|  |  |
| --- | --- |
| 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 | 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 | Nominal |
| 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 | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans: The sample space for tossing the 3 coins is

S={hhh, ttt, hht, thh, tth, htt, hth, tht}

P(2h’s,t)=2/8 =0.25

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

Ans:

1. 0
2. 6/36 =0.166
3. (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Probability for Sum is divisible by 2 and 3 =5/36 =0.138

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:

2 red, 3 green and 2 blue balls =7 balls

Number of ways drawing 2 ball is= = 21

2 balls drawn ,none of the ball is blue=5 10

The probability that none of the balls drawn is blue=10/21 =0.478

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:

Expected number of candies for a randomly selected child

P(RC) =  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

      P(RC) =3.090

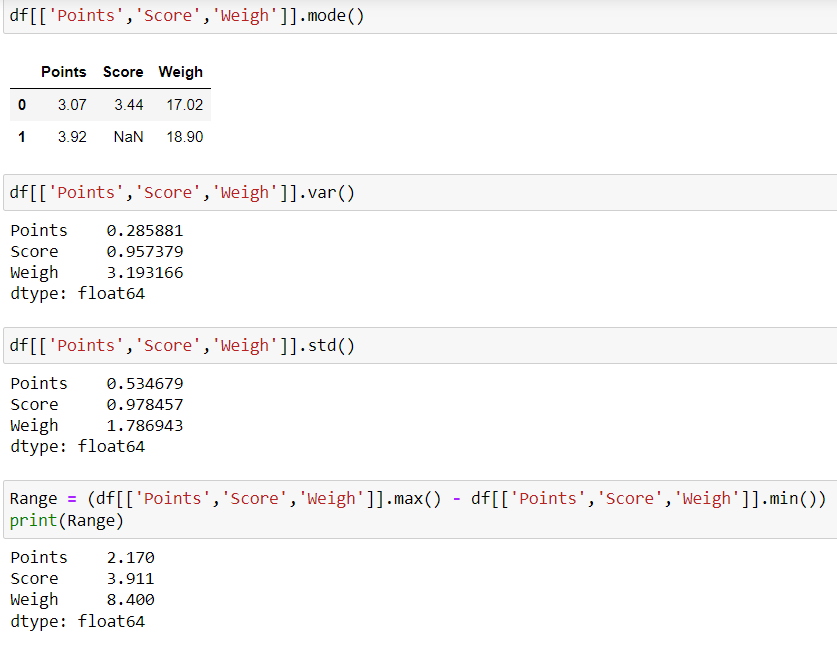
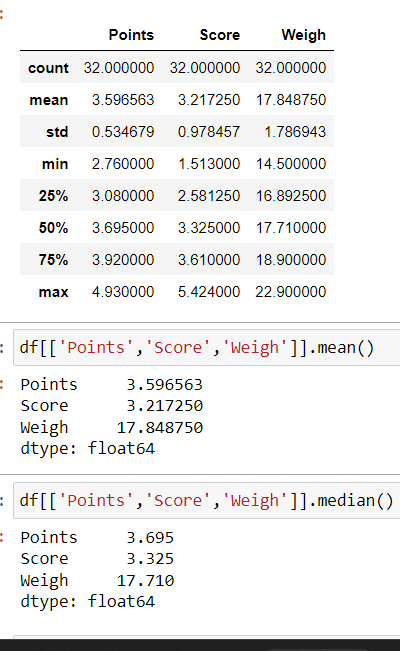
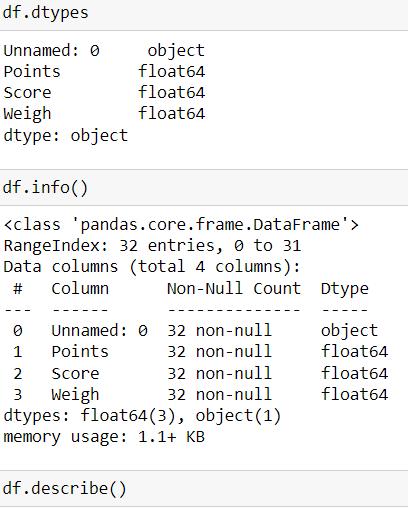
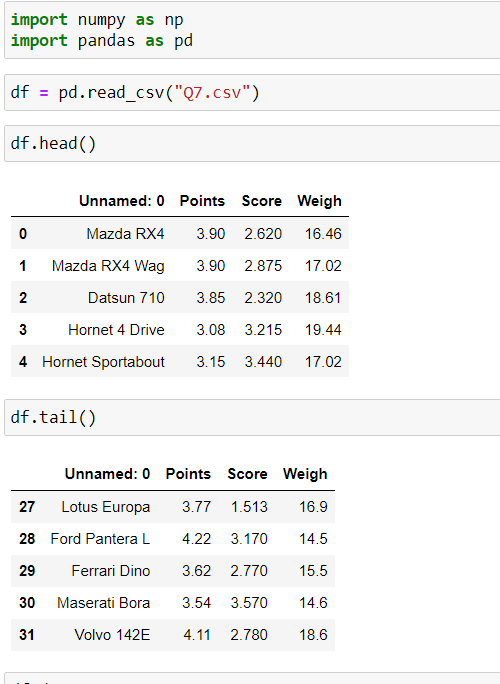
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

Ans:



The given data set Q7.csv has 4 columns and 32 rows. The attributes are not-null. In this data set the first column is an objective type data and the remaining 3 columns are float type data.

