Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q2) Identify the Data types, which were among the following, Nominal, Ordinal, Interval, and 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 | Nominal (Ordinal) |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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 number of events is = 8(HHH/TTT/ HTH /THH/HHT/HTT/THT)

No. of interested events for 2 heads and 1 tail is = 3(HTH/THH/HHT)

Probability= No. of interested events/Total no of events

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

Ans. a) Equal to 1 -Ans: 0

b) Less than or equal to 4 - Ans:1/6

c)Sum is divisible by 2 and 3 – Ans: 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.: =p (2R, 3G, 2B)

=p (5/7, 4/6) = 20/42

=10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of the 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 – the probability of having 4 candies = 0.20

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

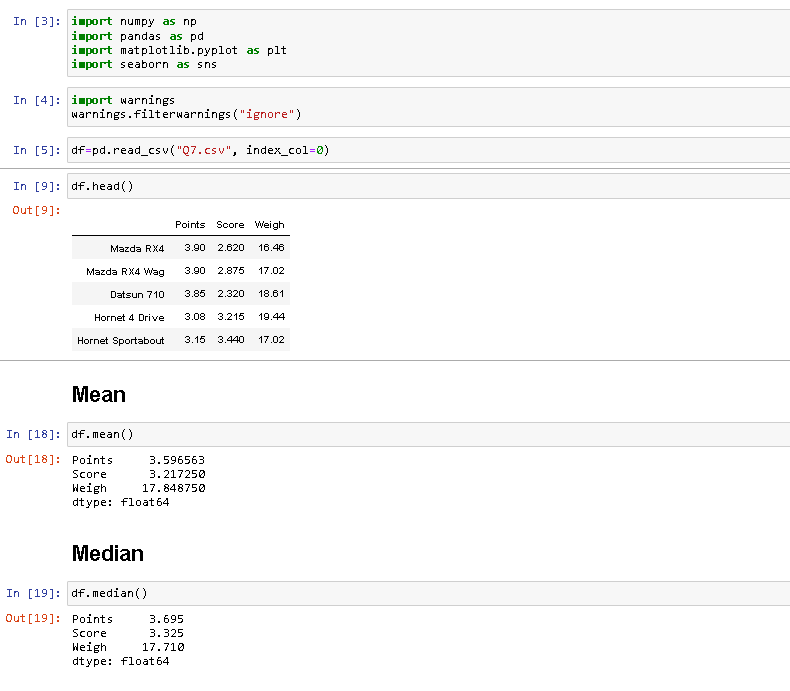
= 3.09

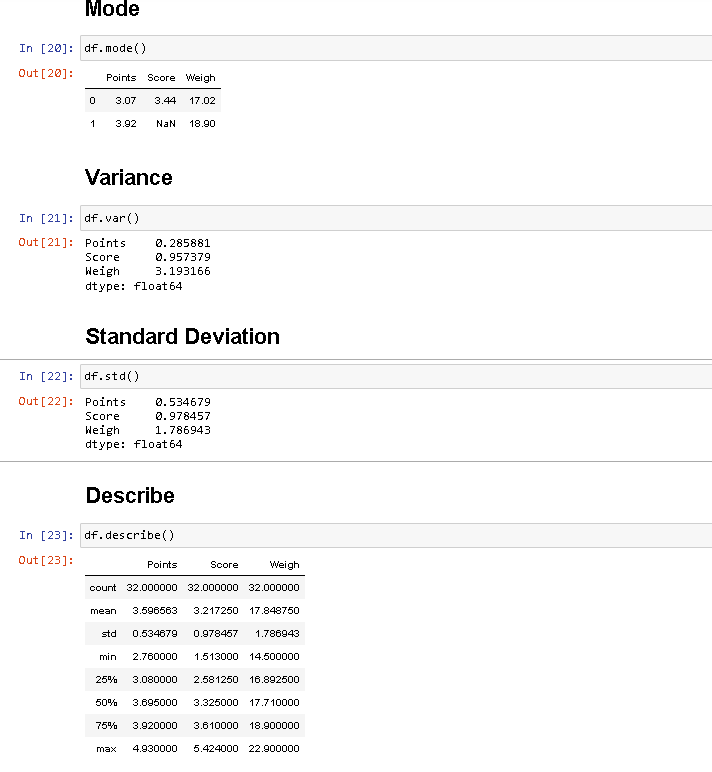
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

