|  |  |
| --- | --- |
| 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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| 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 possible outcomes are 8 - HHH, TTT, HTT, HHT, TTH, THH, THT, HTH

Possible outcomes getting two heads and one tail are 3 – HHT, THH, HTH

So, Probability is = 3/8

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

1. Equal to 1

Ans: No situation to get the sum is 1, because minimum value is 1+1=2

Probability = 0

1. Less than or equal to 4

Ans: Total number of possibilities are 36

Possible outcomes equal to 4 are

(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

Probability of sum is less than or equal to 4 = 6/36 = 1/6

1. Sum is divisible by 2 and 3

Ans : Total number of possibilities are 36

Possible outcomes sum is divisible by 2 and 3

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

Probability of sum is divisible by 2 and 3 = 6/36 = 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:

Red balls = 2

Green balls = 3

Blue balls =2

Total number of balls are 7

Probability of none of the balls drawn is blue is = (2R+3G)/7

=5C2/7C2

=(5\*4/2)/(7\*6/2)

=20/42

=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: Sum of Pi \* Xi

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

= 0.015+0.8+1.95+0.025+0.06+0.240

=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:

Mean : Mean is average value of the data

Median : Median is middle value in the data, If the n is odd. Average of two middle values if n is even.

Mode : Mode is most occurring value in the data.

Variance : Variance is variety of information of the data. Variance is squared of standard deviation.

Standard deviation : Standard deviation is distance or deviation from the mean. Standard deviation is square root of variance.

Range : Range is the difference between maximum and minimum value of the data.

Please check in Jupiter note book.

For Points and Score Mean < Median

For Weigh Mean > Median

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

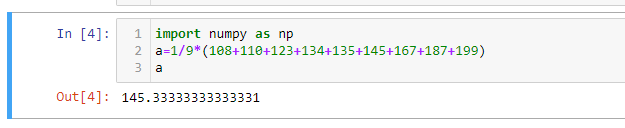
1. The weights (X) of patients at a clinic (in pounds), are

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?

Ans: Total patients are 9

Expected value of randomly selected patient weight is =



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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skewness :** Skewness is a measure of asymmetry of distribution of data**.**

**Kurtosis :** Kurtosis is a measure of peakedness of distribution of data.

install.packages("moments")

library (moments)

skewness(Q9\_a$speed)

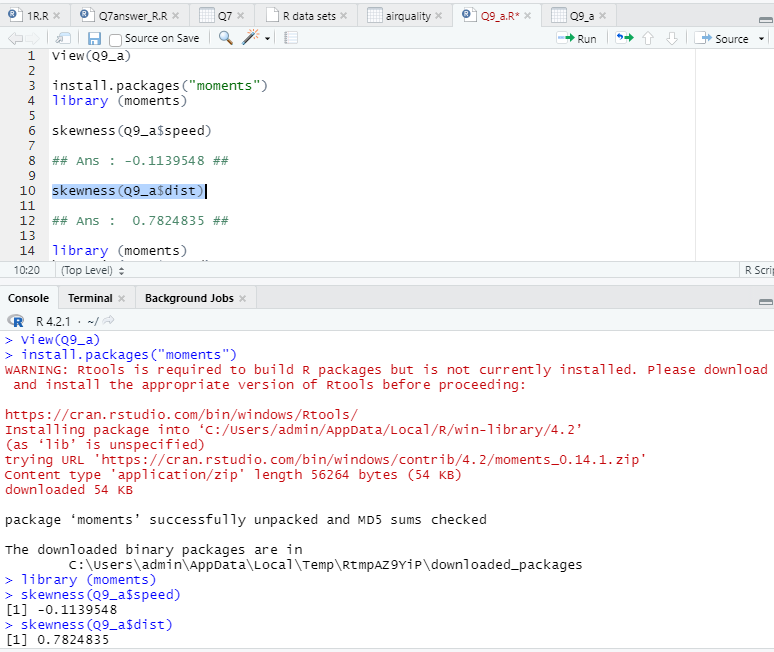
Ans : -0.1139548

Skewness of car speed is negative, left skewed data and mean is less than the median.

skewness(Q9\_a$dist)

Ans : 0.7824835

Skewness of car distance is positive, right skewed data and mean is greater than the median.



