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

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

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

**S = {HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}**

**A = 2 heads and 1 tail occurs**

**A = {HHT, HTH, THH}**

**P (A) = 3/8 = 0.375**

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

**S = {11,12,13,14,15,16,21,22,23,24,25,26,31,32,33,34,35,36,41,42,43,44,45,46,51,52,53,54,55,56,61,62,63,64,65,66}**

1. **Equal to 1**

**A = {}**

**P(A) = 0/36 = 0**

1. **Less than or equal to 4**

**B = {11, 12, 13, 21, 22, 31}**

**P(B) = 6/36 = 0.1666**

1. **Sum is divisible by 2 and 3**

**C = {11,12,13,15,21,22,24,26,31,33,35,36,42,44,45,46,51,53,54,55,62,63,64,66}**

**P(C) = 24/36 = 0.6666**

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?

**S = {RR, RG, RB, GG, GB, BB}**

**A = None of the balls are blue**

**A = {RR, RG, GG}**

**P(A) = 3/6 = 0.5**

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.**

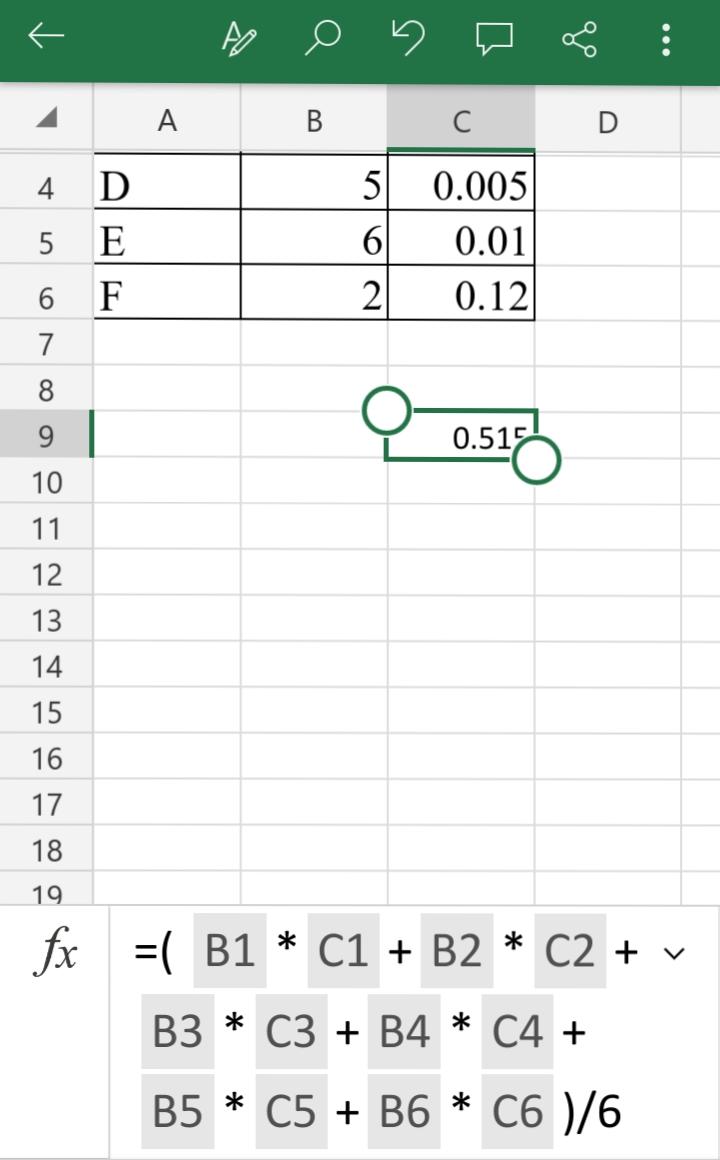
**Child B – probability of having 4 candies = 0.20**

