|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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 | Discreate |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Continuous |
| Gender (Male or Female) | Discreate |

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 | Ordinal |
| 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 | Ratio |
| Years of Education | Nominal |

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

Ans= Total sample space = {HHH, TTT, THH, TTH, THT, HTT, HHT, HTH}

Probability of two head and one tail is= {THH, HHT, HTH}

Probability is =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

Ans=

1. Probability of sum is equal to 1= 0
2. Probability of sum is less than equal to 4=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 =

Total combination=nCr =n!/r!\*(n-r)

7C2=7!/2!(7-2)

=21

For no blue ball is drawn=5C2=5!/2!(5-2)

=10

The Probability of no blue ball is drawn is = 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= for expected number of candies for randomly selected child is

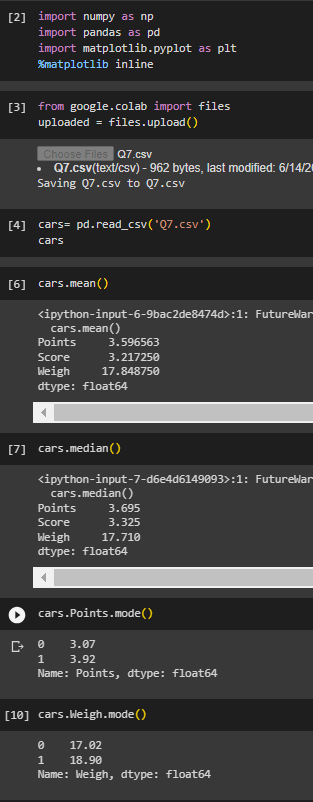
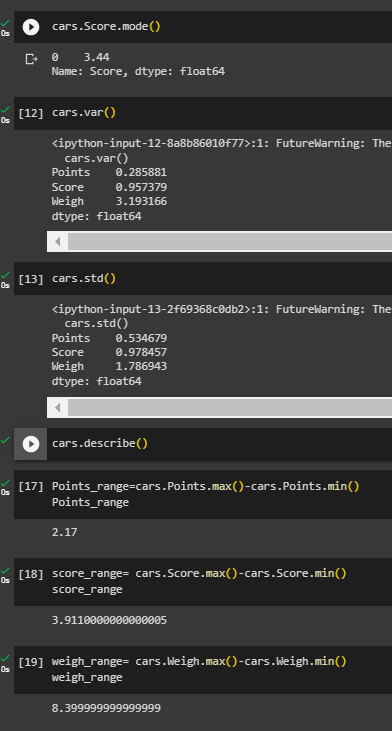
=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

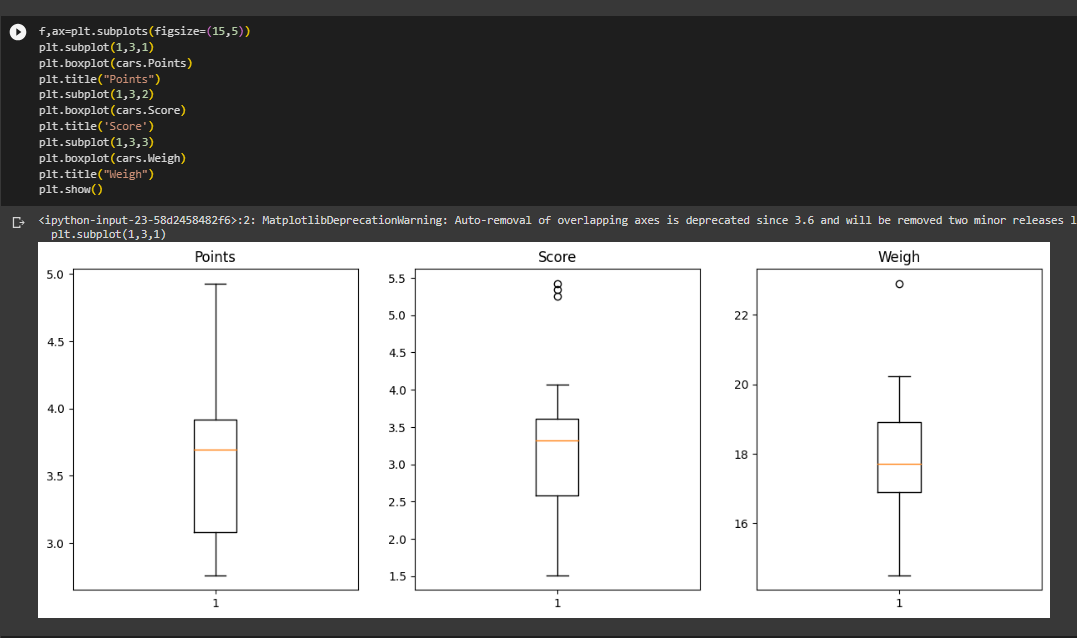
=3.09

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**

****

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= for average weight of patients=

