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 | Discrete |
| 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, 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) | 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

A) The total number of possible outcomes when tossing three coins are {HHH, TTT, HHT,HTH,THH,TTH,THT,HTT}

Therefore the total number of outcomes =8

Number with two heads and one tail =3

Probability of two heads and one tail =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 2and 3
4. The total possible outcomes when two dice are rolled are 36

The maximum sum is (1,1)=2

Therefore p(1)=0/36=0

1. The total outcomes are 36

The number of combinations for which sum is less than or equal to 4 are {(1,1),(1.2),(1,3),(2,1),(2,2),(3,1)} i.e 6

Therefore p(sum<=4)=6/36=0.1666

1. The total possible outcomes when two dice are rolled are 36

The sum is divisible by 2 and 3 are {(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),(4,6),(5,1),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4)}

Therefore p(sum is divisible by 2 and 3)=23/36=0.6388

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?

1. Total number of balls=(2+3+2)=7

Let S be the sample space.

Then n(S) =number of ways of drawing 2 balls out of 7=7 C2=21

Let E =Event of drawing 2 balls, none of which is blue.

n (E)=number of ways of drawing 2 balls out of 5=5 C2=10

p(E)=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. Expected number of candies for a randomly selected child=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

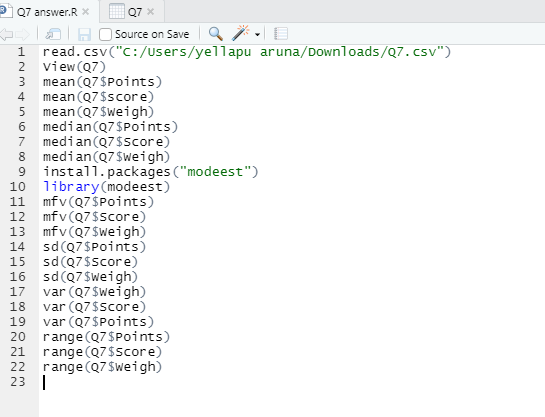
=3.09

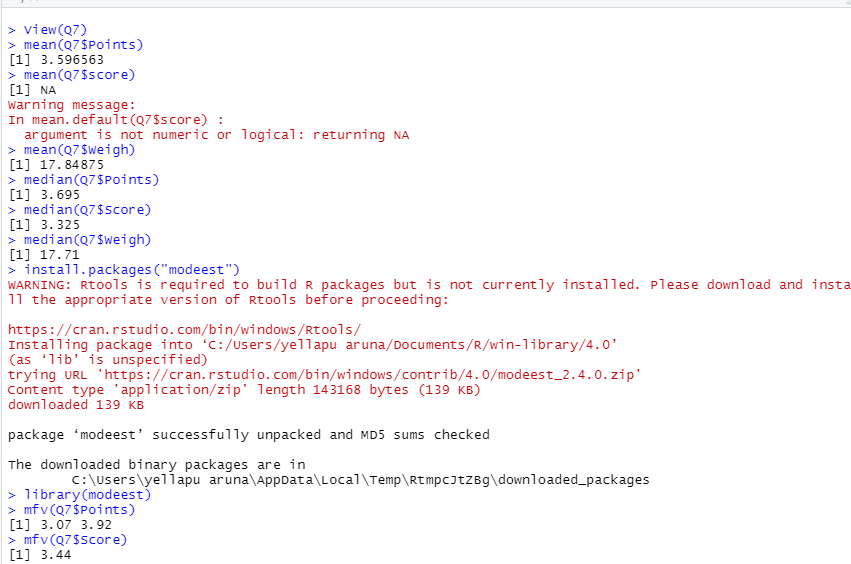
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

