|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continues |
| Weight of Gold | Continues |
| Distance between two places | Continues |
| Length of a leaf | Continues |
| Dog's weight | Continues |
| Blue Color | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

ANS: Sample Space = {TTT,HTT,THT,THH,HHH,TTH,HTH,HHT} = 8

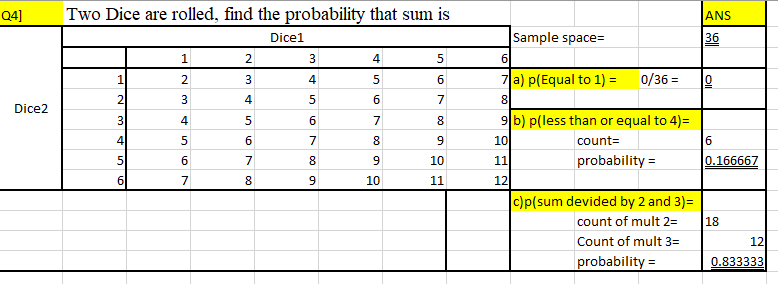
P(2Heads&1Tail) = {THH,HTH,HHT} = 3

= 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



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 : There are total 7 balls out of which 2 non blue balls are drawn.

Here we will find the probability for 2 balls are drawn from the 5 balls(2 red + 3green).

P(2 non blue balls are drawn) = 5C 2  / 7C 2  (By using permutation and combination)

Probability of non blue balls are drawn is 0.47619

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

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

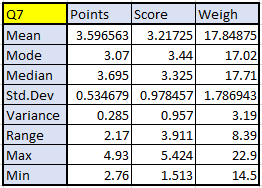
* For Points,Score,Weigh>

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

ANS : *Kindly refer the Assignment\_1 python code file attached for the detail answer.*

Also kindly check the answers below.

**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 : If any one patient choose randomly from the group of patients.

Then the expected value of the weight of that person is:

*145.33 would be the expected weight of that person.*

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

**Cars speed and distance Use Q9\_a.csv**

**ANS : *kindly find the below Screen shot for answers and also refer the attached python cod file.­­***

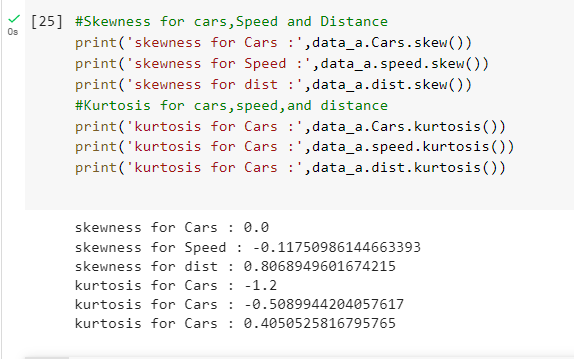
**For cars as you can see the skewness is zero hence there is no deviation for cars from its mean value.**

**For Speed skewness is -0.117 which is not that much deviated from the mean then we can conclude that speed across the all cars is not highly different.**

**For distance skewness is 0.806 which is highly skewed and highly deviated from the mean**

**#kurtosis**

**All the three value is highly skewed and highly deviated fron the mean value of the respective column data**

****

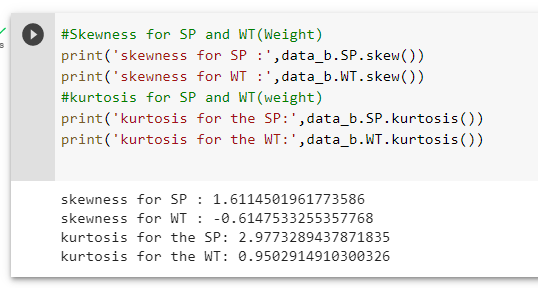
**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS : Skewness for SP is highly positive skewed which is highly deviated or partial to the data**

**Skewness for WT is negatively skewed near 0.5 which we can say is highly deviate from the mean of the data.**

**For the kurtosis both the values are very high which indicates the deviation of data from the mean.**

****

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



ANS :

* Above histogram data which is highly right skewed we also called positvely skewd data.where mean > median.
* We have outliers on greater side.
* Data is partial to the right hand side.
* By observing above data we ca say that as the weight is icreases the frequency got decreases.
* Data is higly deviated from the mean of the data.

