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

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 | Ratio |
| Fahrenheit Temperature | Interval |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

HHT, HTH, THH

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

1. Equal to 1 Ans =0
2. Less than or equal to 4
3. Sum is divisible by 2 and 3
4. No of possible sum equal to 1 = 0

Probability of sum equal to 1 = 0

1. No of possible outcomes with sum less than equal to 4 = (1,3),(2,2)(3,1)(1,2),(2,1)(1,1) = 6

Total possible outcomes =36

Probability = 6/36 = 1/6

1. No of possible outcomes with sum divisible by 2 & 3 = (3,3),(4,2),(1,5),(5,1),(2,4),(6,6)=6

Total possible outcomes =36

Probability = 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?

Total balls = 2+3+2= 7

n Cr

where n = total no of balls

r = no of balls to be drawn at random

7C2 = 21

Now,

We find two balls drawn are blue = 5C2 = 10

Probability = 5C2 /7C2  = 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 |

Ans =

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

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?

**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**

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

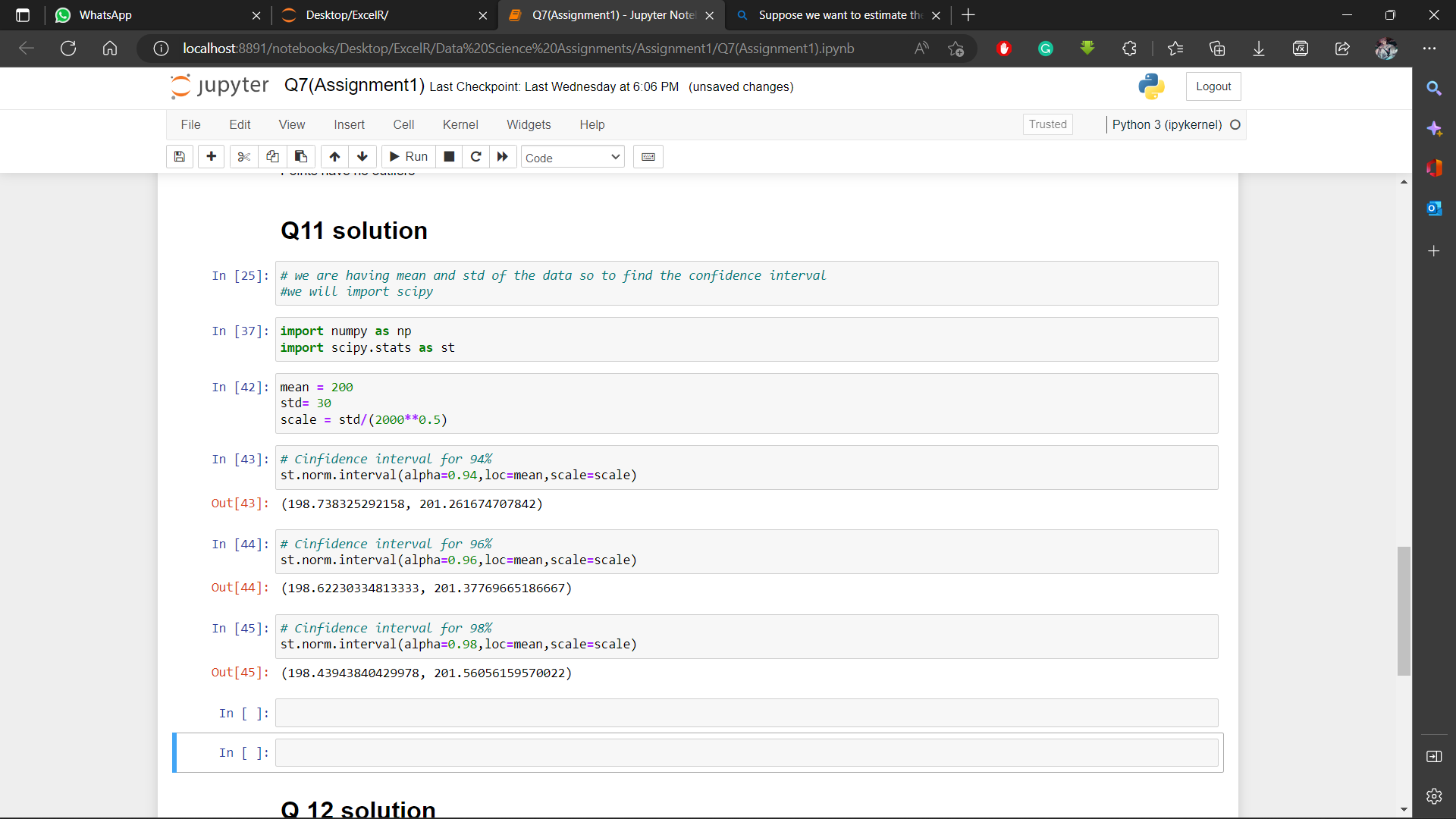


**Ans :-** The Histogram is Positively skewed as the tail of the curve is widely spreading across positive side and the data may contain some outliers on the same side.



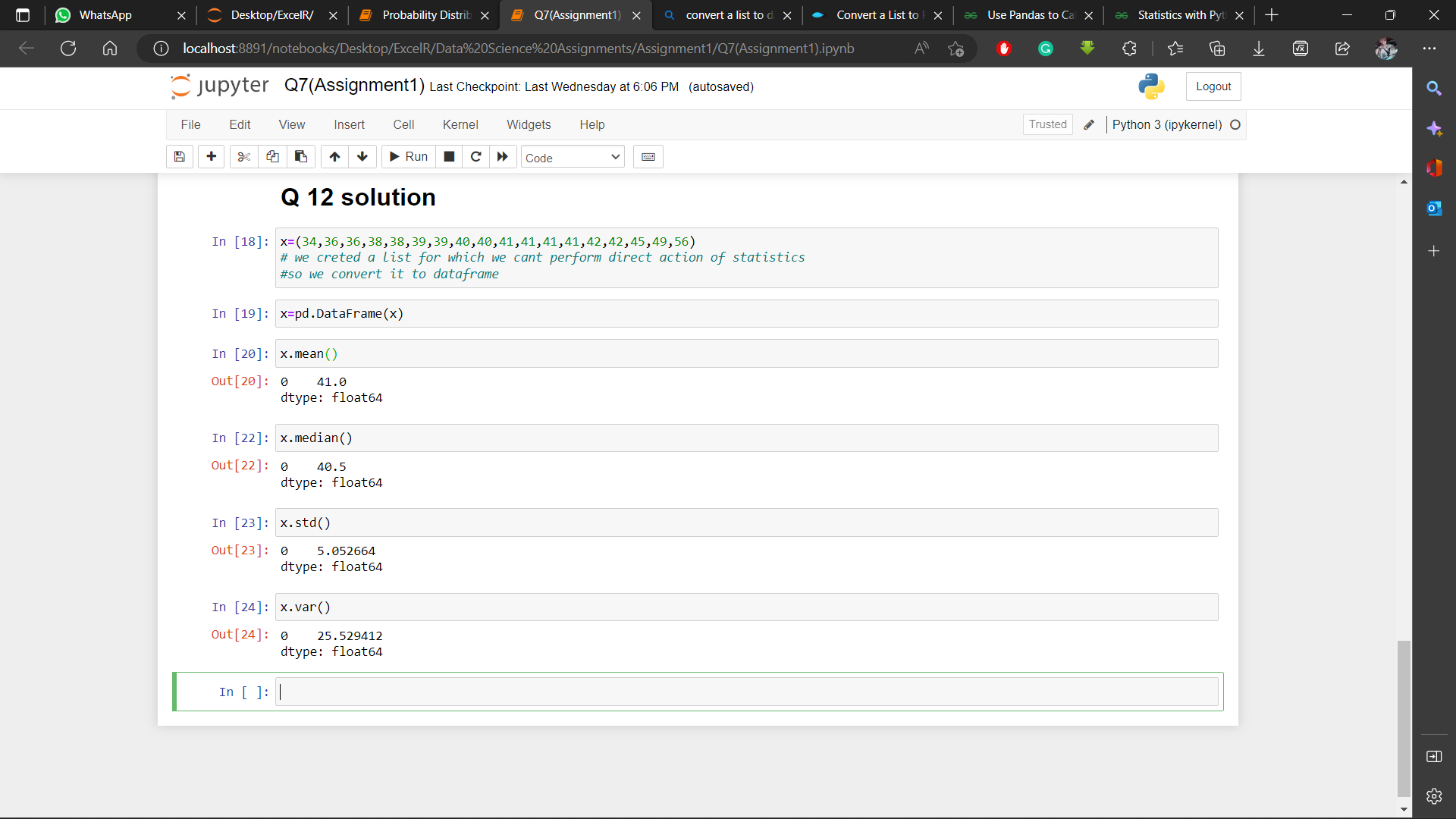
**Ans :-**The boxplot is showing that the data contains many outliers (7)on the Upper Extreme region and the whiskers length is spreading on the Upper extreme side which tells Data is not Normally Distributed .