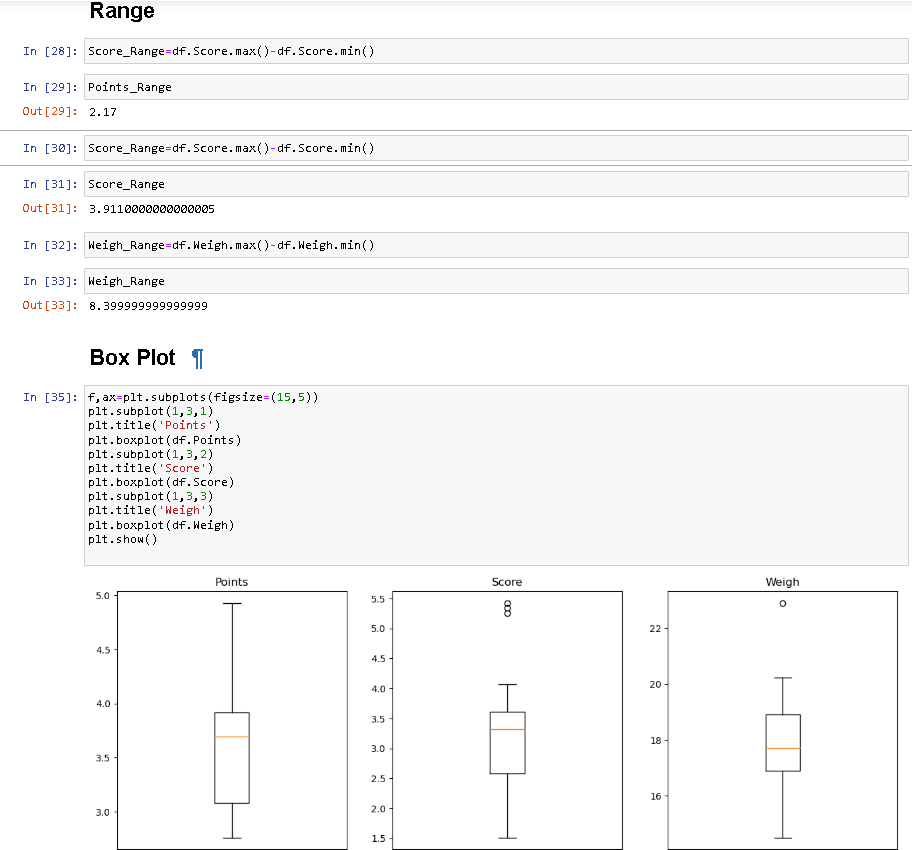
* For Points,Score,Weigh>

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

**Use Q7.csv file**



****

****

**Sol:** Points:

Mean=3.596563, Median=3.695, Mode=” numeric”,

Variance=0.2858814, Standard deviation=0.5346787.

Score:

Mean=3.21725, Median=3.325, Mode=” numeric”,

Variance=0.957379, Standard deviation=0.9784574.

Note: Mean values are closer for both ‘Point’ and ‘Score’.

Weight:

Mean=17.84875, Median=17.71, Mode=” numeric”, Variance=3.193166, Standard deviation=1.786943.