library (moments)

kurtosis(Q9\_a$speed)

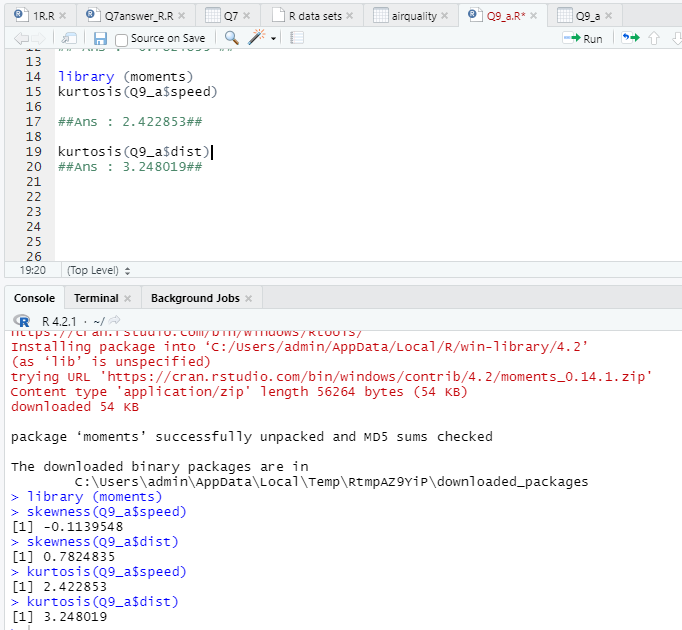
Ans : 2.422

Kurtosis for speed is positive and less than 3.

kurtosis(Q9\_a$dist)

Ans : 3.248

Kurtosis for distance is positive and more than 3, it is having excess or more kurtosis. So will get sharp peak and heavier tail than the normal distribution.



**SP and Weight(WT)**

skewness(Q9\_b$SP)

Ans : 1.581

Skewness of SP is positive, right skewed data and mean is greater than the median.

skewness(Q9\_b$WT)

Ans : -0.603

Skewness of WT is negative, left skewed data and mean is less than the median.

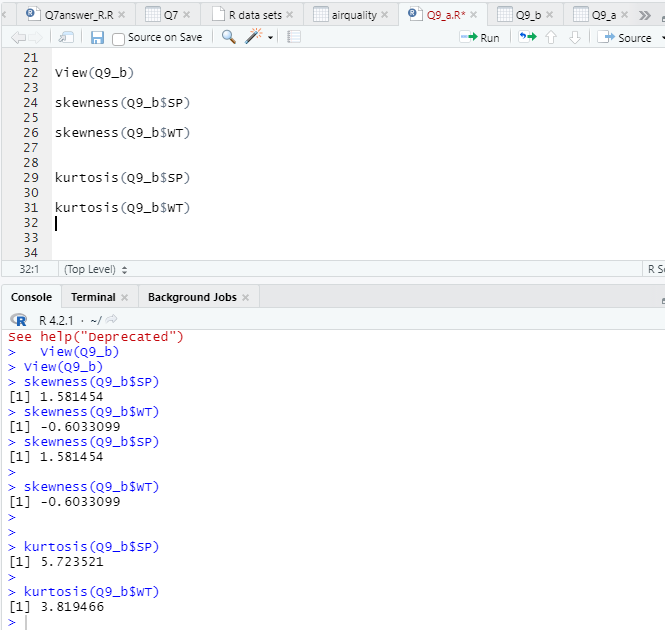
kurtosis(Q9\_b$SP)

Ans : 5.723

kurtosis(Q9\_b$WT)

Ans : 3.819

Kurtosis for SP and WT data is positive and more than 3, so both are having is excess or more kurtosis. So will get sharp peak and heavier tail than the normal distribution.



**Use Q9\_b.csv**

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



Ans:

1. Histogram is right skewed, most of the observations are present on left side and very few observations are present on right side.
2. Skewness is positive.
3. Mean is greater than Median



Ans : Box plot is a visualization used to detect outliers.

Box plot also provides information about the following

Q1= First Quartile

Q2= Median

Q3= Third Quartile

IQR= Inter Quartile Region = Q3-Q1

Lower Limit = LP= Q1-1.5(IQR)

Upper Limit = UP= Q3+1.5(IQR)

Box plot divides data into 4 regions. Every region contains 25% of data. Any observation which is lying beyond the box plot is called as outliers.

