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

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 | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| 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 | Ordinal |
| Years of Education | Interval |

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

*Sample =HHT,HTH,THH* ***/*** *HHH ,TTT,HHT,HTH,THH,TTH,THT,HTT*

*probability that two heads and one tail are :* ***3/8***

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
4. *0*

*b) P(sum is less than or equal to 4) =6/36 =* ***1/6***

*c) P(sum = 6,12) =(5+1)/36=* ***7/36***

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 chances = 7c2*

*no of chances to draw two balls other than blue are = 2C2 + 3C2 + 2C1x3C1 =1+3+6=10*

*5c2 /7c2 = 10/21 =* ***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

Expected number of candies for a randomly selected child =

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.

**Use Q7.csv file**

**Mean** for Points = 3.596563 Score = 3.21725 Weigh = 17.84875

**Median** for Points = 3.695 Score = 3.325 Weigh = 17.71

**Mode** for Points = 3.92 Score =3.44 Weigh = 17.02

**Variance** for Points =0.285881 Score =0.957379 Weigh = 3.1931666

**Standard Deviation** for Points = 0.5346787 Score =0.978457 Weigh = 1.786943

**Range** for Points = 2.17 Score =3.911 Weigh = 8.4

Inferences :- **Large variations in points and weigh**

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?

E(x)= <X \* p(X) = (108+ 110+ 123+ 134+135+ 145+ 167+ 187+ 199) \* (1/9) =

**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**

**Q9\_a.csv Cars speed and distance**  
data.skew()Out[29]:

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

*Skewness inference:*

*Speed distribution is left skewed (negative skewness)*

*Distance distribution is right skewed(positive skewness)*

data.kurt() Out[30]:

Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

*Kurtosis inference:*

*Speed distribution is platykurtic (negative kurtosis : flatter than normal distribution)*

*Distance distribution is leptokurtic (positive kurtosis : peaked than normal distribution)*

**Q9\_b.csv SP and Weight(WT)**

data2.skew()

SP 1.611450

WT -0.614753

dtype: float64

*Skewness Inference:*

*1. SP distributioon is Right skewed (Positive skewness)*

*2. WT distribution is Left skewed (Negative skewness)*

data2.kurt() Out[16]:

SP 2.977329

WT 0.950291

dtype: float64

*Kurtosis Inference: Both the SP and WT distributions are leptokurtic (have positive kurtosis i.e. Peaked than normal distribution)*

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



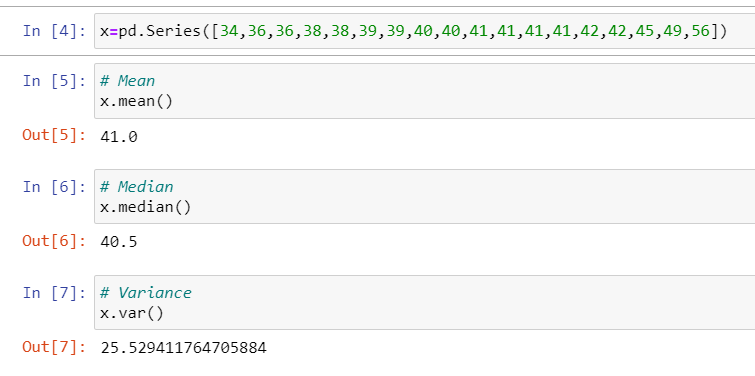
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