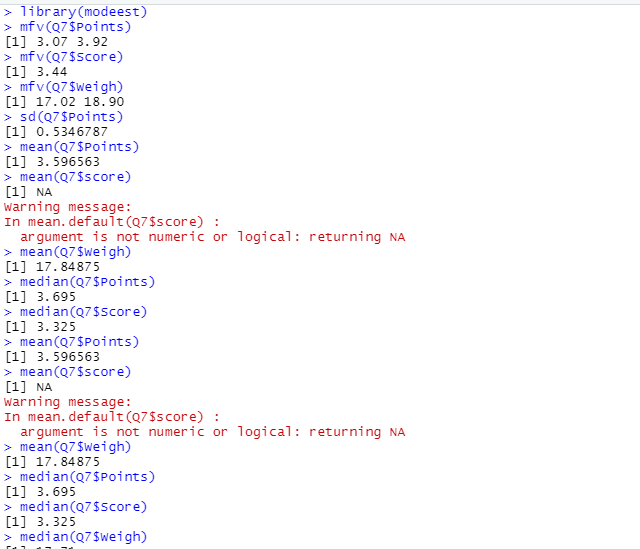
* For Points, Score, Weight

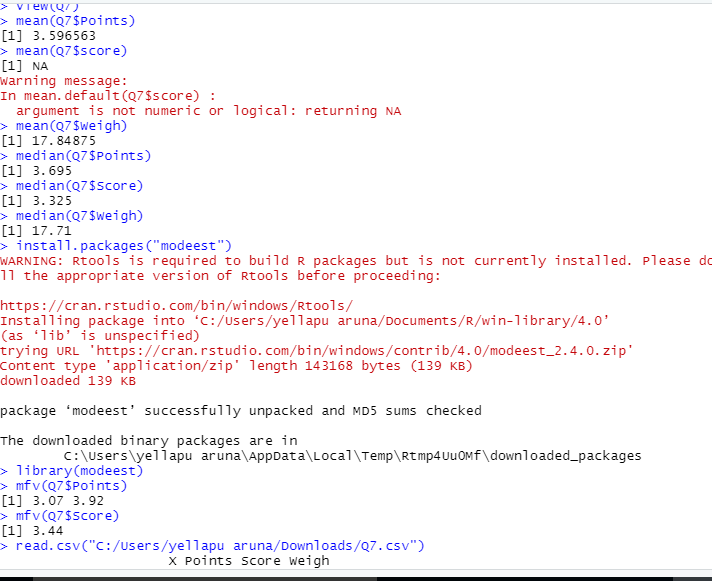
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**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?

1. expected value=∑ P(x).E(X)

Probability of selecting each patient is 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

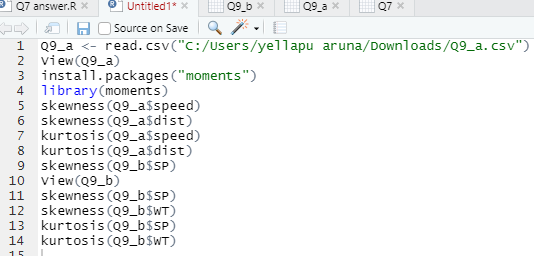
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

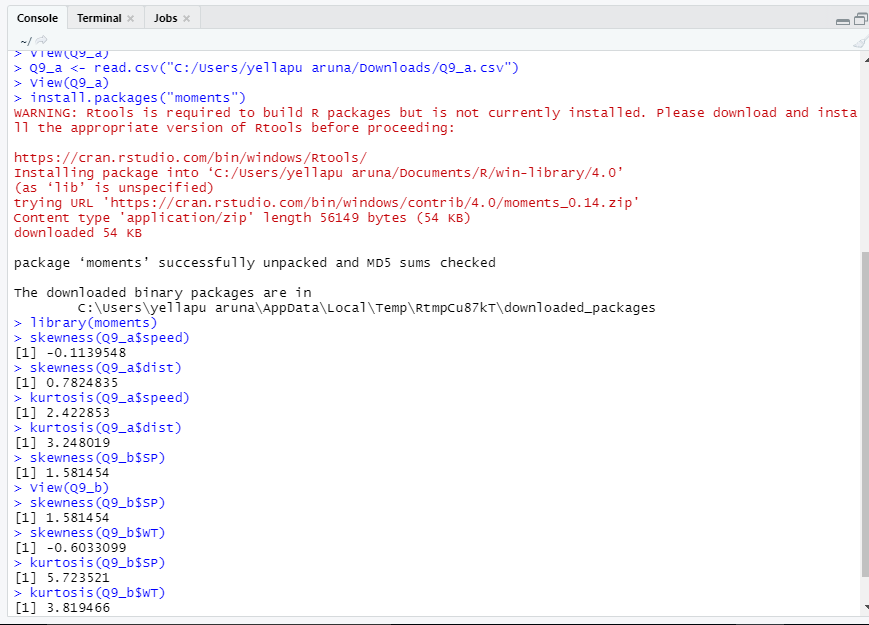
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

