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 | 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 | Ratio |
| Years of Education | Ratio |

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

Total possible outcome=8 (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT)

Combination of two heads and one tail=3 (HHT, HTH, THH)

Probability = favorable outcome/ total possible outcome

= 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

Total possible outcome = 36

1. Equal to 1 = favorable outcome/ total possible outcome

= 0/36

= 0

1. Less than or equal to 4

Favorable outcome: 4 => (1, 3), (3, 1), (2, 2)

3 => (1, 2), (2, 1)

2 => (1, 1)

Total favorable outcome = 6

Total possible outcome = 36

Probability = favorable outcome/ total possible outcome

= 6/36

= 0.166666

1. Sum is divisible by 2 and 3

Favorable outcome = 2 (6, 12)

Total possible outcome = 36

Probability = favorable outcome/ total possible outcome

= 2/36

= 0.0555555

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 possibilities = 7C2 ways (2+3+2)

= 7\*6/2\*1

= 21 ways

None of the balls drawn is blue = 5C2 ways (2+3)

= 5\*4/2\*1

= 10 ways

Probability = favorable outcome/ total possible outcome

= 10/21

= 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

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,Weigh>

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

**Use Q7.csv file**

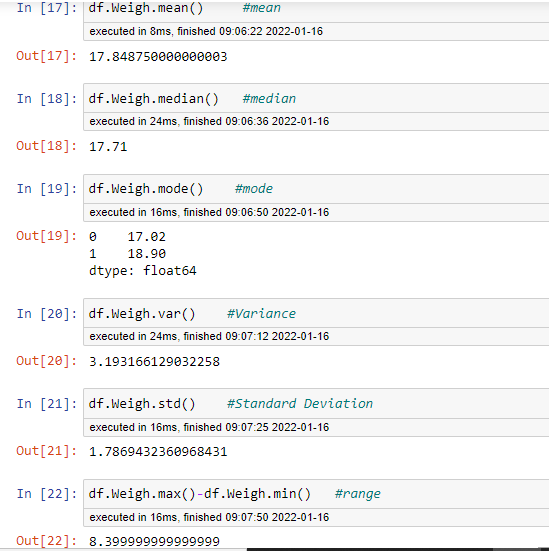
Mean, Median, Mode, Variance, Standard Deviation, and Range for Points column



Mean, Median, Mode, Variance, Standard Deviation, and Range for Score column



Mean, Median, Mode, Variance, Standard Deviation, and Range for Weigh column



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?

Probability of selecting one person=1/9

= (1/9)( 108+110+123+134+135+145+167+187+199)

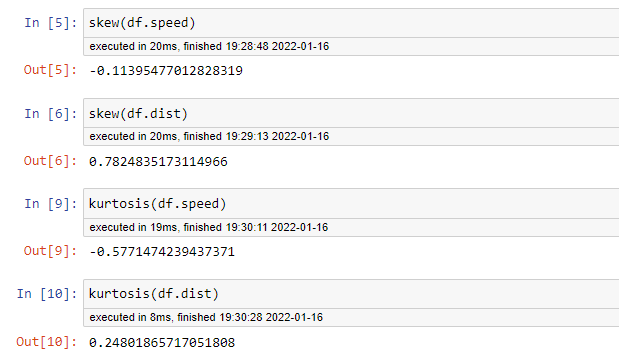
= (1/9)(1308)

= 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

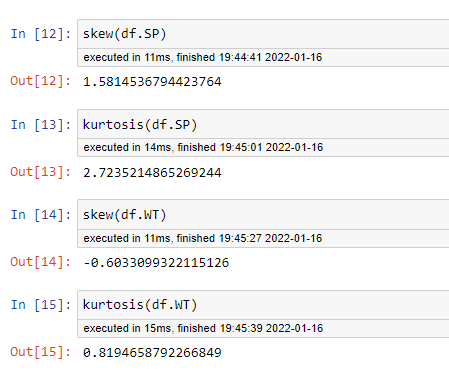


The skewness value for the speed column is –ve and the data is left skewed. Kurtosis value is –ve and it indicates data distribution has a flatter peakedness.

The skewness value for the dist column is +ve and the data is right skewed. Kurtosis value is +ve and it indicates data distribution has a sharper peakedness.

**SP and Weight(WT)**

**Use Q9\_b.csv**



The skewness value for the SP column is +ve and the data is right skewed. Kurtosis value is +ve and it indicates data distribution has a sharper peakedness.

The skewness value for the WT column is -ve and the data is left skewed. Kurtosis value is +ve and it indicates data distribution has a sharper peakedness.

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

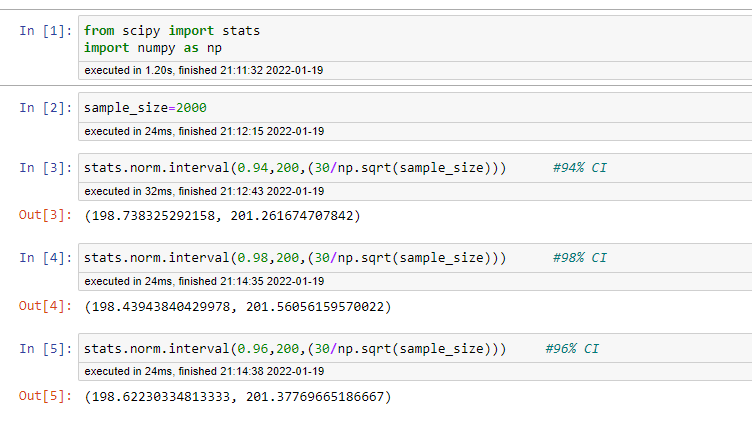


The Above histogram is right skewed. Skewness value for the above histogram is +ve.