**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?

**Ans:-** we can say the most students have scored around 40 marks

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

**Ans:-**  Symmetrically Distributed Data has equal value of mean and median.

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

**Ans:-** The Nature of skewness is Positively skewed when the mean of the data is greaten than the median of the data.

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

**Ans:-** The Nature of skewness is Negatively skewed when the median of the data is greater than the mean of the data.

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

**Ans:-** A Positive kurtosis indicates that the distribution has a sharp peak and wide tails like a high mountain.

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

**Ans: -** Negative Kurtosis indicates that the peak is widely spread and tinner tails like a plateau .

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



What can we say about the distribution of the data?

**Ans: -** The Data is not normally distributed as the Boxplot has the values near to the maximum and medians are also not at equal distance.

What is nature of skewness of the data?

**Ans: -** The nature of Data is Negatively skewed as we can see the boxplot is showing the median near to the maximum value of the data .

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

**Ans: -** LQR(Lower Quartile Range)= 10 UQR(Upper Quartile Range)=18

IQR= UQR-LQR

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 Inter-Quartile range for plot one is approximately 25

and that of plot 2 is around 90

The mean of plot1 is between 260 to 265 and of plot2 is also around same

By view of the graphs, we can say that the data of plot is widely spread in comparison with the plot 1

And on normal distribution curve plot2 has far wider tail than plot1

Both the plots do not have any outliers and plot 1 seems to be little bit of negatively skewed as its mean tends towards maximum.

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**:- total no of outcomes = 81

Outcomes for mpg>38=33 Probability=33/81=0.407407

Outcomes for mpg<40=61 Probability= 61/81=0.7530864

Outcomes for 20<mpg<50=69 Probability=69/81=0.851851

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans**:- No the data in MPG of cars is not normally distributes it is Negatively skewed

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

Dataset: wc-at.csv

**Ans**:- AT is Positively skewed as the median is more closer towards the minimum value .

Waist is sightly Positively skewed as median is closer to the minimum as compared to maximum

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

**Ans**:- confidence interval for 90% = 1.644

confidence interval for 94% = 1.880

confidence interval for 60% =0.253

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

**Ans**:- the probability that 18 randomly selected bulbs would have an average life of no more than 260 days will be 0.32167

**I have solved all the required questions in my Jupyter Notebook**

**And have sent it with this.**

**And uploaded them on my GitHub**

[Ajay3736pandey/Assignment-1-DS- (github.com)](https://github.com/Ajay3736pandey/Assignment-1-DS-)