**Let, X = number of candies for a child**

**Then –**

**E(X) = SUM of (candies count \* probability)**

**E(X) = 0.515**

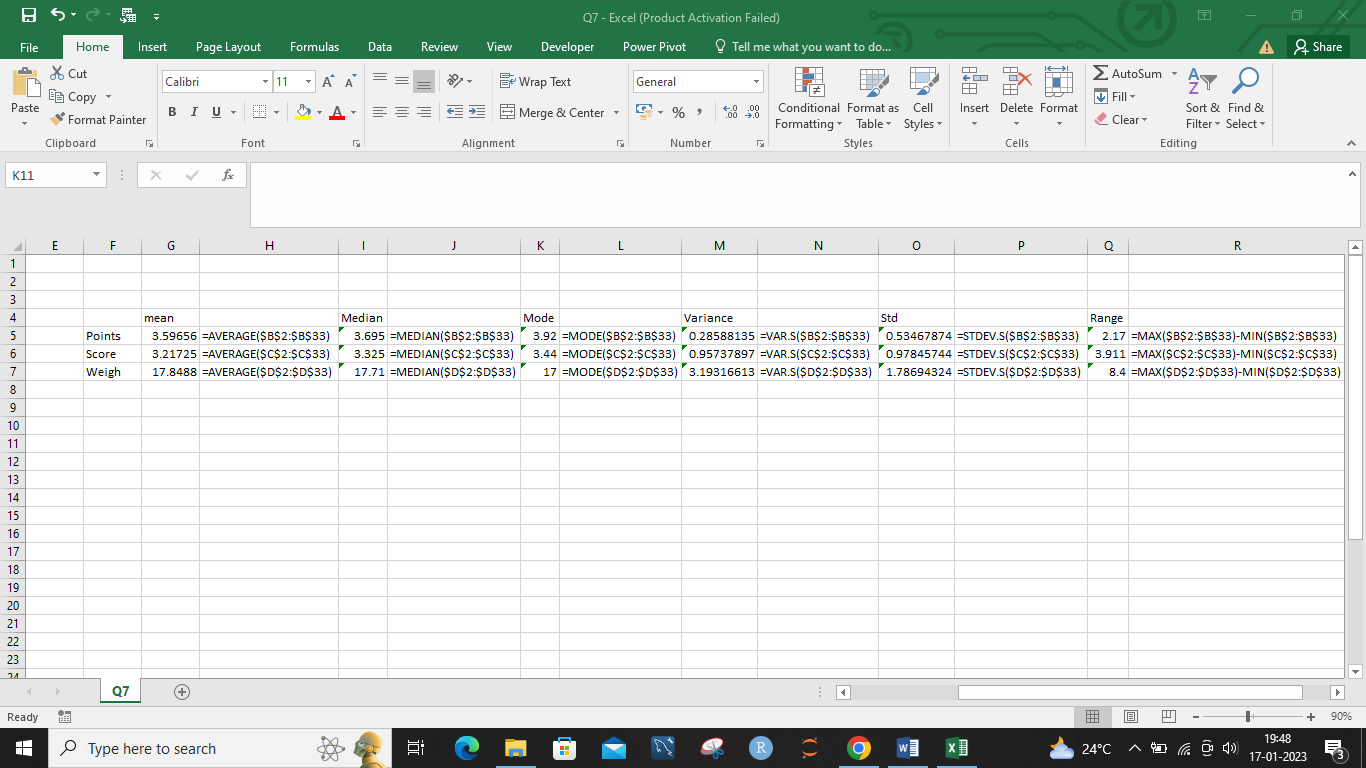


Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

****

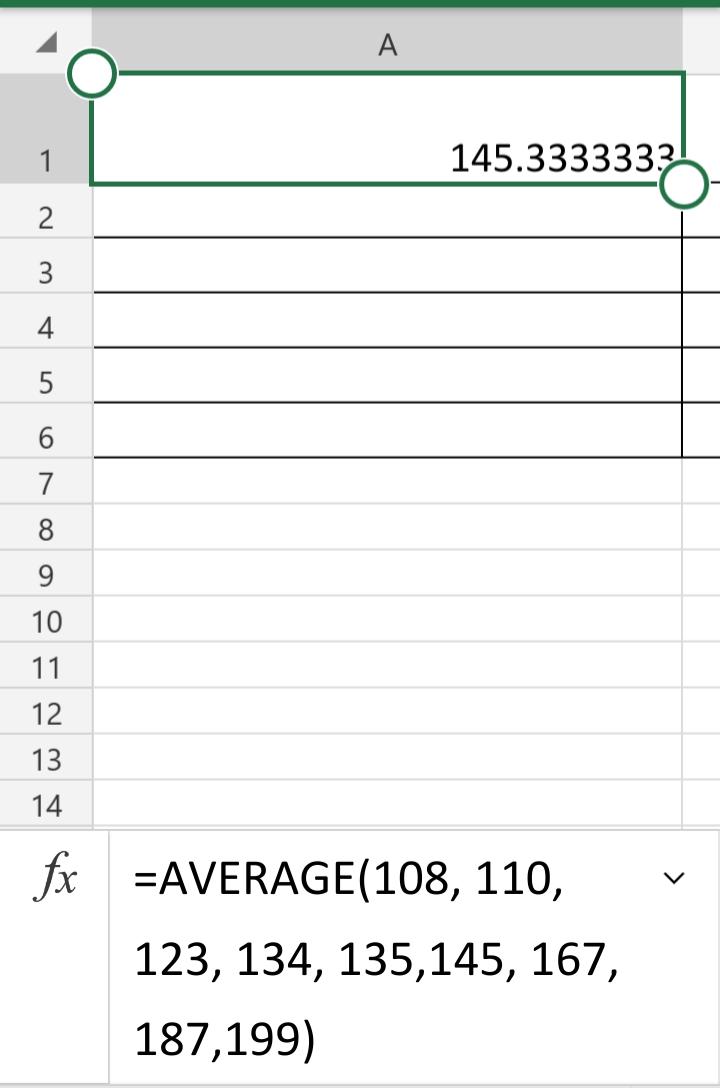
Q8) Calculate Expected Value for the problem below

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?