Q8) Calculate the Expected Value for the problem below

1. The weight
2. 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 = Sum of(Probability \* value)

There are 9 patients

The probability of one random patient = 1/9

Expected Value =

(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) \*1308

= 145.33

The Expected value of random = 145.34

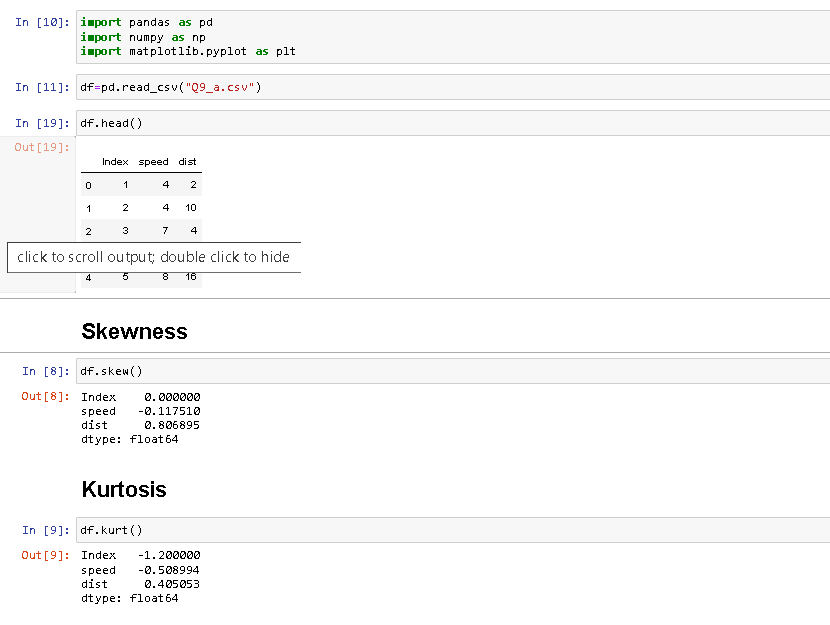
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

