|  |  |
| --- | --- |
| **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 |

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 |  |
| IQ(Intelligence Scale) |  |
| Sales Figures |  |
| Blood Group |  |
| Time Of Day |  |
| Time on a Clock with Hands |  |
| Number of Children |  |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education |  |

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

HTH , THH, HHT

**Ans:** Probability of getting two heads an done tail = 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

**Ans:** 0/36

1. Less than or equal to 4 =

**Ans:**

(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)

Probability = 1/6

1. Sum is divisible by 2 and 3

**Ans:**

Divisible by 2 : (1,1) (1,3) (1,5) (2,2) (2,4) (2,6) (3,1) (3,3) (3,5) (4,2) (4,4) (4,6) (5,1) (5,3) (5,5) (6,2) (6,4) (6,6) =18/36=0.5

Divisible by 3 :(1,2) (1,5) (2,1) (2,4) (3,3) (3,6) (4,2) (4,5) (5,1) (5,4) (6,3) (6,6) =12/36=0.33

Total probablility = 0.5+0.33 = 0.83

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:** 5/7 \* 4/6 = 20/42 = 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** : Expected number =

=1\*0.015 + 4\*0.2 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120

= 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.

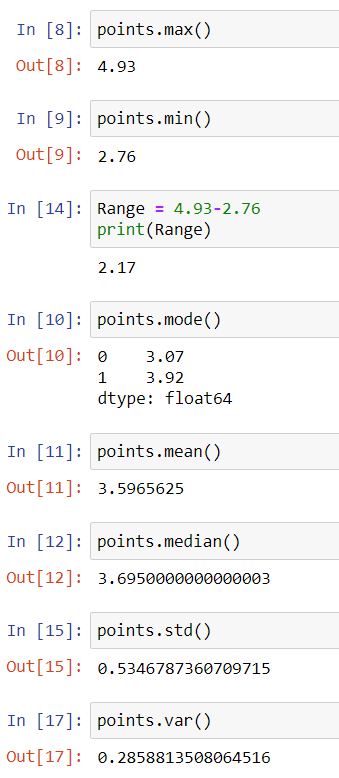
**Ans:**

* No case has shown Mean = Median = Mode
* Thus as seen in the graph dataset “score” and “Weigh” has outliers

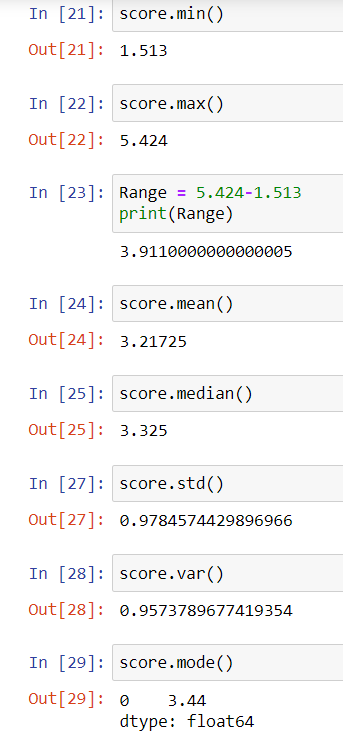




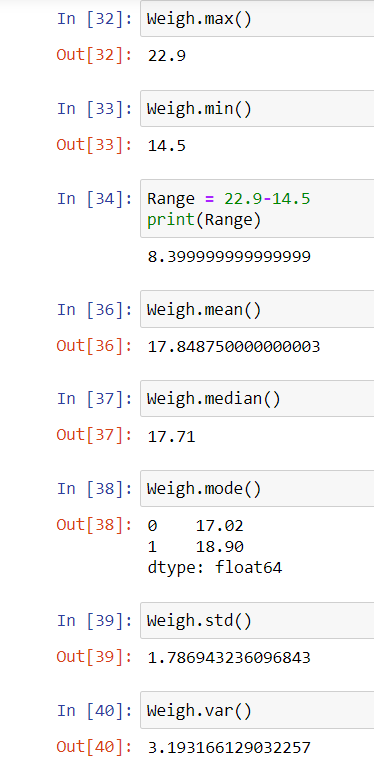












Q8) Calculate Expected Value for the problem below( google)