The Above boxplot data is right skewed. Skewness value for the above boxplot data is +ve. There are outliers present in the data.

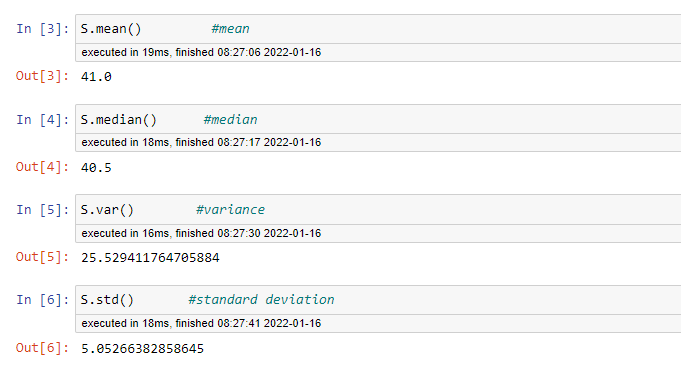
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate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

Data is perfectively symmetrical. Skewness value is 0.

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

Data is right skewed. Skewness value is +ve.

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

Data is left skewed. Skewness value is –ve.

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

It indicates data distribution has a sharper peakedness.

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

It indicates data distribution has a flatter peakedness.

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



What can we say about the distribution of the data?

Most of the data are distributed towards the right of the curve.

What is nature of skewness of the data?

The data is left skewed. Skewness value value is –ve.

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

= 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.

a) Both the data have the same median or Q2 i.e. have the approximate value of 262.5

b) Both the data are normally distributed.

c) Both the data do not have outliers.

d) Boxplot1 has lower IQR value when compared to the IQR value of Boxplot2.

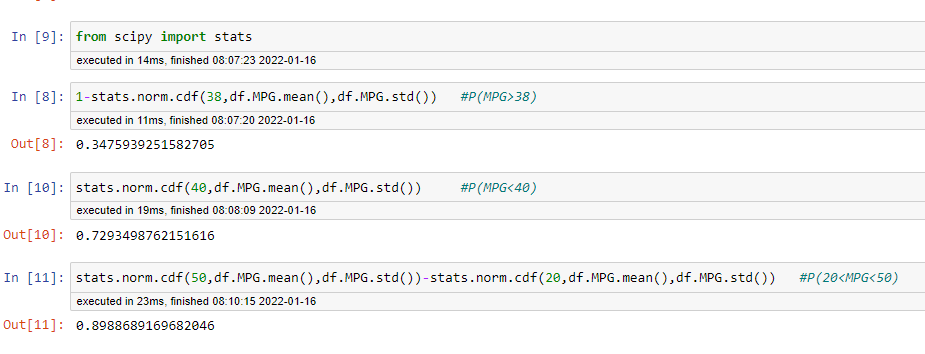
Q 20) Calculate probability from the given dataset for the below cases

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)

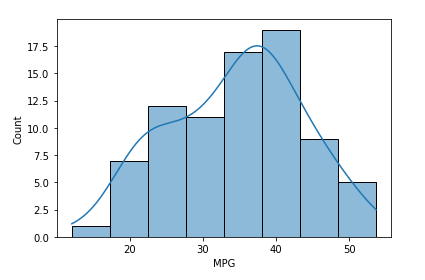


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

MPG data of cars doesn’t follow normal distribution.

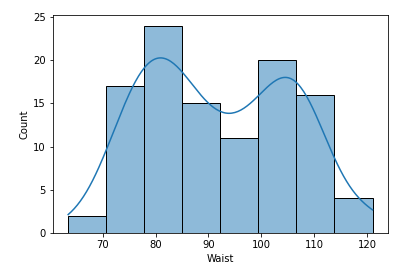


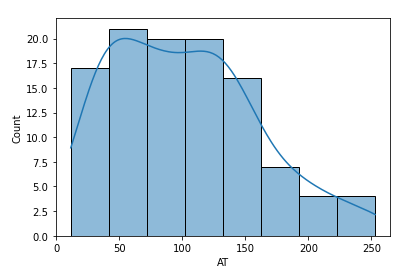
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

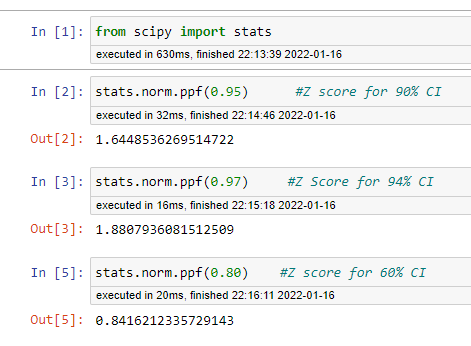
Adipose Tissue (AT) and Waist circumference (Waist) from WC-at data set

doesn’t follow normal distribution.

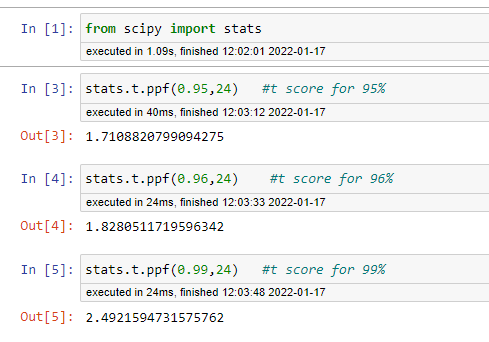




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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