The mean of the attributes ‘Points’, ’Score’, ’Weigh’ are 3.596563, 3.217250,

17.848750 respectively.

The median of the attributes ‘Points’, ’Score’, ’Weigh’ are 3.695, 3.325, 17.710

respectively.

The mode of the attributes ‘Points’, ’Score’, ’Weigh’ are 3.07, 3.44, 17.02

respectively.

The variance of the attributes ‘Points’, ’Score’, ’Weigh’ are 0.285881,

0.957379,3.193166 respectively.

The standard deviation of the attributes ‘Points’, ’Score’, ’Weigh’ are 0.534679,

0.978457,1.786943 respectively.

The range of the attributes ‘Points’, ’Score’, ’Weigh’ are 2.170,

3.911, 8.400 respectively.

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: Given the weights of the 9 patients at a clinic

X : 108, 110, 123, 134, 135, 145, 167, 187, 199

The probability of selecting any patient is

P(X)= 1/9

Expected value of weight of choosing one patient at random is

E(X)=∑ X P(X)

E(X) = 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)

=12+12.22+13.66+14.88+15+16.11+18.55+20.77+22.11

E(X) = 145.3

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

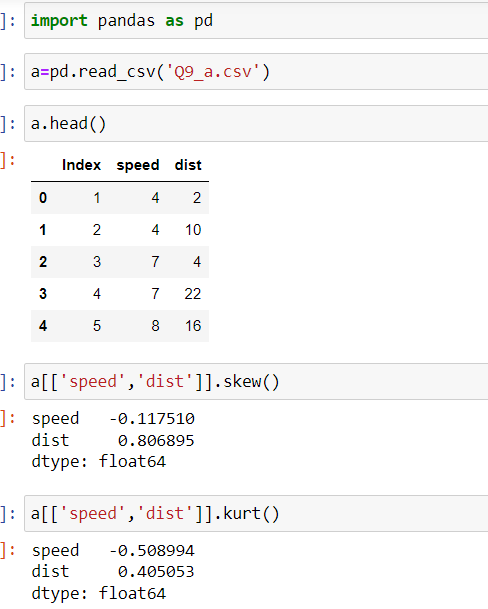
**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans:

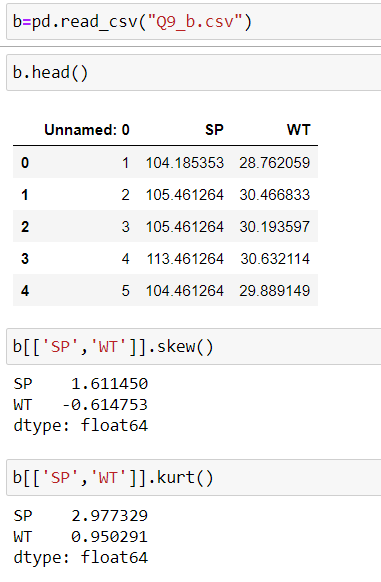


For speed we got skewness and kurtosis are negative values which means the speed of a car is negatively skewed that is the data falls on the right hand side.

Kurtosis is also negative value so we can say the shape of the speed of a car is platykurtic.

For distance we got skewness and kurtosis are positive values which means the distance of a car is positively skewed that is the data falls on the lefthand side.

Kurtosis is also positive value so we can say the shape of the distance of a car is leptokurtic.



For SP we got skewness and kurtosis are positive values which means the SP of a car is positively skewed that is the data falls on the left hand side.

Kurtosis is also positive value so we can say the shape of the SP of a car is leptokurtic.

For WT we got skewness negative values which means the WT of a car is negatively skewed that is the data falls on the right hand side.

Kurtosis is positive value so we can say the shape of the WT of a car is leptokurtic

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



Ans:

The given Histogram is plotted for the ChickWeights frequency

* The data is positively skewed.
* More than 50% Chick Weight is between 50 to 150.
* Most of the chick weight is between 50 to 100 with frequency of 200.

The box plot

* The data is right skewed or positively skewed.
* There are outlies at upper 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: Given

Sample size=2000

Mean =200

Standard deviation=30

Using the t-distribution

94% confidence level, with 200 - 1 = 199 df, the critical value is t = 1.8916

96% confidence level, with 200 - 1 = 199 df, the critical value is t =2.0673

98% confidence level, with 200 - 1 = 199 df, the critical value is t = 2.3452

The 94% confidence interval is (198.73, 201.27).

The 96% confidence interval is (198.61, 201.39).

The 98% confidence interval is (198.43, 201.57).