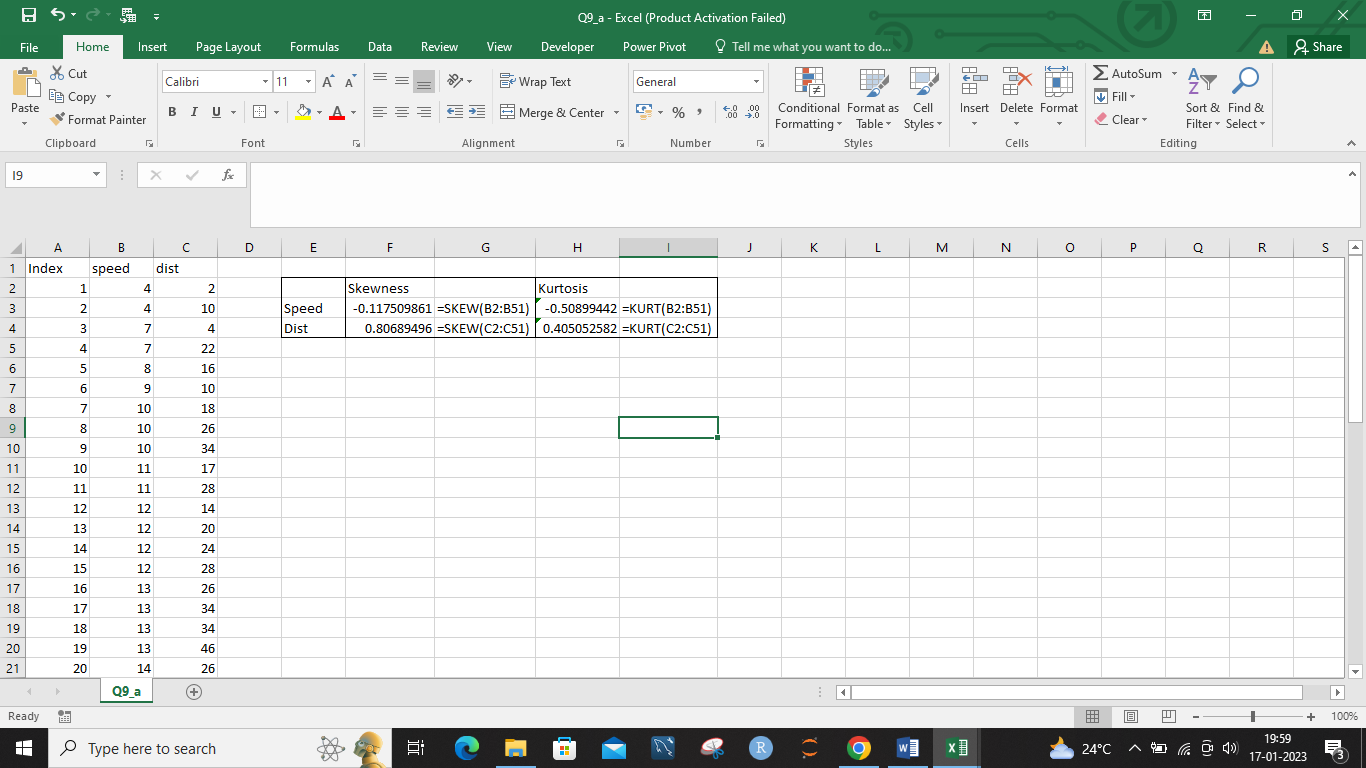
Expected value = 145.3333



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

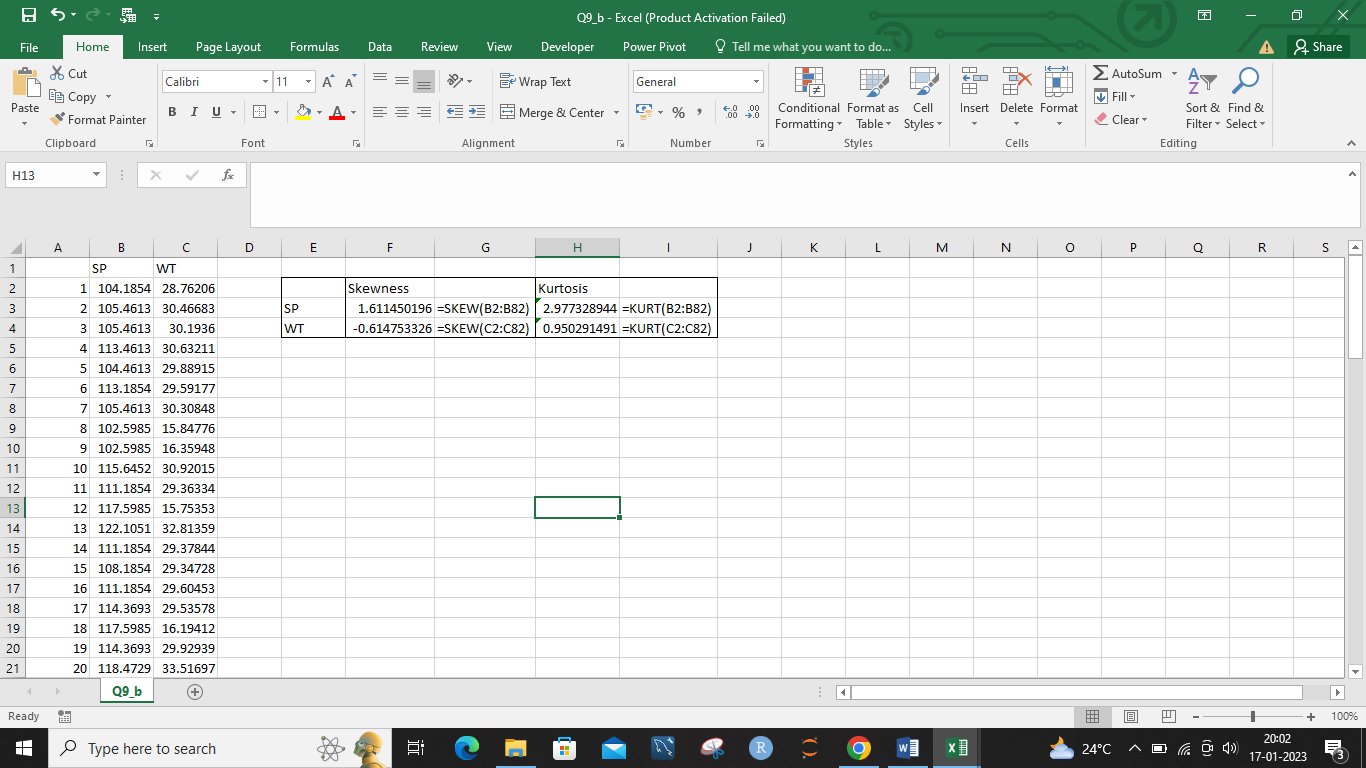
**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



