|  |  |
| --- | --- |
| 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, Interval |
| 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 , Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval, Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal & Interval |
| Years of Education | Interval/Ratio |

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

Total events = 2\*2\*2=8

Interested events (Possibilities of getting 2 heads & 1 tail) = HHT, HTH, THH=3

Probability=Interested events / Total no of events = 3/8=37.5%

**Ans: 3/8=37.5%**

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 = **Zero**
2. Less than or equal to 4 = **6/36=1/6**
3. Sum is divisible by 2 and 3 = **2/12=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?

Total no of balls = 7

Total possibilities drawing 2 balls out of 6 = 7C2=21

Possibilities getting no blue balls = 5C2=10

Probability none of the balls are blue=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

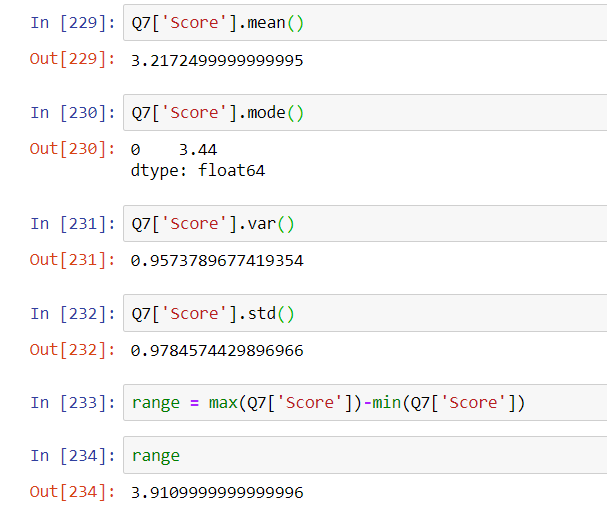
**Ans: 3.09**

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.

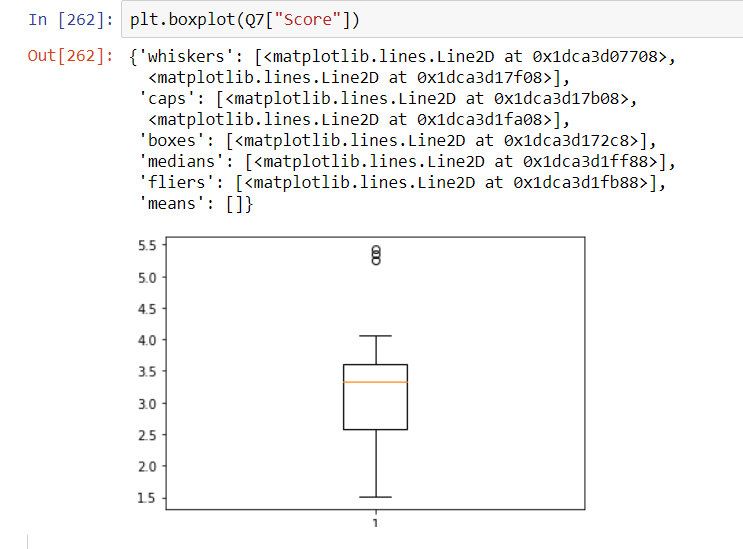
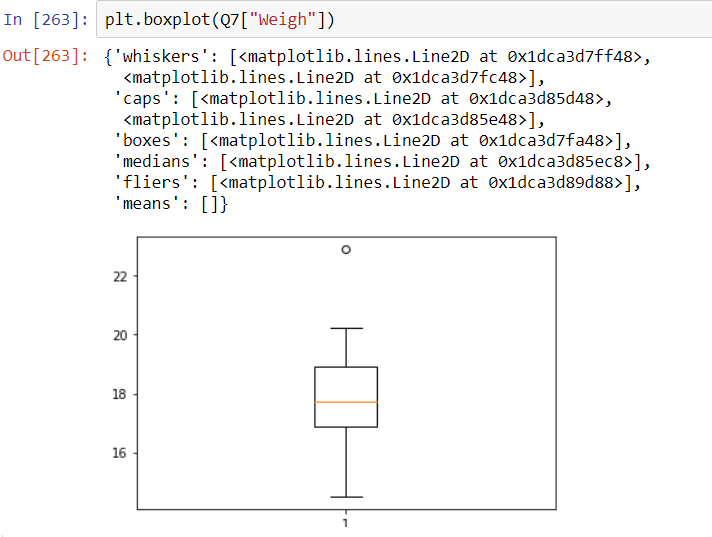