**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?

Ans:

1) Mean of the scores = 41

Median of the scores = 40.5

Variance = 24.1111

Standard deviation = 4.9103

2)The student scores were fluctuated with the standard deviation of 5.

The highest score obtained by students is 56 and the lowest score is 34.

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

Ans: mean=median=0 data is symmetric.

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

Ans: mean > median data is positively skewed.

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

Ans: median > mean data is negatively skewed.

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

Ans: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

Ans: A distribution with a negative kurtosis value indicates that the distribution has lighter tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Ans:

* The above Boxplot is not normally distributed the median is towards the higher value.
* The data is a skewed towards left. The whisker range of minimum value is greater than maximum.
* The Inter Quantile Range = Q3 Upper quartile – Q1 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: There are no outliers in the first Boxplot.

Both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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:



Q 21) Check whether the data follows normal distribution

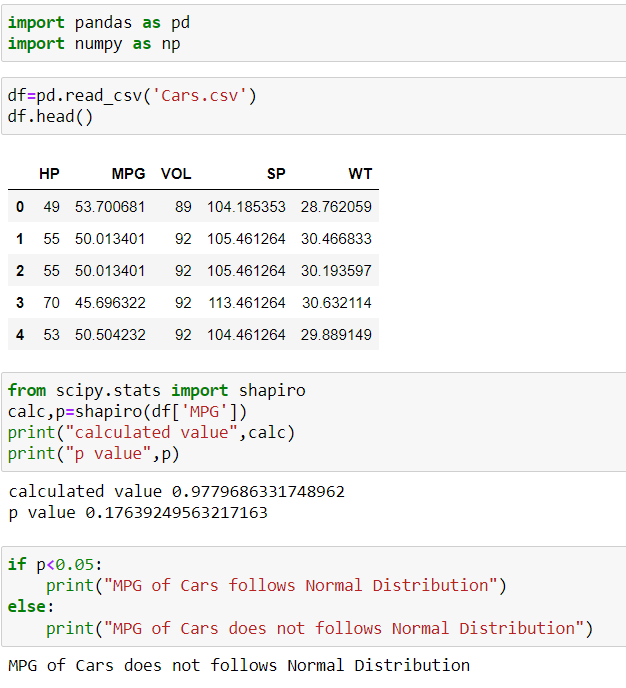
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

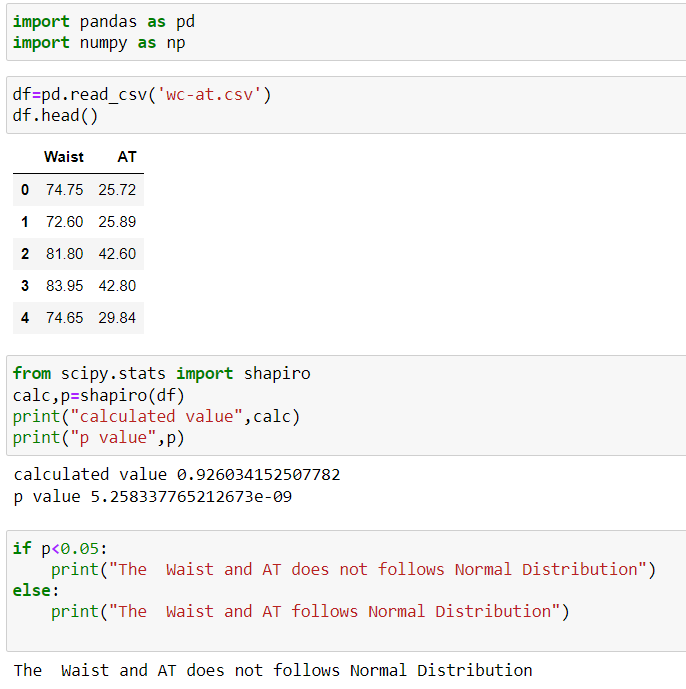
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: a)

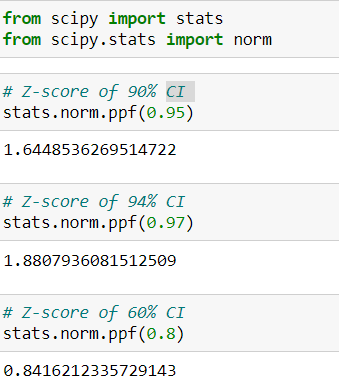


b)



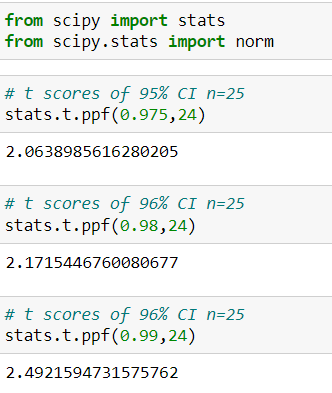
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Ans:



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

Ans:



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: Null Hypothesis is: Ho=Avg life of Bulb>=260 days

Alternate Hypothesis is: H1 = Avg life of Bulb<260 days