**From histogram we can say that, the maximum frequency of chick weights lies in 50-100 weights.**

**The chick weights looks positively skewed.**



**The above boxplot shows seven outliers.**

**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?

**Answer:**

**Standard error of the mean= (30 pounds) / sqrt(2000) = 3 pounds**

**A 94% confidence interval would be calculated as:**

**(Sample mean) - (critical value \* SEM) < (population mean) < (sample mean) + (critical value \* SEM)**

**Critical value for 94% confidence is 1.88**

**200 - (1.88 \* 3) < (population mean) < 200 + (1.88 \* 3)**

**191.24 < (population mean) < 208.76**

**So, the 94% confidence interval for the average weight of adult male in Mexico is 191.24-208.76 pounds.**

**Similarly, A 98% confidence interval would be calculated as:**

**(Sample mean) - (critical value \* SEM) < (population mean) < (sample mean) + (critical value \* SEM)**

**Critical value for 98% confidence is 2.33**

**200 - (2.33 \* 3) < (population mean) < 200 + (2.33 \* 3)**

**190.33 < (population mean) < 209.67**

**So, the 98% confidence interval for the average weight of adult male in Mexico is 190.33-209.67 pounds.**

**A 96% confidence interval would be calculated as:**

**(Sample mean) - (critical value \* SEM) < (population mean) < (sample mean) + (critical value \* SEM)**

**Critical value for 96% confidence is 2.05**

**200 - (2.05 \* 3) < (population mean) < 200 + (2.05 \* 3)**

**192.15 < (population mean) < 207.85**

**So, the 96% confidence interval for the average weight of adult male in Mexico is 192.15-207.85 pounds.**

**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?

**Answer:**

1. **To calculate the mean, add up all the scores and divide by the number of scores: Mean = (34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56) / 18 = 39.6**

**To find the median, we need to arrange the scores in numerical order and find the middle value. Since there are 18 scores, the median will be the 9th value: Median = 39**

**To find the variance, we need to calculate the deviation of each score from the mean, square it, and take the average**

**variance = 31.6**

**To find the standard deviation, we take the square root of the variance: standard deviation = sqrt(31.6) = 5.6**

1. **From the above analysis, we can say that the student's mean score is 39.6 which is relatively close to the median score of 39. The standard deviation of 5.6 indicates that the scores are relatively close to the mean with most of the scores falling within 5.6 points of the mean. However, there are some high scores like 56 that are farther away from the mean. Overall, the student's scores are relatively consistent and they have performed well.**

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

**Answer: when the mean and median of a dataset are equal, it means that the distribution is symmetric and the skewness is zero.**

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

**Answer: When the mean is greater than the median in a data set, it indicates that the distribution of the data is positively skewed or right-skewed.**

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

**Answer: When the median is greater than the mean in a data set, it indicates that the distribution of the data is negatively skewed or left-skewed.**

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

**Answer: A positive kurtosis value indicates that a distribution has a higher peak than a normal distribution and the tail is "heavier" or has more observations than a normal distribution. This means that the distribution has more frequent extreme values (outliers) than a normal distribution. This type of distribution is called a leptokurtic distribution.**

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

**Answer: A negative kurtosis value indicates that a distribution has a flatter peak than a normal distribution and the tail is "lighter" or has fewer observations than a normal distribution. This means that the distribution has less frequent extreme values (outliers) than a normal distribution. This type of distribution is called a platykurtic distribution.**

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



What can we say about the distribution of the data?

**Answer:** **The median is closer to one end of the box, and the box is skewed, it suggests a skewed distribution**

What is nature of skewness of the data?