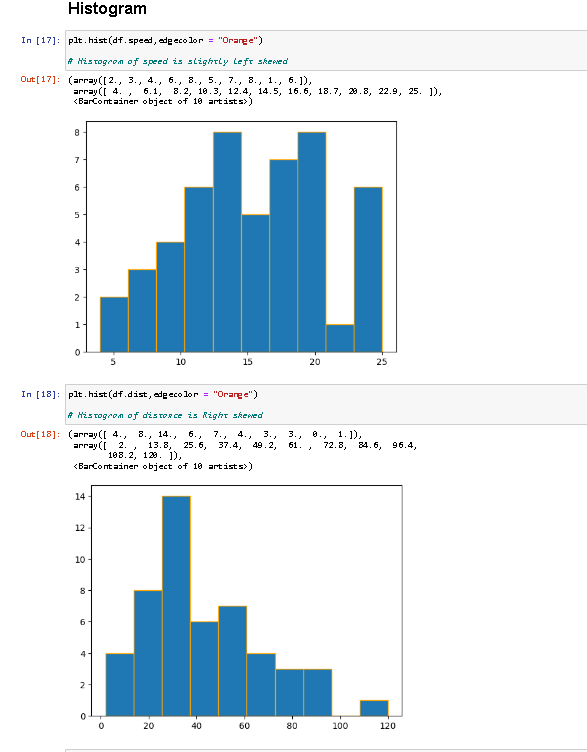
**Car’s speed and distance**

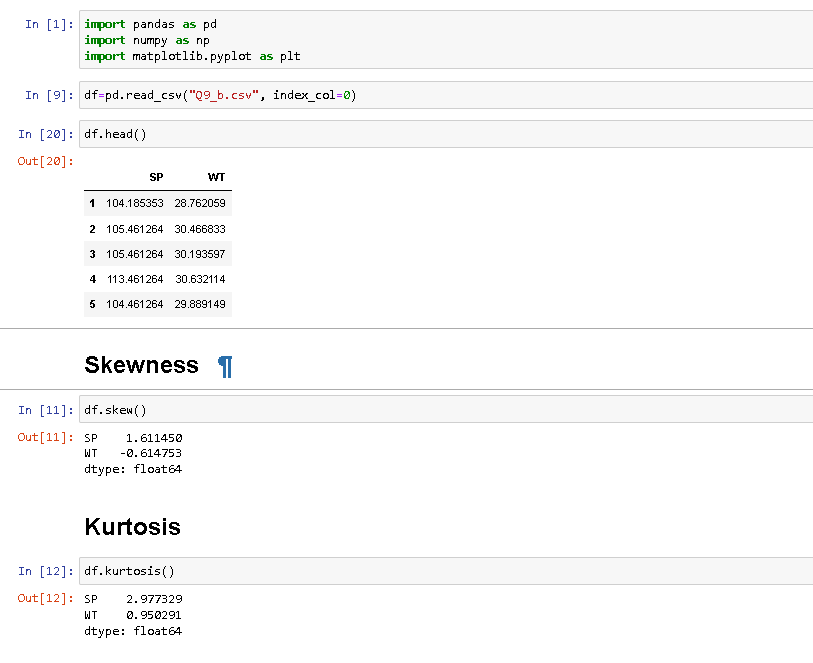
**Use Q9\_a.csv**

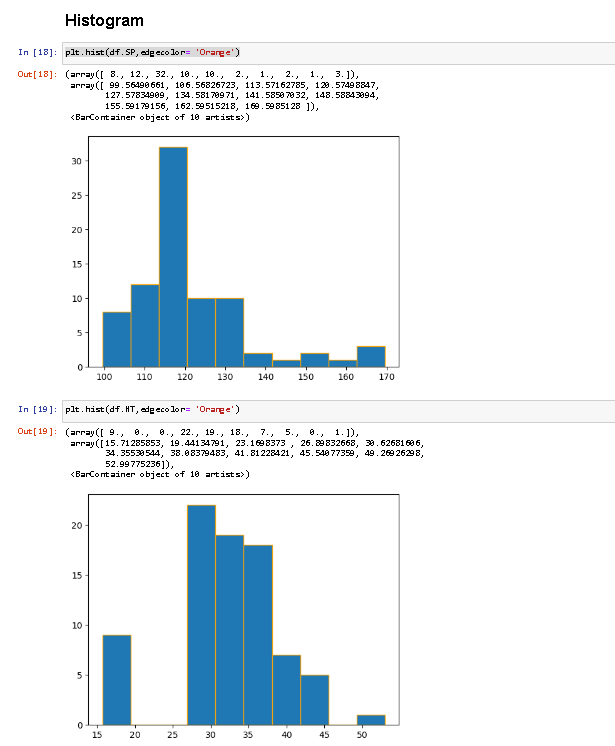
**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans 9(A)**

****

Ans 9(b)

****

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



**Sol:** The most of the data points are concerted in the range 50-100 with a frequency of 200. And the least range of weight is 400 somewhere around 0-10.

So, the expected value of the above distribution is 75.

Skewness- we can notice a long tail towards the right so it is heavily right skewed.

**Sol:** Medicant is less than mean right skewed and we have an outlier on the upper side of the box plot and there are fewer data points between Q1 and the bottom point.

**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%, and 96% confidence interval.?

**Sol:** X+/-(Z1-α.σ/sqrt (n)

Degrees of freedom=2000-1=1999

Confidence interval=94%

(1-σ/2) =1-0.03) =0.97

For confidence interval for 94% is 1.882

Confidence interval for 98%=2.33

Confidence interval for 96%=2.05.

**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): Mean=41, Median=40, variance=24.111, Standard deviation=4.910

2): There is no Outlier in the data

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

Ans.: Positively or negatively skewed distribution.

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

Ans.: Skewness and tail is towards the right

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

Ans.: Skewness and tail is towards left.

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

Ans.: Positive kurtosis means the curve is more peaked and it is Leptokurtic

.

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

Ans.: Lighter tails and a flatter peak

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



What can we say about the distribution of the data?

**Sol:** The above Boxplot is not normally distributed the median is towards the higher value

The above box plot is about the ages of the students in a school.

50% of the people are above 10 years old and the remaining are less.

And students whose age is above 15 are approx. 40%.

What is the nature of the skewness of the data?