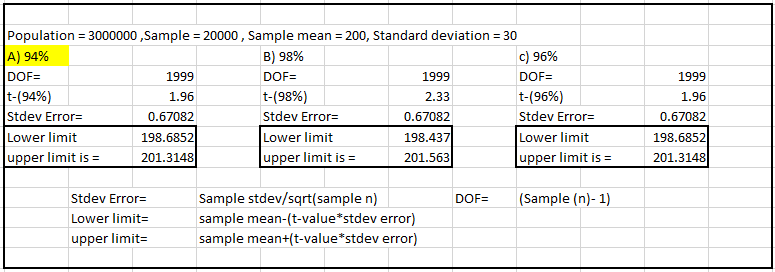


ANS:

* By observing above box plot we can say that the outliers are above the maximum 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 : Population = 3000000 ,Sample = 20000 , Sample mean = 200, Standard deviation = 30

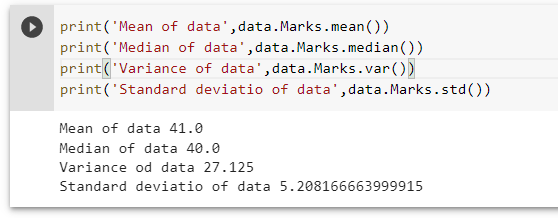


**Q12)** Below are the scores obtained by a student in tests

**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 the Answers kindly check the attached python code file



1. What can we say about the student marks?

ANS : As per above figure we can say that data is right skew data where mean> median.

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

ANS : There is no any skewness data symmetrical to the mean.

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

ANS : Data is right skewed (positively skew) when median less than mean

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

ANS : Data is left skewed(Negatively skewed ) when Median is greater than mean

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

ANS : Positive kurtosis also known as leptokurtic it highly peaked and possess thick tail. In positive kurtosis more numbers are located in the tail.

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

ANS : Negative kurtosis also known as platykurtic it has lighter tails low peaked

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



What can we say about the distribution of the data?

ANS: Data is not normally distributed median of data is = 15.6 (Approximately) towards the high value.

What is nature of skewness of the data?

ANS : The skewness of data is left skew as the whisker is extending towards the negative number hence we can also say that is negatively skewed data.

What will be the IQR of the data (approximately)?   
ANS : IQR(Inter quartile range) = Quartile 3 – Quartile 1

From Figure = 18 -10

IQR = 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 the both Box plot is approximately same which is 262.5
* Both the box plots are normally distributed. There is no any skewness neither at minimum side nor maximum
* Q1 for 1st box plot is 255(Approx) and Q1 for 2nd box plot is 225(Approx)
* Q3 for 1st box plot is 280(Approx) and Q3 for 2nd box plot is 305(Approx)
* IQR for 1st box plot is (280-255) (approx.)
* IQR for 2nd box plot is (305-225) = 80 (approx.)
* There is no any outliers for the both box plots
* Whiskers are equally extended.

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

Data \_set: Cars.csv

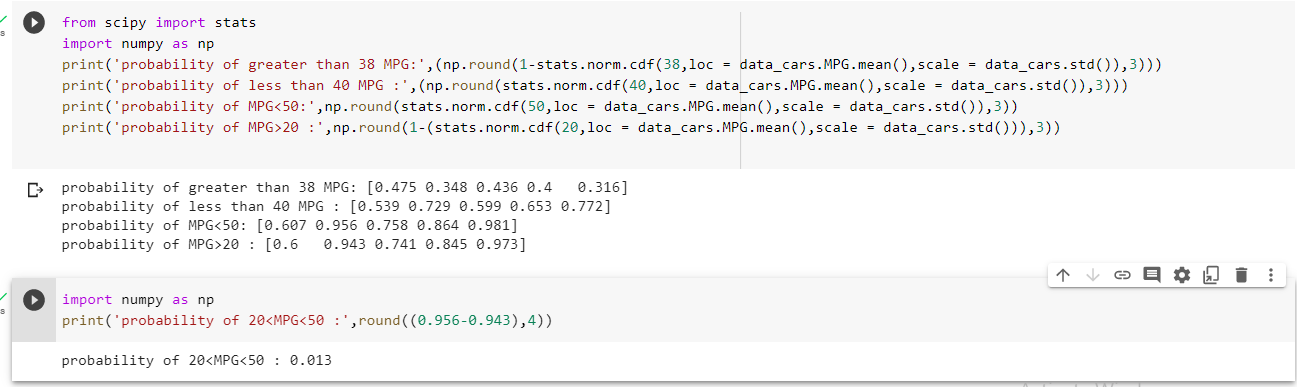
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

ANS : Kindly check the attached python code file for detailed ans.

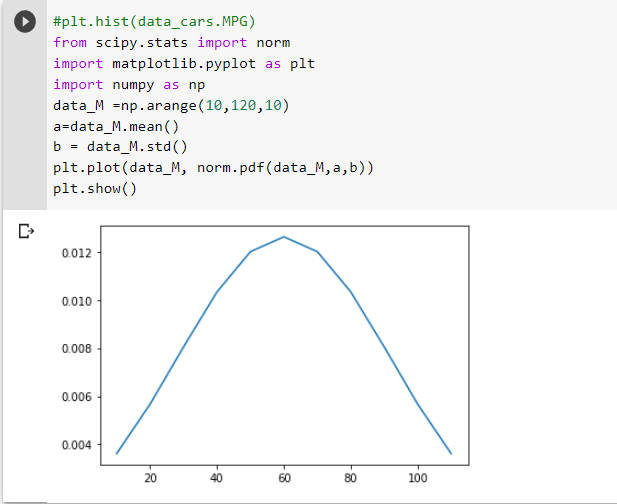


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

ANS : Acording to observation of histogram of the MPG data we can say that the data is normally distributed.