**Answer:** **The distribution of above boxplot is negative skewed distributed.**

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

**Answer: IQR = Q3 - Q1  
  
IQR = 17-12 = 5**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Answer: The median of both the box plots are same.**

**The data of the first box plots is spread between 250-288 and the second box plot is spread between 200-338.**

**The whisker of boxplot 1 is less than boxplot 2.**

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

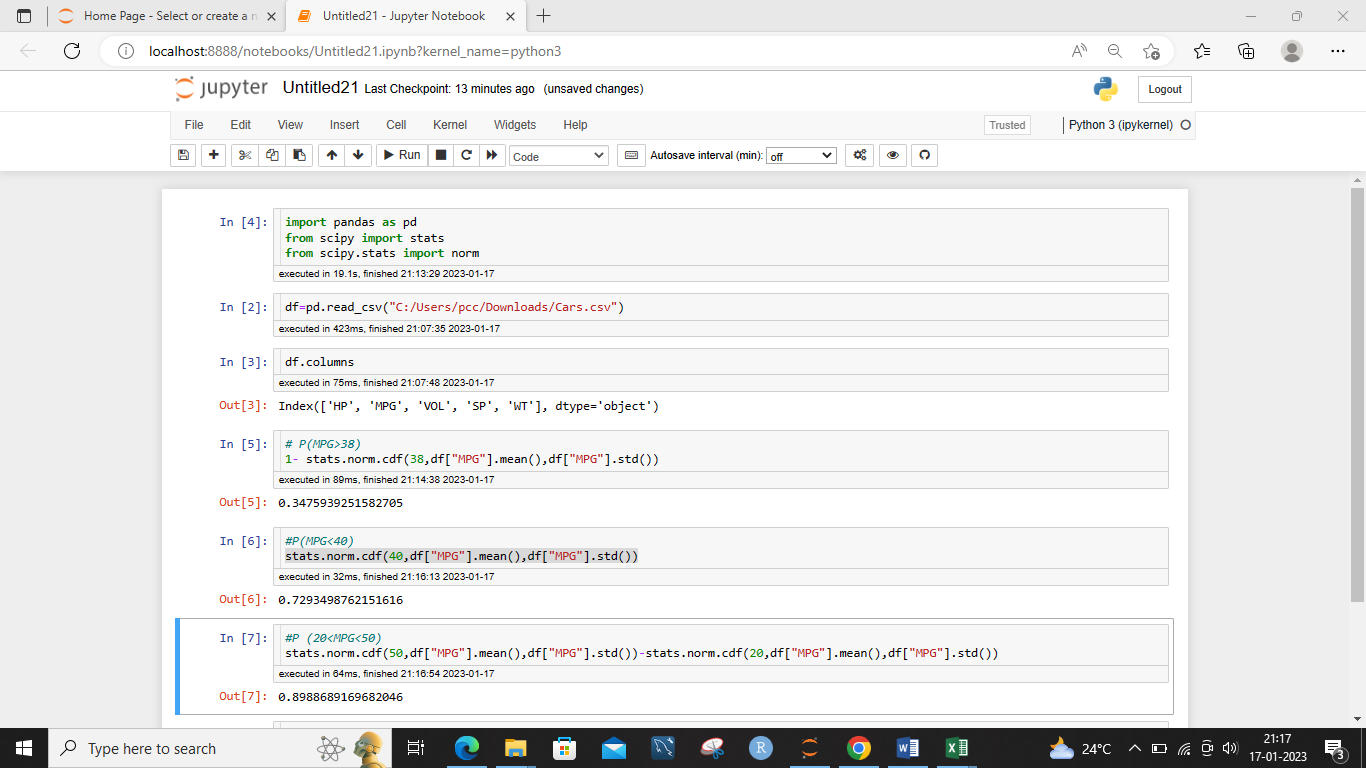
Data \_set: Cars.csv

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

MPG <- Cars$MPG

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

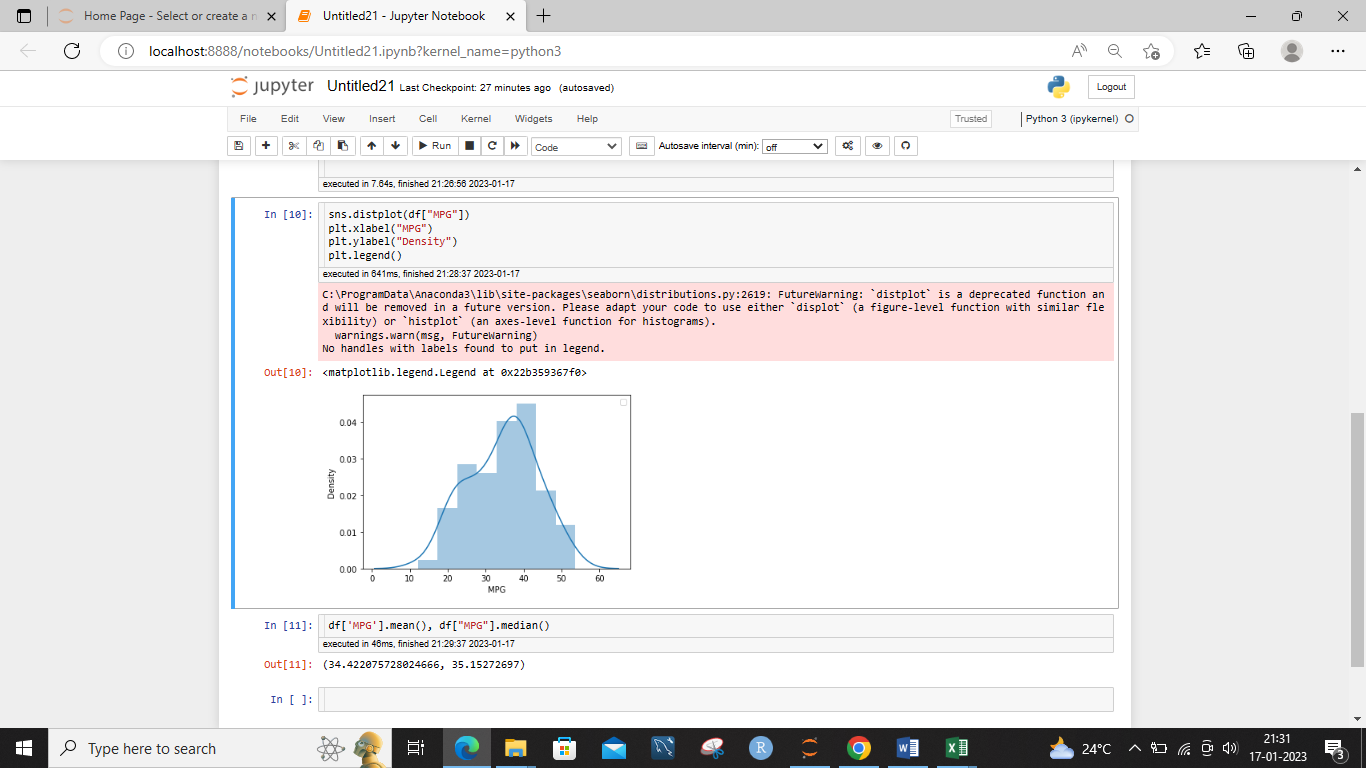
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