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 :**

Expected value

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

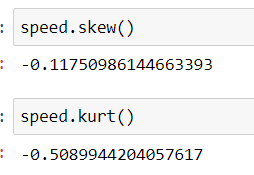
=1/9 \* 1308

=145.33

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

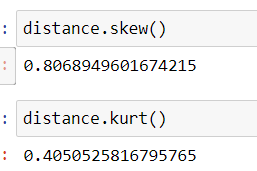
**Cars speed and distance**





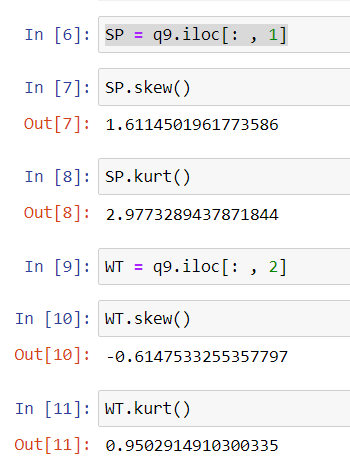
* “Distance” is positively skewed where as “speed” is negatively skewed .Thus distance has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the graph
* Distance has positive kurtosis and speed has negative Kurtosis ,which means its data much more varied compared to distance





**SP and Weight(WT)**

* SP” is positively skewed where as “WT” is negatively skewed Thus SP has distribution of data concentrated on the left whereas WT has distribution on the right. As seen in the graph
* Both WT and SP has positive Kurtosis



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



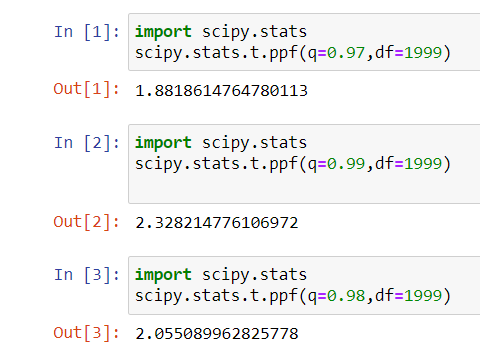
**Ans :** Most of the values of the data are towards the right . The tail on the right side of the distribution is longer , hence it is a positively skewed data . The mean and median will be greater than the mode.

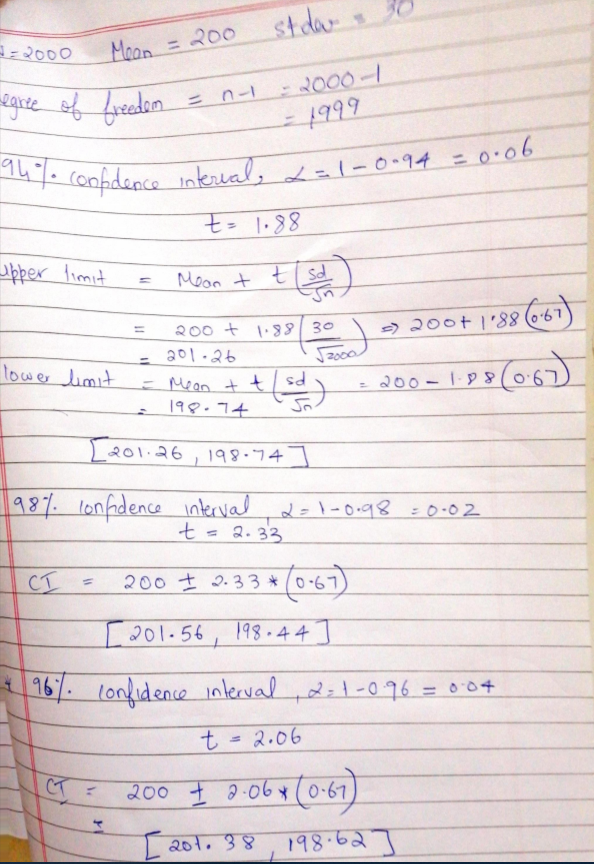


**Ans** :

When there are outliers in the data mainly on the right side , they are dotted outside the whiskers. we have outliers on upper extreme of box plot and there is less data points between Q1 and lower extreme. . The top whisker is much longer than the bottom whisker . SO this data is Positively skewed which means that long tail on right side and more data concentrate on left 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?