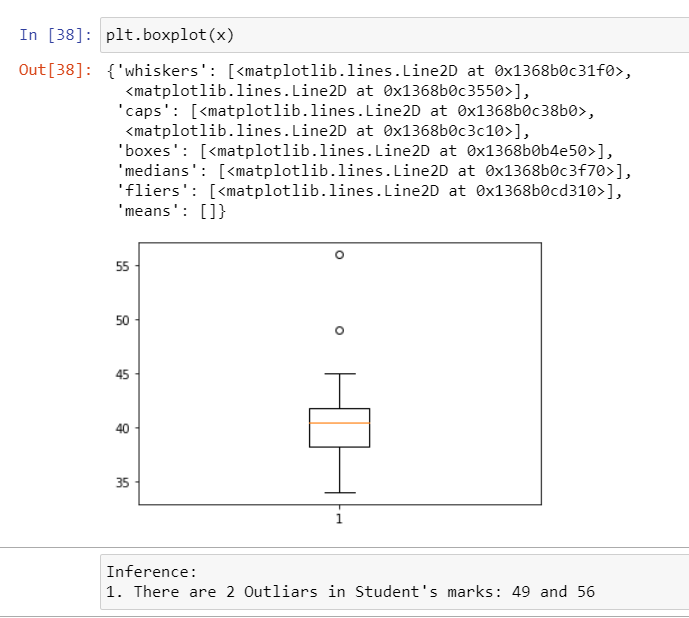


**Q12)** Below are the scores obtained by a student in tests

**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?





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

*Ans) There will be no skewness, It will be symmetric*

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

*Ans) When mean>median , It is tailed on the right side. Ie :- Right skewed.*

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

*Ans) When mean<median, It is tailed on the left side. Ie :- Left skewed.*

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

*Ans) Positive Kurtosis ,means sharp peak and less variation.*

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

*Ans) Negative Kurtosis ,means broad peak and more variation.*

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



What can we say about the distribution of the data?

*Ans) It is not a Normal Distribution .*

*50% of the values are above 10 and remaining are less.*

*And points with value above 15 are approximately 40%.*

What is nature of skewness of the data?

*Ans) It is left skewed. Median is greater than mean.*

What will be the IQR of the data (approximately)?   
*Ans) Approximately= 8  
 Inter Quartile Range =Upper Quartile- Lower Quartile => 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) mean and median are equal hence distribution is symmetrical.*

*The median of the two boxplots are same approximately 260.*

*2) The boxplots are not skewed in +ve or –ve direction.*

*3) Outliers doesn’t exist in both.*

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

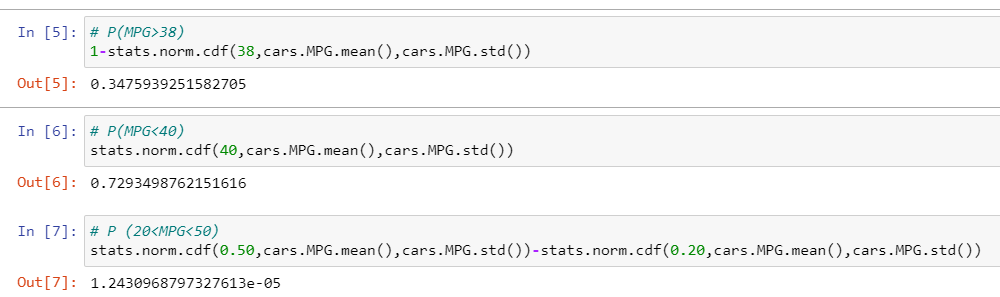
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

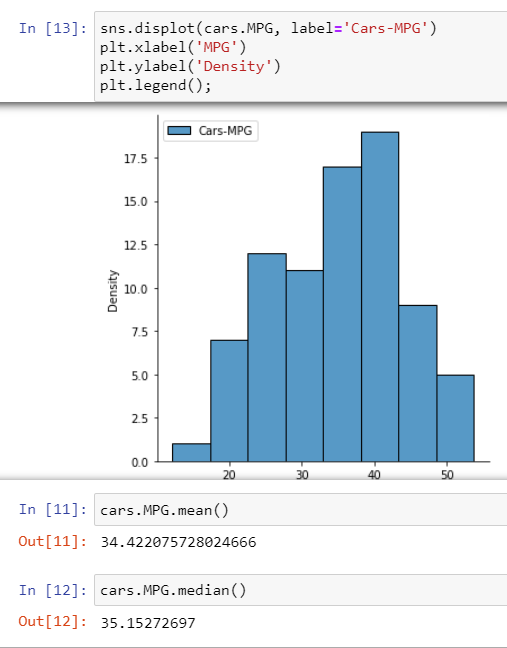
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