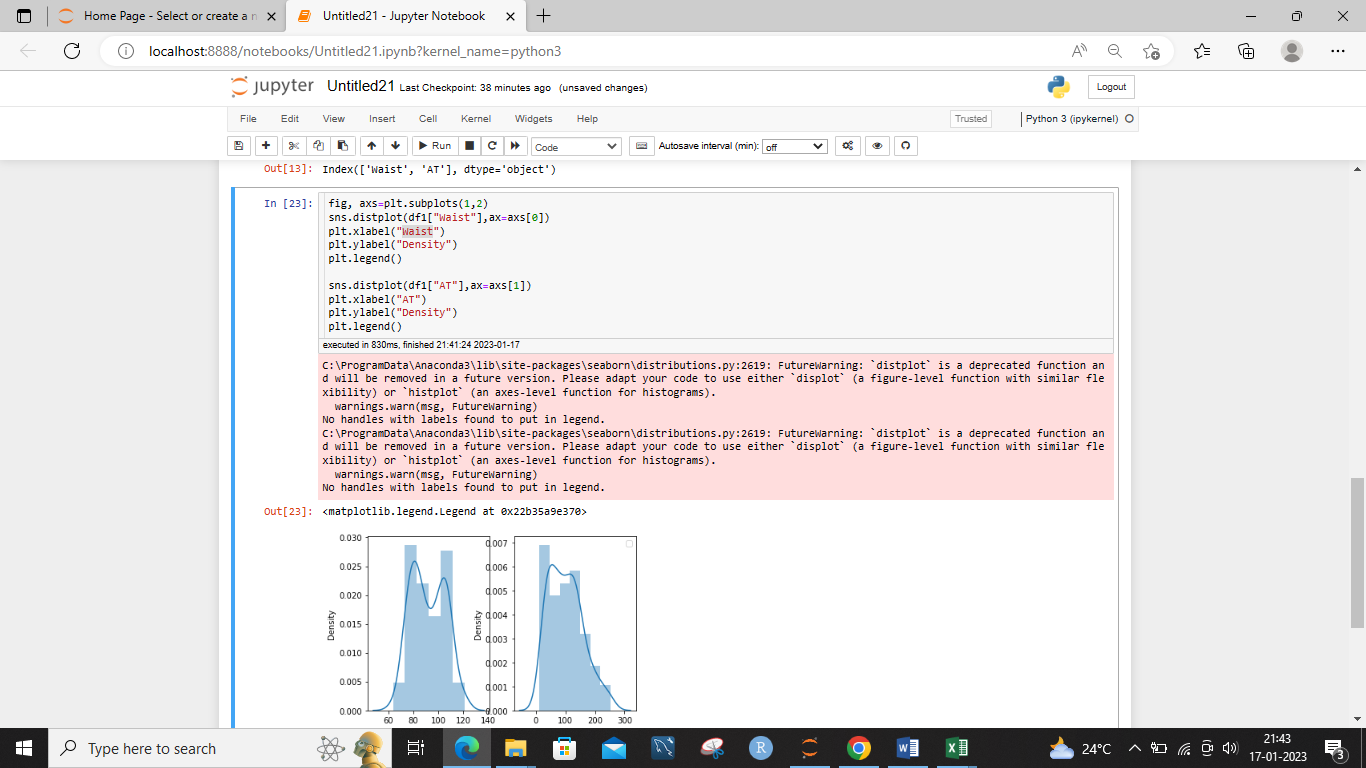
Dataset: Cars.csv

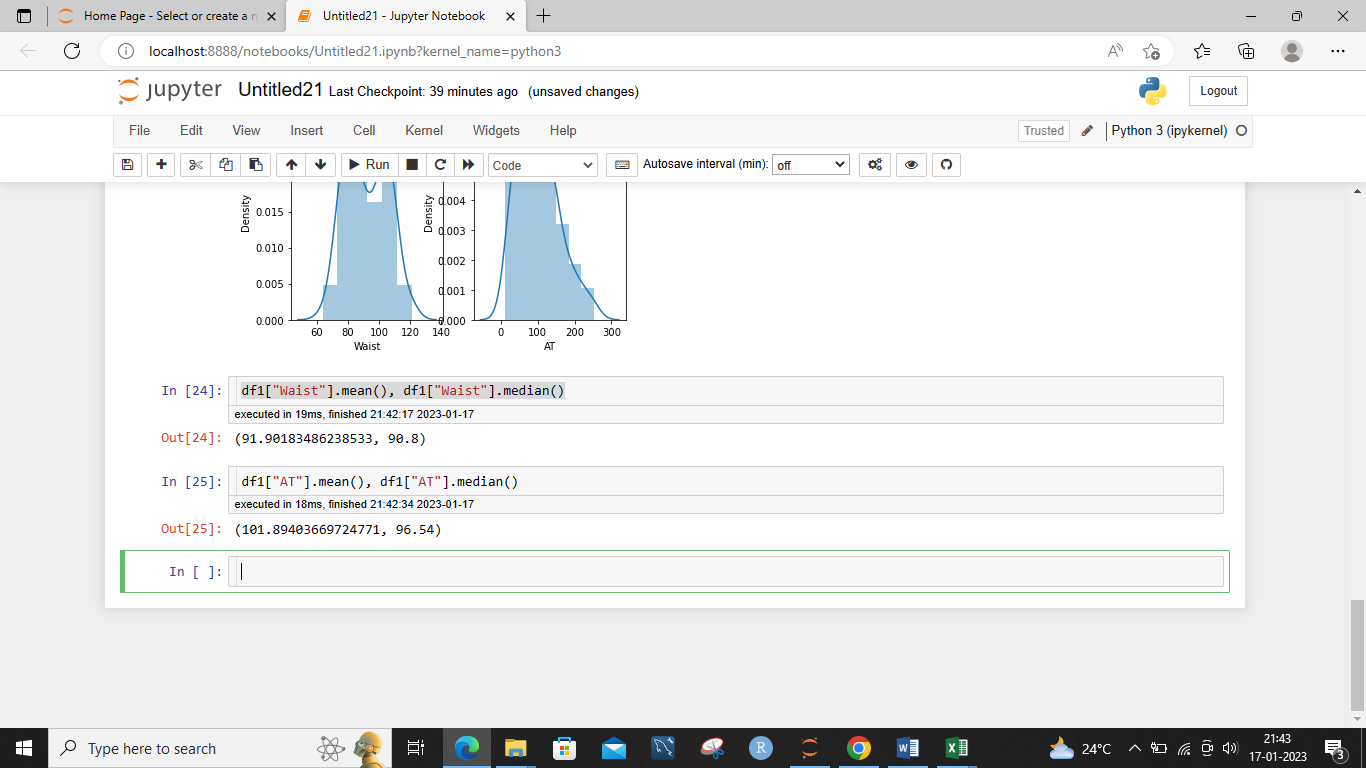


**The MPG of cars is normally distributed.