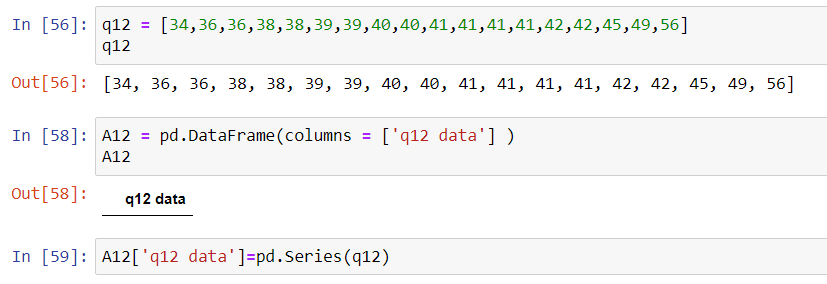


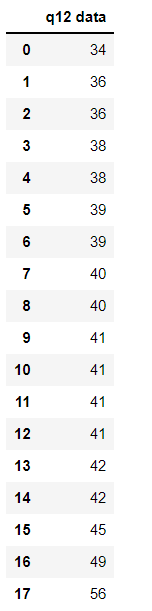
**Q12)** Below are the scores obtained by a student in tests

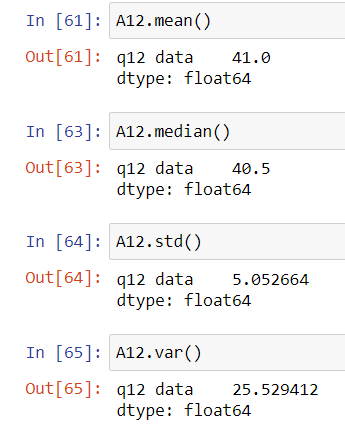
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

One of the students scored a very high mark compared to the other , that value is an outlier value , therefore it increases the mean







* The data is not normally distributed
* Data has outliers
* Majority of the students scored between 35 – 45 Marks

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

**Ans :** No skewness

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

**Ans :** Right skewness

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

**Ans :** Left skewness

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

**Ans :** It means that majority of the values in the data are closer to the mean . data does not have much variation . the curve has a sharp peak

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

**Ans :** It means that the majority of the values are spread far away from the mean . There is a flat curve , indicating more variation in the data

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



What can we say about the distribution of the data? Most of the values are on the left side of the median

What is nature of skewness of the data? This data is left skewed (negatively skewed data)

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

**Ans :** The data is not symmetric. Data is more concentrated towards right side . Hence it is positively skewed data

Q1 = 10

Q3= 18

IQR = Q3-Q1 = 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 :**

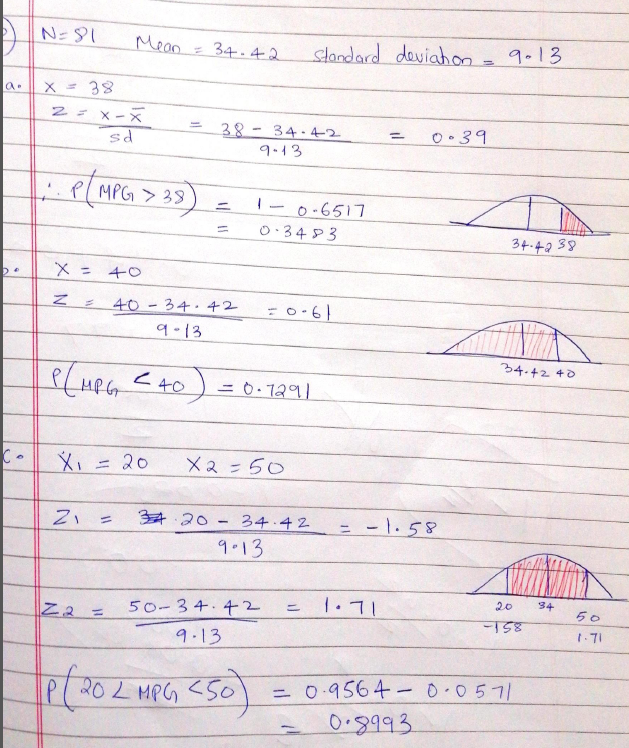
The boxplot 1 has a shorter box , Short boxes mean their data points consistently hover around the center median value and does not have much variance. But compared to the boxplot 1 , boxplot 2 has a bigger box .Taller boxes imply more variable data. The data points are spread out more away from the mean compared to the first one.