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

=145.33 pounds

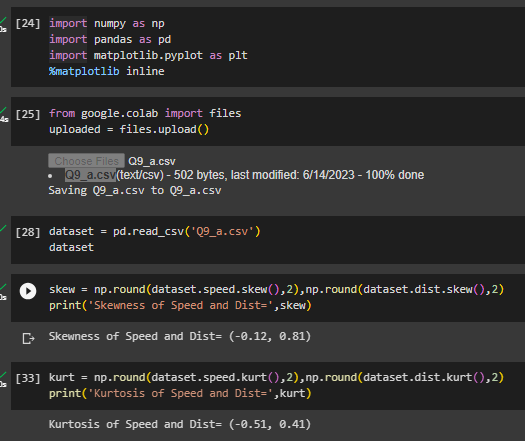
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

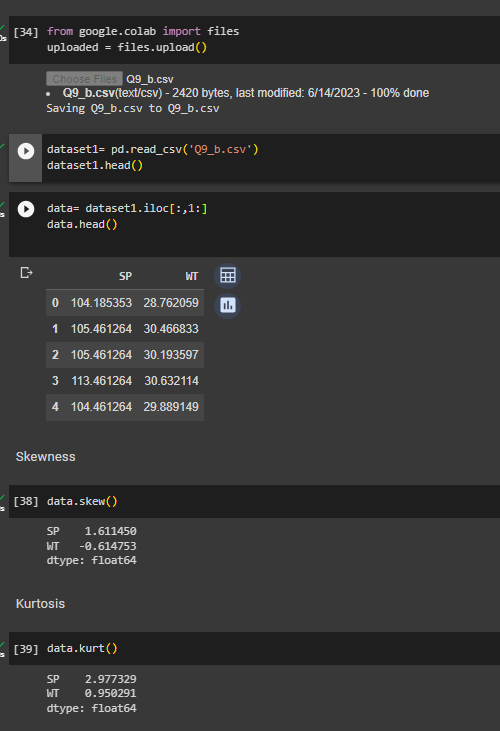
**Use Q9\_a.csv**

**Ans= Skewness of SP & WT is 1.611 & -0.614**

**Kurtosis of SP & WT is 2.977 & 0.950.**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness of SP & WT is 1.611 & -0.614**

**Kurtosis of SP & WT is 2.977 & 0.950.**

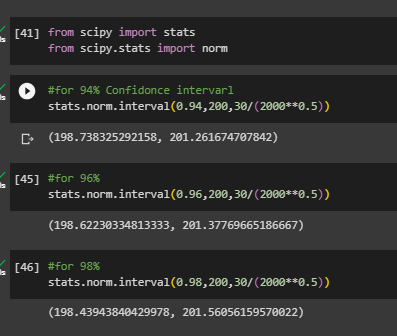
**Q10) Draw inferences about the following boxplot & histogram**



**Answer:**  Given histogram is right skewed, which means that there are lesser number of concentration of chic weight in the 300 – 400 category.

In the given boxplot, median is less than mean, which clearly mean that it has right skewed distribution. Also, there are some outliers on the upperside of the boxplot.

**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?

A screenshot of a computer

Description automatically generated

Here, we can say that the mean of student’s marks is greater than median which means the data is slightly skewed towards right. Also, two outliers can be observed in student’s marks dataset.

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

Ans= Zero skewed

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

Ans= Right skewed/ positively skewed.

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

Ans= Left skewed/ negatively skewed.

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

Ans= Distribution has thick tail compare to normal distribution

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

Ans= Distribution has lighter tail compared to normal distribution

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



What can we say about the distribution of the data?

Ans= The above boxplot is following normal distribution, the median is towards higher value

What is nature of skewness of the data?

Ans= The data is left skewed; the whisker range of minimum value is greater than maximum value.

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

Ans= IQR=Q3-Q1(Upper quantile – Lower quantile)

IQR=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 =

* Both the boxplot has no outliers.
* Both the boxplots are normally distributed, hence no skewness can be seen.
* Both the boxplots shares the same median.

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

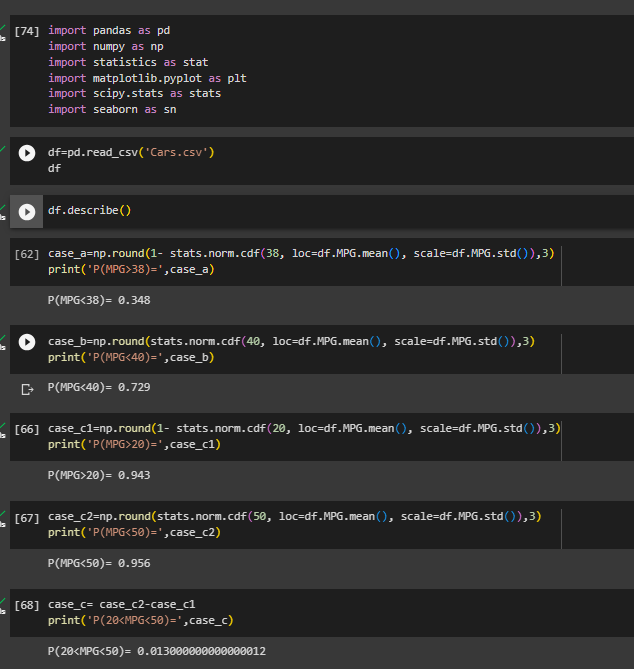
Data \_set: Cars.csv

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)