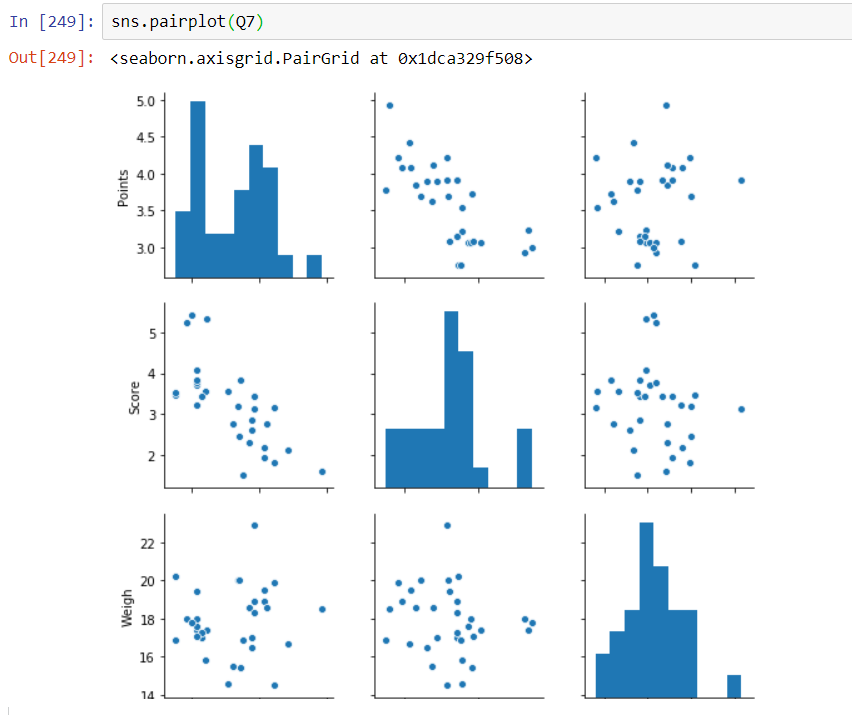




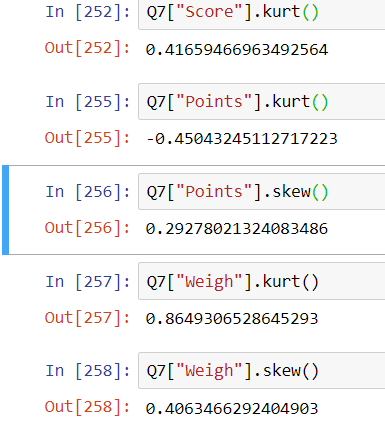
Outliers:

No outliers in Points but there are few outliers in Score (5 to 5.5) & Weigh (above 22) which can be interfered by Box plot as shown below.



Note: From above graph it is understood that positive kurtosis has narrow peak & negative kurtosis has wide range of peak.