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



1. Above histogram plot is positive skewed so from that we came to know there are some outliers. The positive skewness here tells us asymmetry. It is not following normal distribution because it is not symmetrical about mean. Mean ,median and mode of histogram are not equal.



1. From above box plot we can tell that there are multiple outliers present. from above box plot we can tell that it is not following normal distribution

**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?

1. Given sample=200

Standard error =s/√n

30/√200=0.67082

α =100-confidence interval

=0.06, 0.02, 0.04

Critical probability=1-α/2=0.97,0.99 and 0.98

Degrees of freedom=2000-1=1999

From t table critical values are 1.882, 2.328 &2.055

Margin of error (98%)=critical value of 94%\*standard error

=2.328\*0.67082=1.561

Margin of error (94%)=critical value of 98%\*standard error

=1.882\*0.67082=1.262

Margin of error(96%)=critical value of 96%\*standard error

2.055\*0.67082=1.378

Confidence interval=sample statistics± margin of error

CI(94%)=200±1.262

198.738 to 201.262

CI(98%)=200±1.561

198.439 to 201.561

CI(96%)=200±1.378

198.622 to 201.378

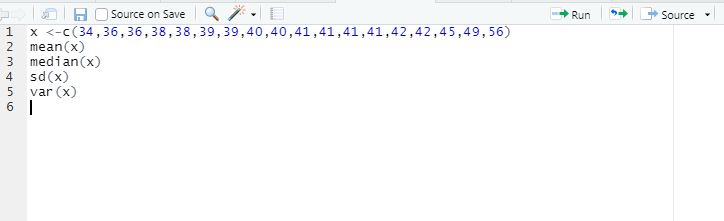
**Q12)** Below are the scores obtained by a student in tests

**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. Mean=(39+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+

56)/18

41

Median=(40+41)/2=40.5

1. What can we say about the student marks?
2. According to the above data students who have scored in between 34-39 are 7 members, students who scored in between 40-45 are 9 members, student who scored between 46-51 is 1 and student who scored in between 52-57 is 1 only.

The skewness and kurtosis of the above data are 1.686 and 3.953 both are positive only.

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

1. when mean =median then the distribution is symmetric and the distribution has zero skewness.

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

1. whenever the mean of data is greater than median of data then the nature of skewness is positively skewed.

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

1. Negatively skewed

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

1. Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

1. A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

Q18) Answer the below questions using the below box plot visualization.



What can we say about the distribution of the data?

1. In above box plot visualization mean is 10 and median is lies between 14 and 16 , mean is less than the median. Hence it is negatively skewed data. Here median is closer to top quartile.

What is nature of skewness of the data?

1. (Q3-Q2)<(Q2-Q1) negative skewed data

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

1. IQR=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.

1. The first box plot is comparatively short ,In both the plots mean=median and the above two box plots overlapping with each other.