**

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

Dataset: wc-at.csv

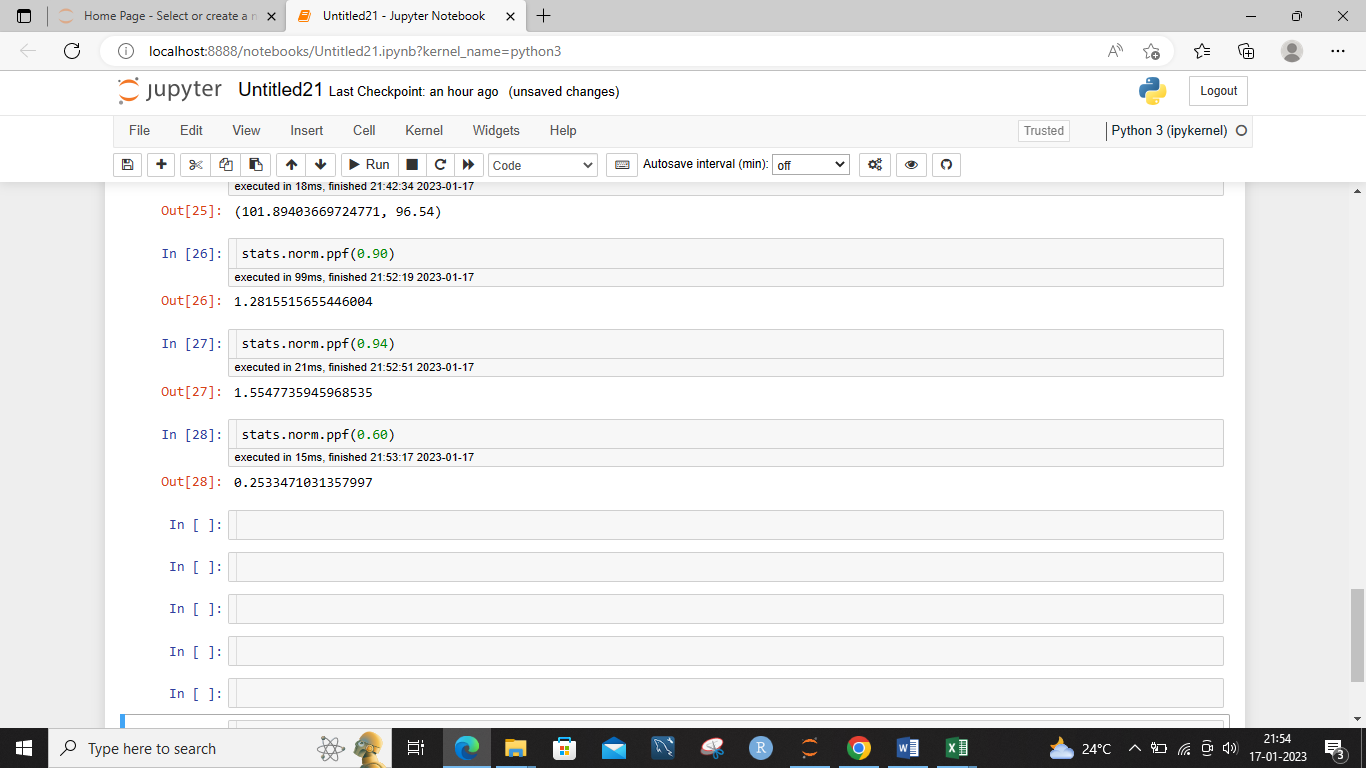