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**

**Formula = X\*P(x)**

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

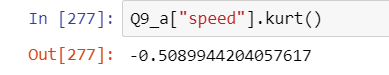
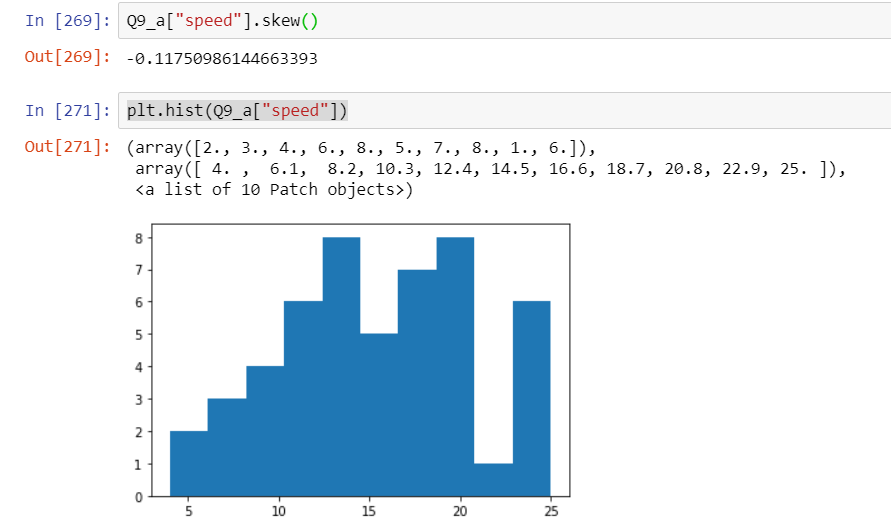
**Cars speed and distance**



**Speed:**

**Skewness:** Negative Skewed

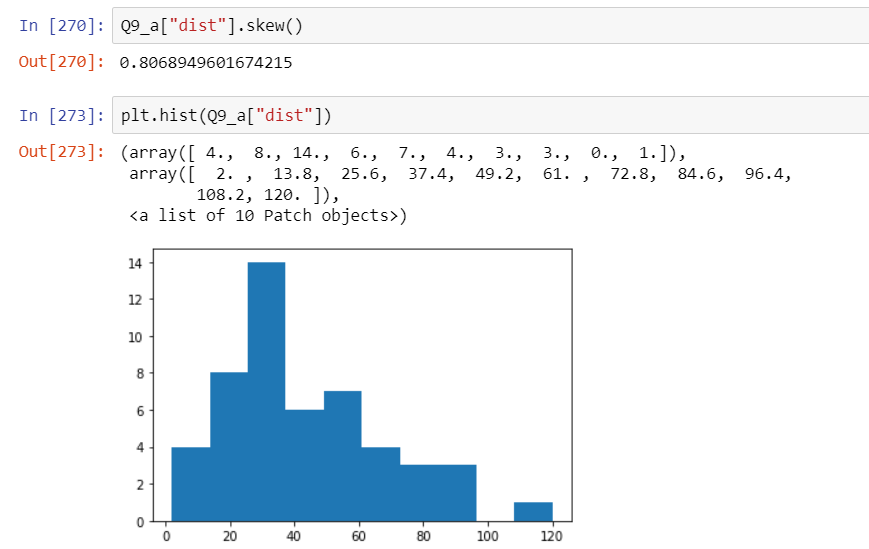
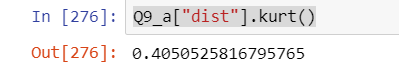
**Kurtosis:** Negative Kurtosis hence has wider range of peaks.