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

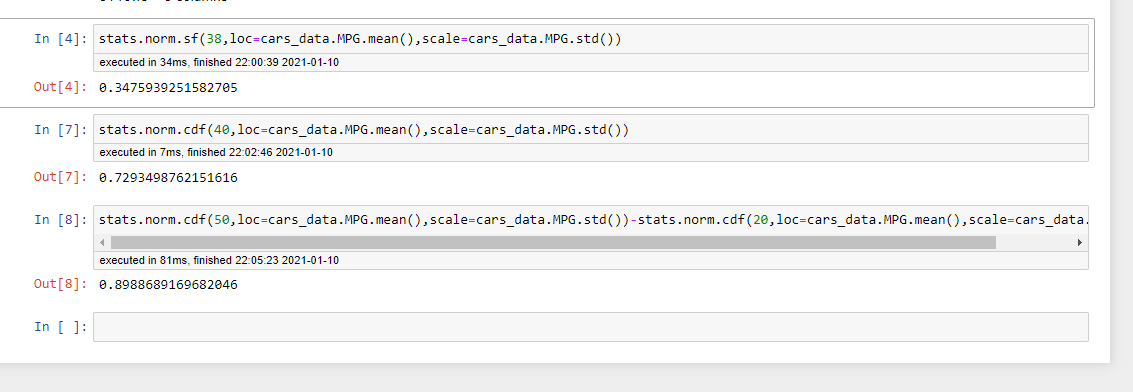
Data \_set: Cars.csv

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

MPG<- Cars$MPG

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

c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

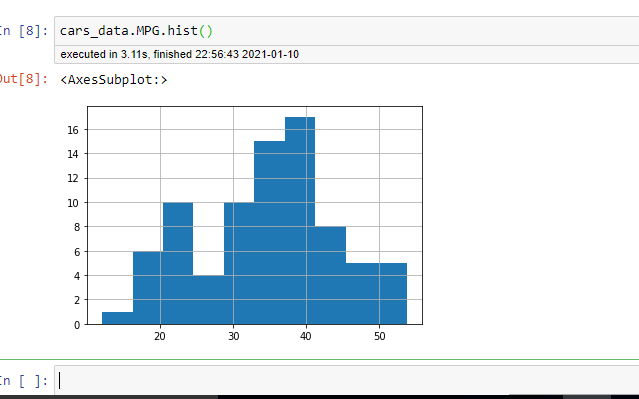
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Here the mean, median and mode are not equal so it is not following normal distribution.

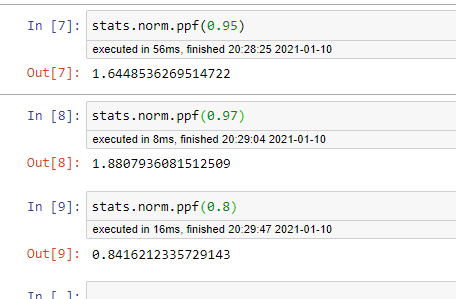
From below histogram it is not symmetric about the mean so it is not following normal distribution.



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

Dataset: wc-at.csv

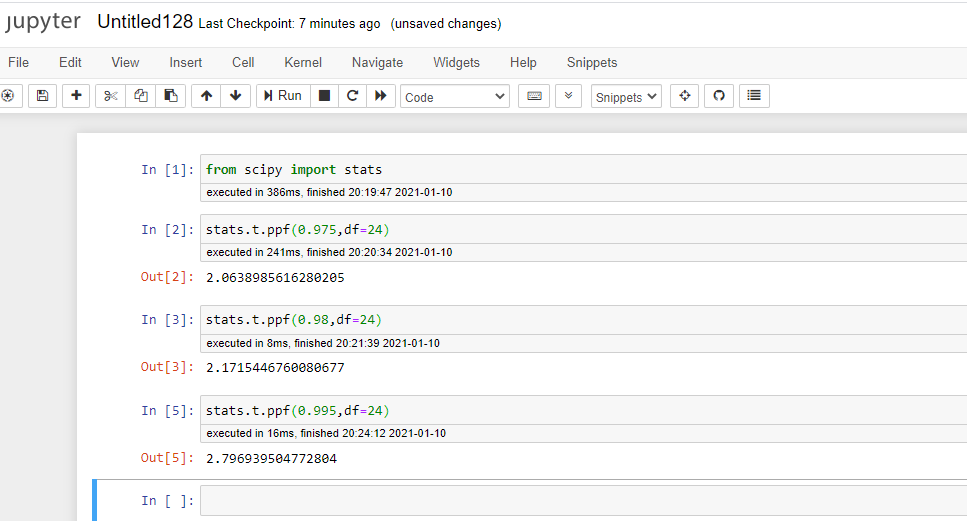
Q 22) Calculate the Z scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

A) 

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

A) Given sample size(n)=25

Degrees of freedom=n-1=24



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

1. Given n=18

S=90

Sample mean= 260

T score =(x-µ)/(s/√n)

T score=0.47

Degrees of freedom=n-1=17