**Sol:** Left skewed, the median is greater than the mean.

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

**Sol:** The Inter Quantile Range = Q3(Upper quartile) – Q1 (Lower Quartile) = 18 – 10 =8  
Approximately=-8

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect to Boxplot 2.

**Sol:** By observing both the plots whisker’s level is high in boxplot2, mean and median are equal hence distribution is symmetrical.

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: P(MPG>38)**

= mean (MPG)=34.4220

= Sd (MPG)=9.131445

= 1- pnorm(38, mean(MPG),sd(MPG)

= 0.330

=33%

**P(MPG<40)**

=pnorm (40, mean (MPG), Sd (MPG))

=0.7293499

=72.3%

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

=pnorm(50, mean(MPG))-pnorm(20,mean(MPG),sd(MPG))

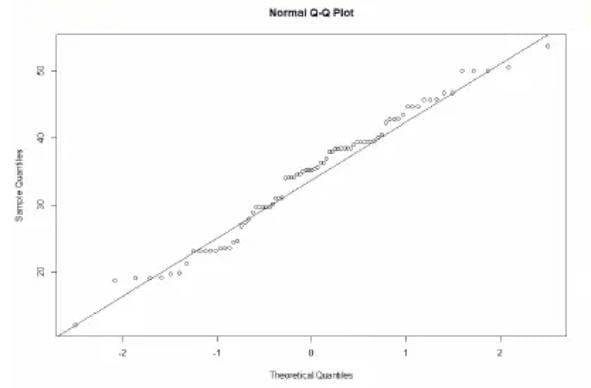
=0.955-0.057

=0.8988689

Q 21) Check whether the data follows a normal distribution

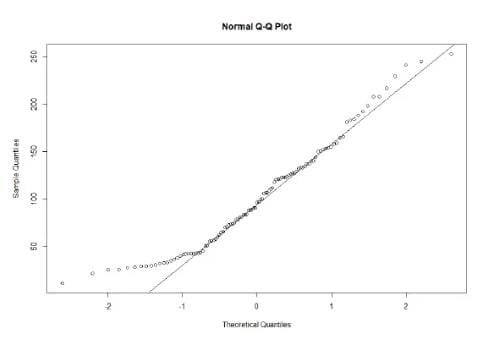
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

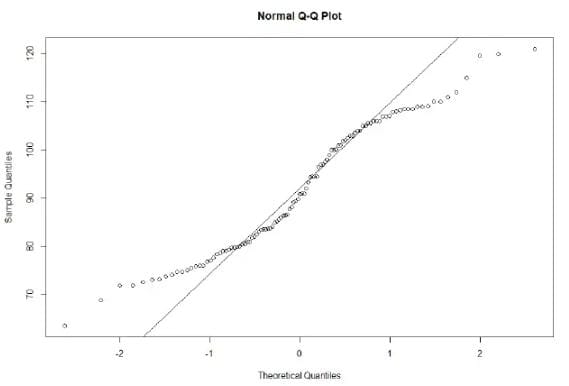


When we plot to check the qqnorm and qqline we can almost get a straight line thus the data is normalized. MPG of cars follows Normal Distribution.

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

Dataset: wc-at.csv

Majority of the data points lie on the qqline hence Normal Distribution.



This data set is not normal because the data points follow an abnormal curve.

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

**Sol:** Z score

**=90%**

=951+2.5

=97.5

=qnorm (0.975)

=1.96

**94%**

=94+4

=97

=qnorm (0.97)

=1.88

**60%**

=60+20

=80

=qnorm (0.80)

=0.841

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

**Sol:** **TSCORE CALCULATION**

T ((1, alpha), (n-1))

Here n = 25

n-1 =24

Hence t score values will be:

**95%**

=qt (0.975,24)

=2.063899

**96%**

=qt (0.98,24)

=2.179694

**99%**

=qt (0.995,24) =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

**Sol:** Sample size = 18 = n

Sample mean = 260 days = x

Sample standard deviation = s = 90days

= 260-270/90/SQRT (18) =-10/9.487=-1.054