In the given box plot we can see the outliers are present above the upper limit.

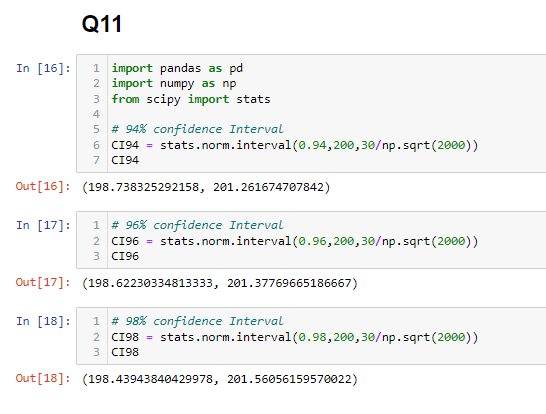
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

A:

Average sample weighs = 200 pounds

Standard deviation of sample is 30 pounds

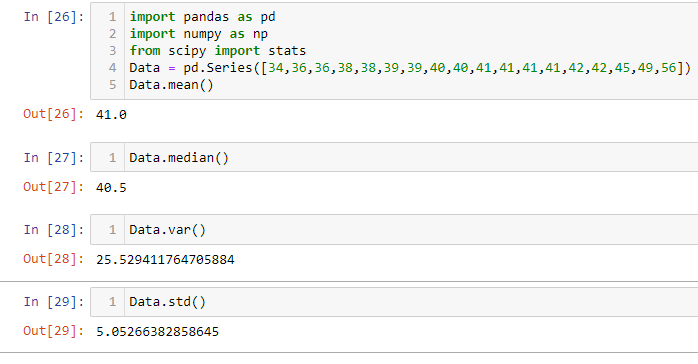
Sample size = 3000



**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.



1. What can we say about the student marks?

Ans : We can see student marks are slightly right skewed, we can see outliers present in boxplot.

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

Ans :

1. When mean, median of data are equal, then the data is perfectly symmetric.
2. 2. Skewness for the data is equal to Zero.

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

Ans: 1. When mean is greater than the median, the data is right skewed.

1. Skewness for the data is Positive.

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

Ans:

1. When median is greater than the mean, the data is left skewed.

1. Skewness for the data is negative.

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

Ans : Positive kurtosis value indicates the distribution of data has sharp peak and thin tails.

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

Ans : Negative kurtosis value indicates the distribution of data has flat peak and fat tails.

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



What can we say about the distribution of the data?

Ans : The distribution of the data is left skewed, skewness is negative. Mean is less than the median

What is nature of skewness of the data?

Ans: Data is left skewed and skewness of the data is negative.

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

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.

Ans :

1. Box plot 1 is shorter than the box plot 2. IQR and whiskers are smaller for boxplot 1 than box plot 2.
2. We can see median(Q2) is same for both the box plots 1 and 2.
3. We could not see any outliers for both the box plots.

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)

Ans : please check in Jupiter notebook

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

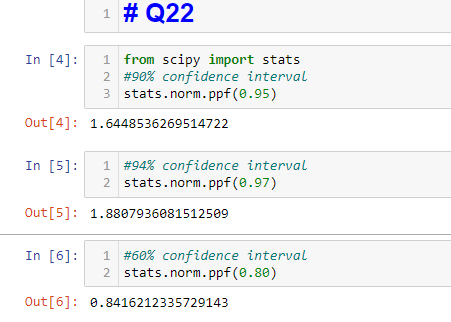
Ans : please check in Jupiter notebook

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

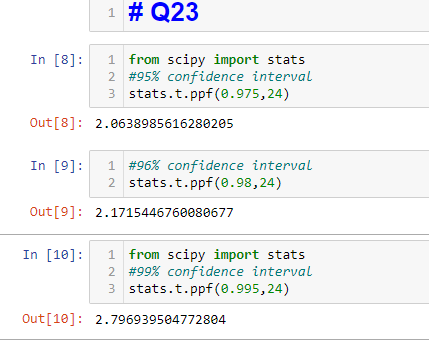
Dataset: wc-at.csv

Ans : please check in Jupiter notebook

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



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



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 : please check in Jupiter notebook