****

**Distance:**

**Skewness:** Positive skewed.

**Kurtosis:** Positive Kurtosis, hence has narrow peak.

**SP and Weight(WT)**



**SP:**

**Skewness:** Positive skewed

**Kurtosis:** Positive kurtosis hence narrow peak.

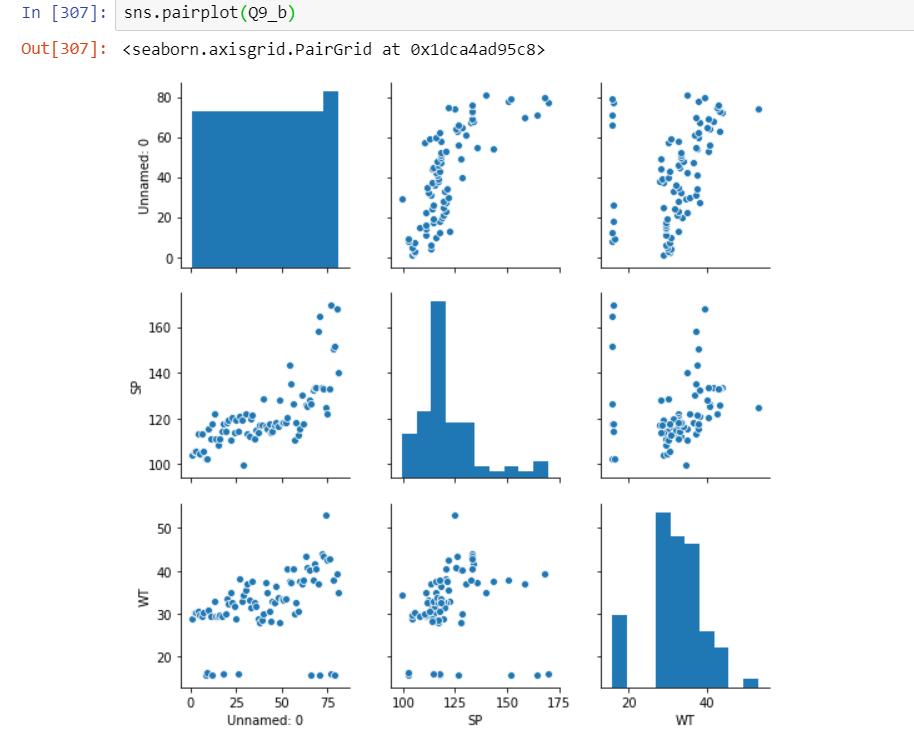
**WT:**

**Skewness:** Negative Skewed

**Kurtosis:** Negative Kurtosis hence has wider range of peaks.



Graphical representation by pair plot.



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



**Histogram:**

Above graph is Weight vs Frequency. As per graph as Weight increases, frequency decreases.

This is positive Skewed graph

Highest Frequency is between weight 50 to 100

From weight 300 to 400 comes under long tail

**Box plot:**

From box plot we can understand that there are 7 no of outliers on positive side.

**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:

Point estimate – 200

No of samples – n – 2000

Std deviation of sample – 30

Confidence interval = Point estimate ± Margin of error

Point estimate ± Z(1-α) \* (Std deviation/√n)

For 94%:

Z = 1.88 (*Using qnorm function in R)*

Upper limit = 201.26

Lower limit = 198.74

For 98%

Z = 2.33 (*Using qnorm function in R)*

Upper limit = 201.56

Lower limit = 198.43

For 96%

Z = 2.05 (*Using qnorm function in R)*

Upper limit = 201.37

Lower limit = 198.62

From above values we can infer that as confidence level increases range also increases.

**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.

***For Population***

**Mean:** Sum(Xi)/n

Sum(xi)=738, n=18

Hence Mean = 738/18=**41**

**Median:** middle most value = **40.5**

**Variance:** sum((X-mean)^2/N) = **24.11**

**Std Deviation:** Root of Variance = **4.910307**

***For Sample***

**Mean:** 41

**Median:** 40.5

**Variance:** 25.529411

**Std Deviation:** 5.05266

1. What can we say about the student marks?

They are multimodal. Positive skewed & positive kurtosis.

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

Ans: Zero skewed

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

Ans: Positive skewed

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

Ans: Negative skewed

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

Ans: Unimodal.

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

Ans: Multi modal

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



What can we say about the distribution of the data?

Ans: Median is 15, more data lies between 14 to 16.

What is nature of skewness of the data?

Ans: Negative skewed

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

Ans: IQR (Inter Quartile range) = 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.

Ans: Median of both box plot = 212.5

As Q1 & Q3 are different IQR for both box plots are different.