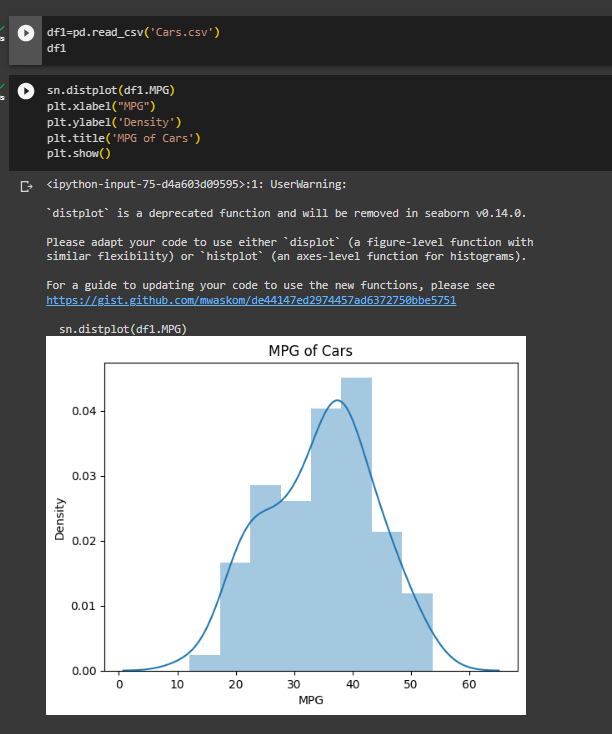


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

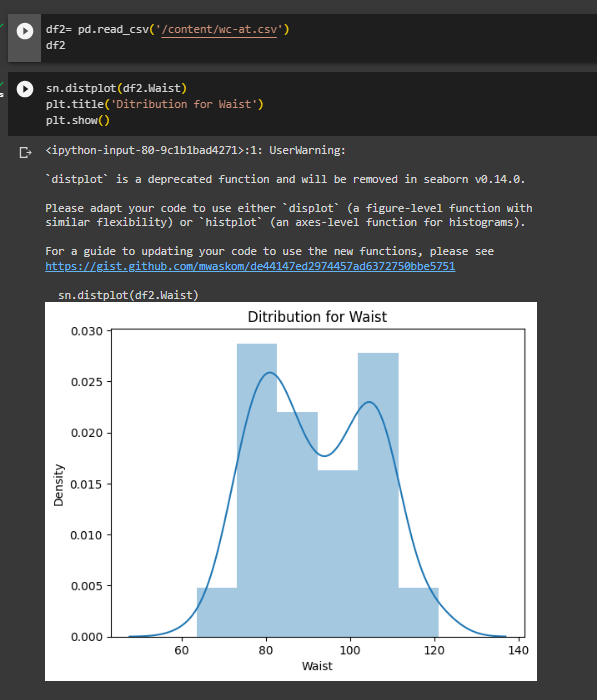
A screenshot of a computer

Description automatically generated

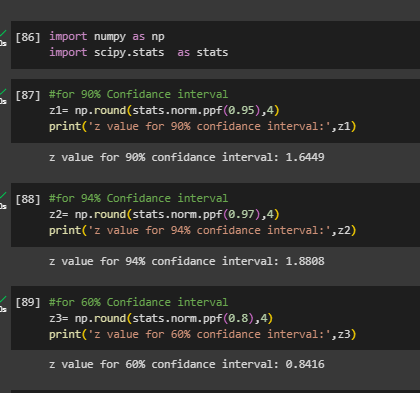
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

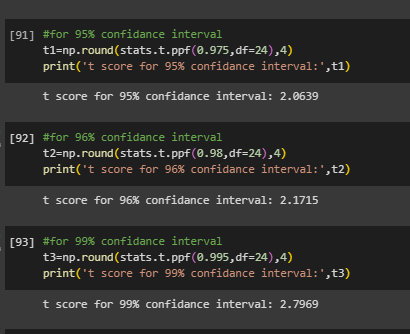
A screenshot of a computer screen

Description automatically generated

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

A screenshot of a computer program

Description automatically generated

Probability of life of a bulb not more than 260 days is **32%**.