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

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

MPG <- Cars$MPG

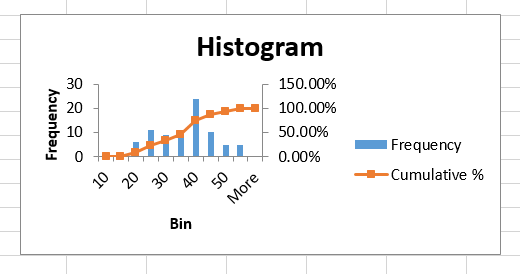
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution(chk using data analytics tool kit)

1. Check whether the MPG of Cars follows Normal Distribution

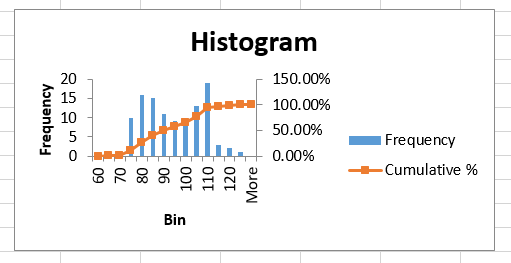
Dataset: Cars.csv



**It follows 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

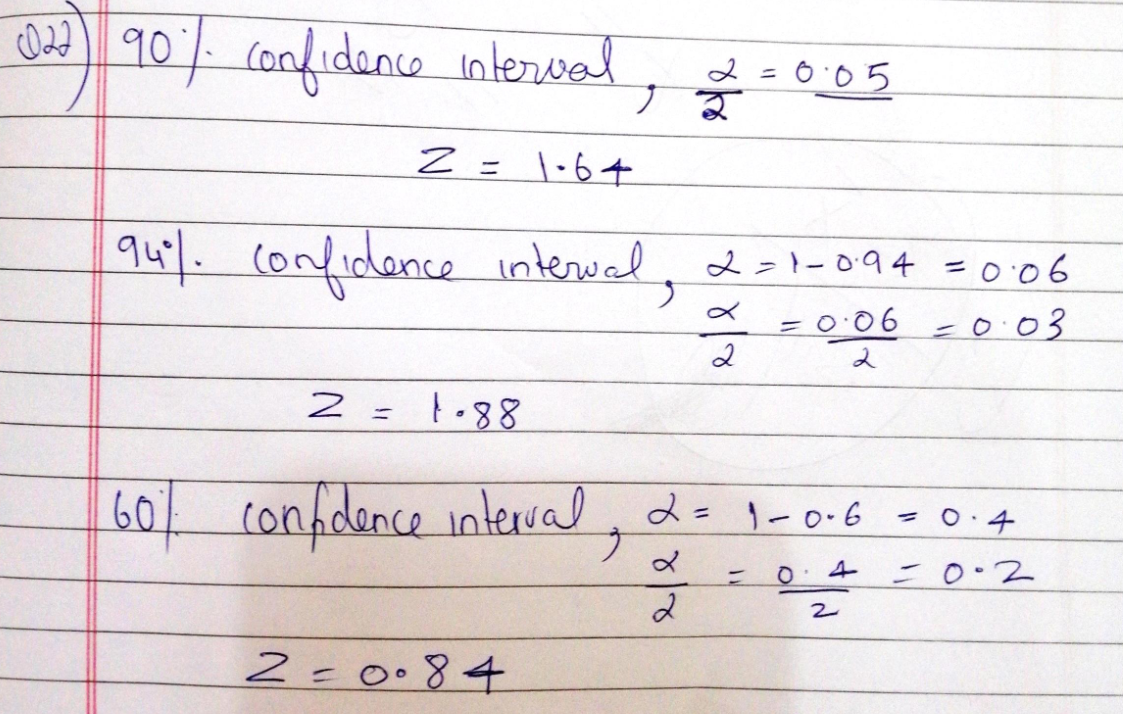


**Waist – it does not follow normal distribution**

**AT – it does not follow normal distribution**

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

**Ans:**



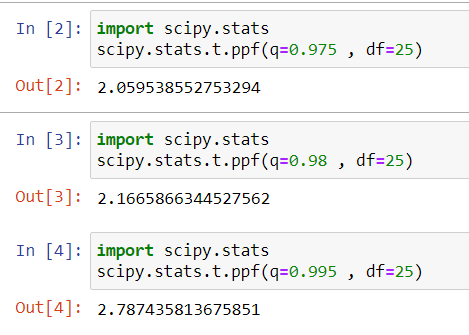
Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

**Ans :**

95% Confidence interval , t = 2.06

96% Confidence interval , t = 2.17

99% Confidence interval , t = 2.79



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 :**

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t =

t =

t = -0.47

degree of freedom = 17

p value = 0.322

therefore , probability that 18 randomly selected bulbs would have an average life of no more than 260 days is = 32%