For Box plot 1:

Q1=250, Q3=275, IQR=25

For Box plot 2:

Q1=310, Q3=225, IQR=85

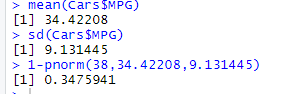
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

* 1. P(MPG>38)



* 1. P(MPG<40)

0.7289871

* 1. P (20<MPG<50)

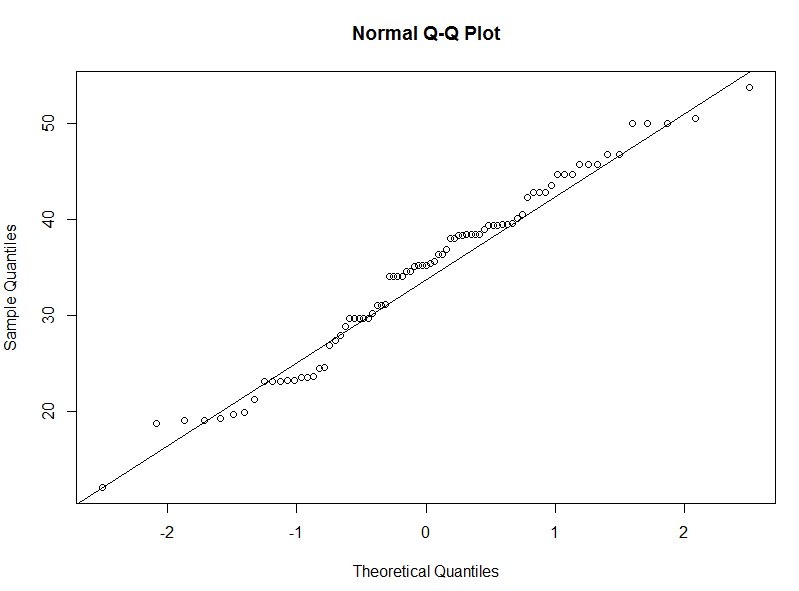
0.8975008

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

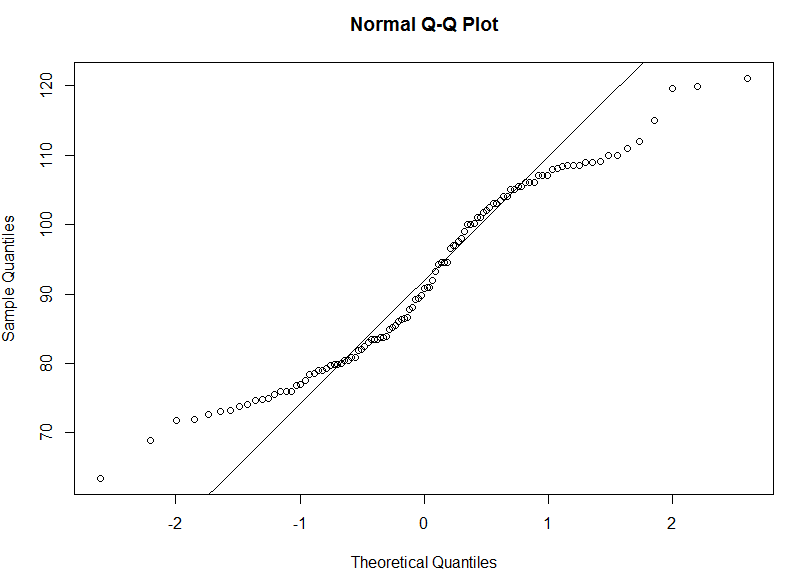
Yes. As per below graph data of Cars@MPG is normally distributed.



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

Dataset: wc-at.csv

Ans: From below Q-Q plot it is understood that Adipose Tissue (AT) and Waist Circumference(Waist) are not distributed normally.



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

Ans: Z scores

90% = 0.8289439

94% = 0.8339768

60% = 0.7881446

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

Ans: t scores

95% = 2.063899

96% = 2.171545

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:

t score = (Point estimate – Population estimate) /(s/√n)

Point estimate-260

Population estimate-270

S(std deviation of sample) – 90

n – 18

t score = (260-270)/(90/√18)

tscore = -0.4714

pt(-0.4714, 17)

Hence probability is **0.32167**