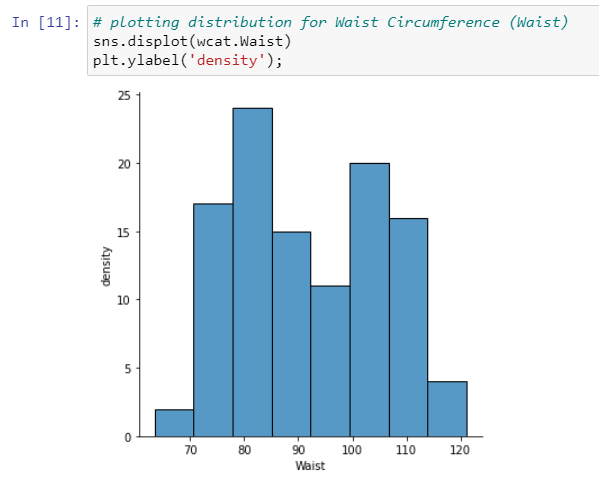
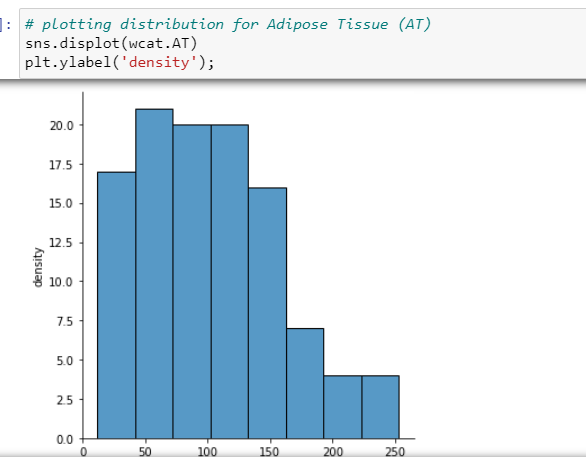
Dataset: Cars.csv

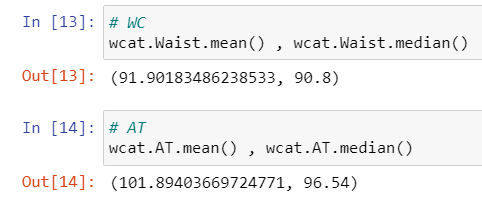


Inference: MPG of Cars does follow normal distribution approximately (as mean and median are approx. same)

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

Dataset: wc-at.csv

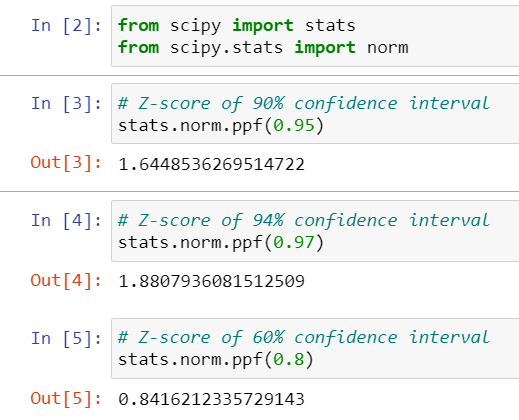
 



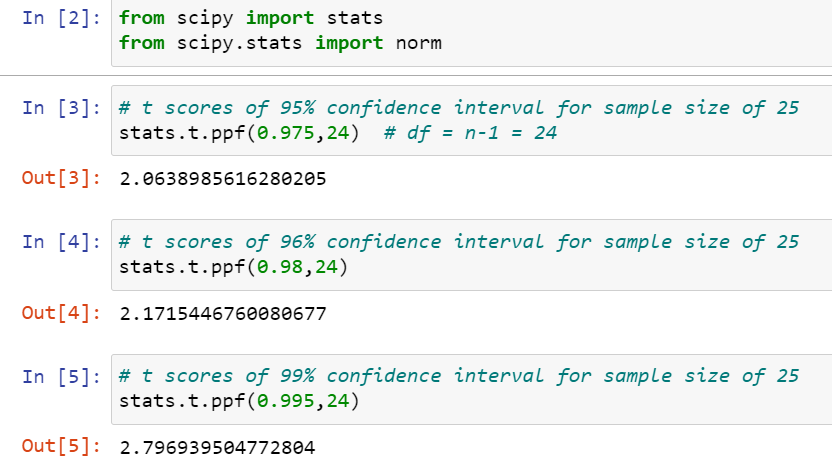
Inference: Both the Adipose Tissue (AT) and Waist Circumference(Waist) data set do follow the normal distribution approximately (as mean and median of both the data are approximately same)

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

nterval, 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