But the median is slightly greater than the mean



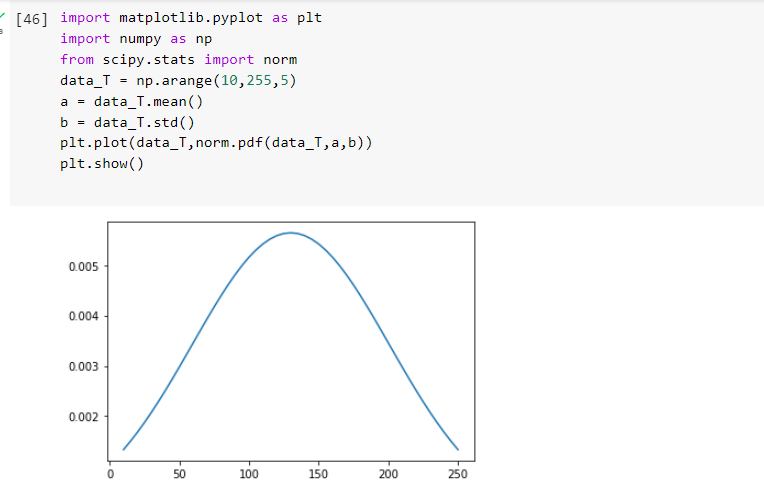
Dataset: Cars.csv

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

Dataset: wc-at.csv

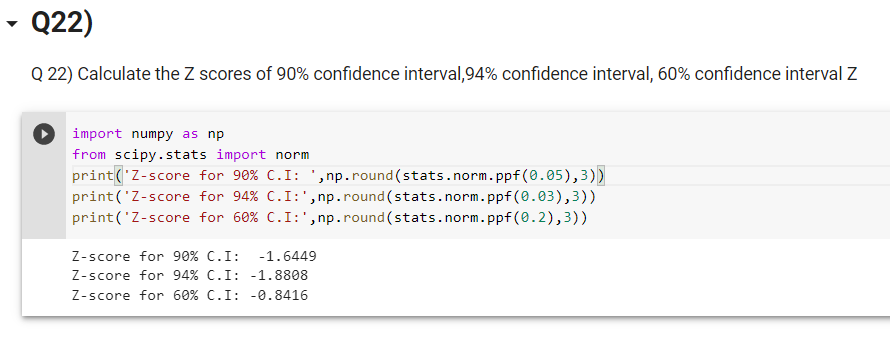
ANS : According to th observation we can say that the WC-AT data is not normally distributed as the data is slightly right skewed.

Kindly refer the python code file for more detailed answer



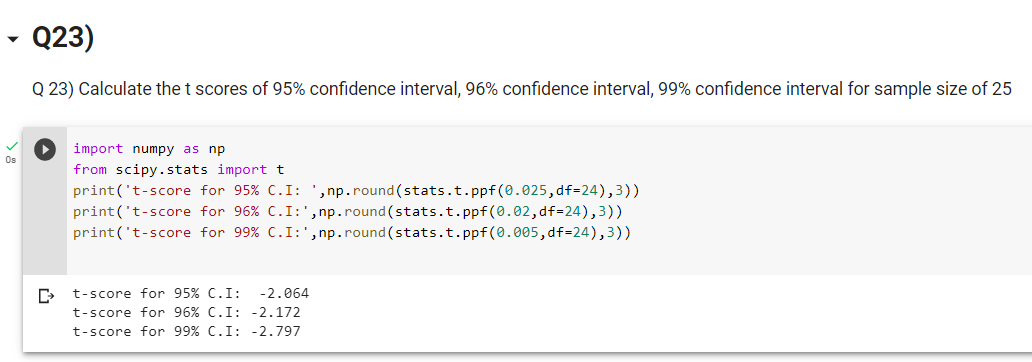
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

ANS : Kindly refer attached python code file for detailed answer



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

ANS : sample size = 25 , DOF = (25-1) = 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

ANS:

pop mean = 270

Sample Size = 18 ,DOF = 17 , sample mean = 260 , Stdev = 90

(Sample mean – pop mean)/(Stdev/sqrt(sample size)

(260-270)/(90/sqrt(18)) = -0.471

For more detailed answer kindly check the attached python code file.

Also kindly find the screenshots below for answers

