiv/sqrt(samp))

> ci <- error\*se

> avg-ci

> avg+ci

[avg-ci, avg+ci]

For 94%, 98% and 96%.

Substitute the value of alpha respectively and we get the points as

[1] > error<- qt(0.97, df= samp-1)

[198.7376, 201.2624]

[2] > error2<- qt(0.98, df= samp-1)

[198.6214, 201.3786]

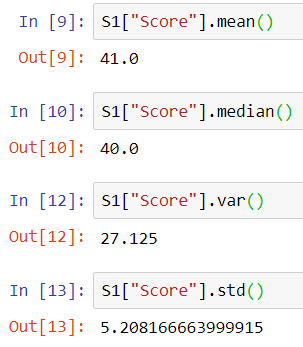
[3] > error3<- qt (0.99, df= samp-1)

[197.6718, 202.3282]

**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?



2)2 Students marks are outliers among the marks of the student

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

Ans. **Skeweness is zero**

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

**Ans. Skewness is Positive**

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

**Ans. Skewness is Negative**

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

**Ans. A distribution with a positive kurtosis value indicates that the data is Unimodel**

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

**Ans. A distribution with a negative kurtosis value indicates that the distribution has flatter peak than the normal distribution**

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



What can we say about the distribution of the data?

Ans. The data is distributed in right region of the graph. More values lies in between 14 and 16 and median is 15.

What is nature of skewness of the data?

Ans. Skewness is Negative as data points are in right side.

What will be the IQR of the data (approximately)?   
Ans. IQR = Q3-Q1

From above graph IQR (approx) = 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. As value changes of upper quartile range and lower quartile range in both the cases. Values of IQR is also change

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)

1. mean(Cars$MPG)
2. [1] 34.42208
3. > sd(Cars$MPG)
4. [1] 9.131445
5. 1-pnorm(38,34,9)
6. [1] 0.3283606
   1. P(MPG<40)
7. pnorm(40,34,9)
8. [1] 0.7475075

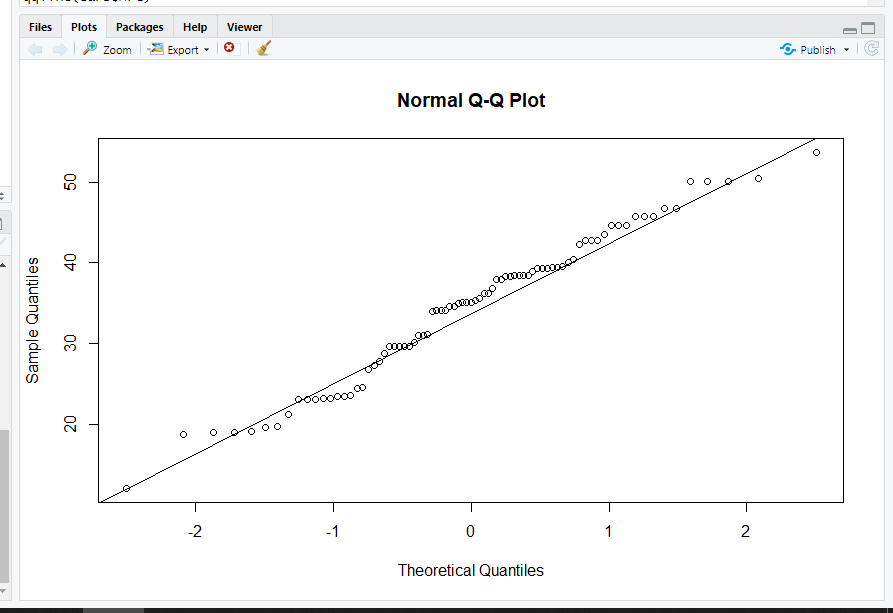
c. P (20<MPG<50)

|  |
| --- |
| pnorm(50,34,9)-pnorm(20,34,9)  [1] 0.9023729 |
|  |
| |  | | --- | |  | |

Q 21) Check whether the data follows normal distribution

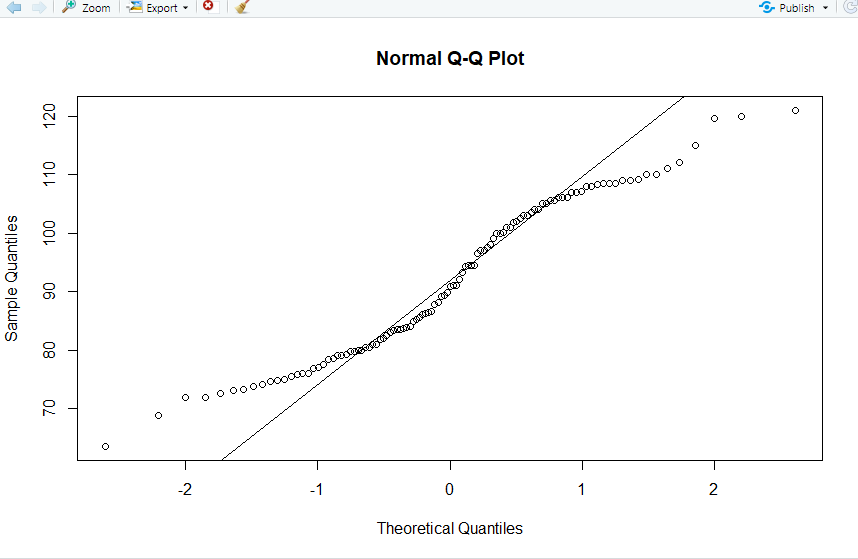
1. Check whether the MPG of Cars follows Normal Distribution

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



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

90% = 1.64

94% = 1.88

60% = 0.84

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

For 95% = 2.063899

For 96% = 2.171545

For 99%=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%