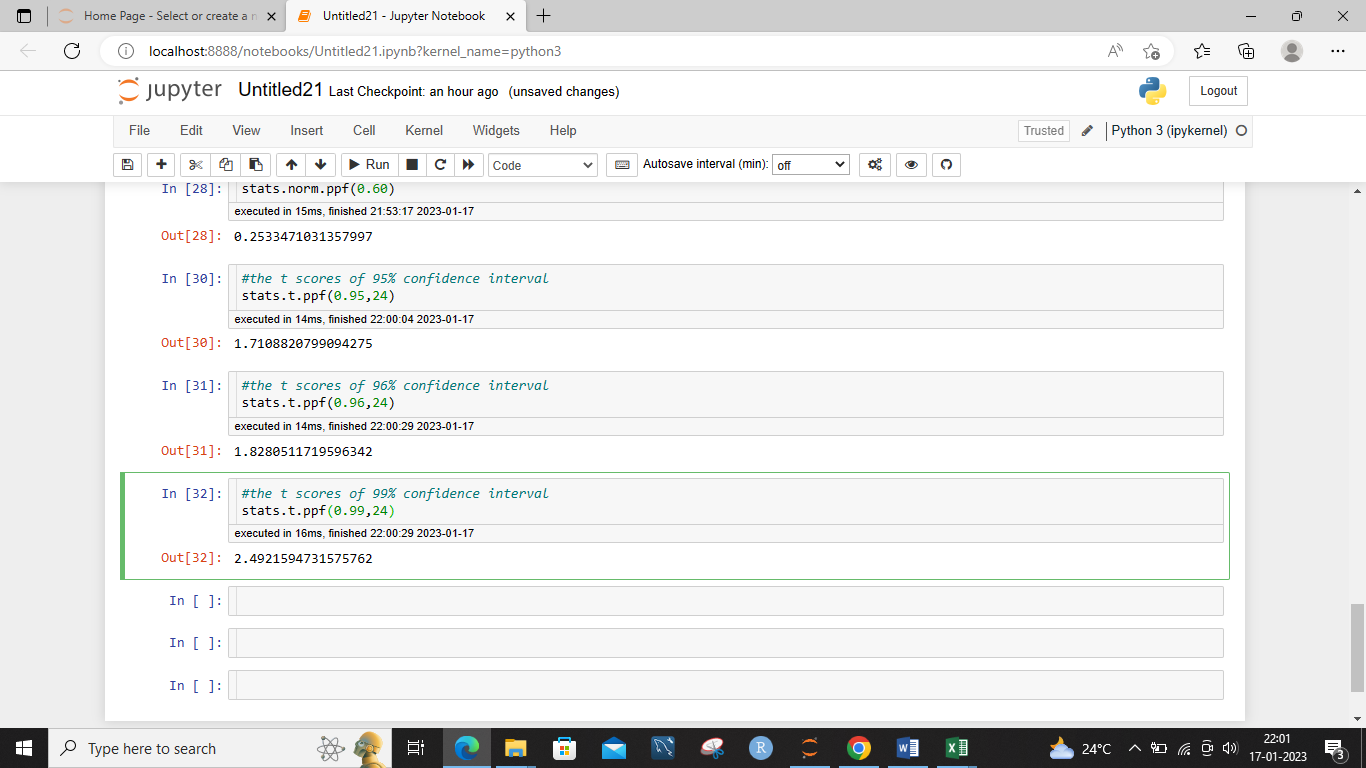
**The AT and Waist are not normally distributed.**

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

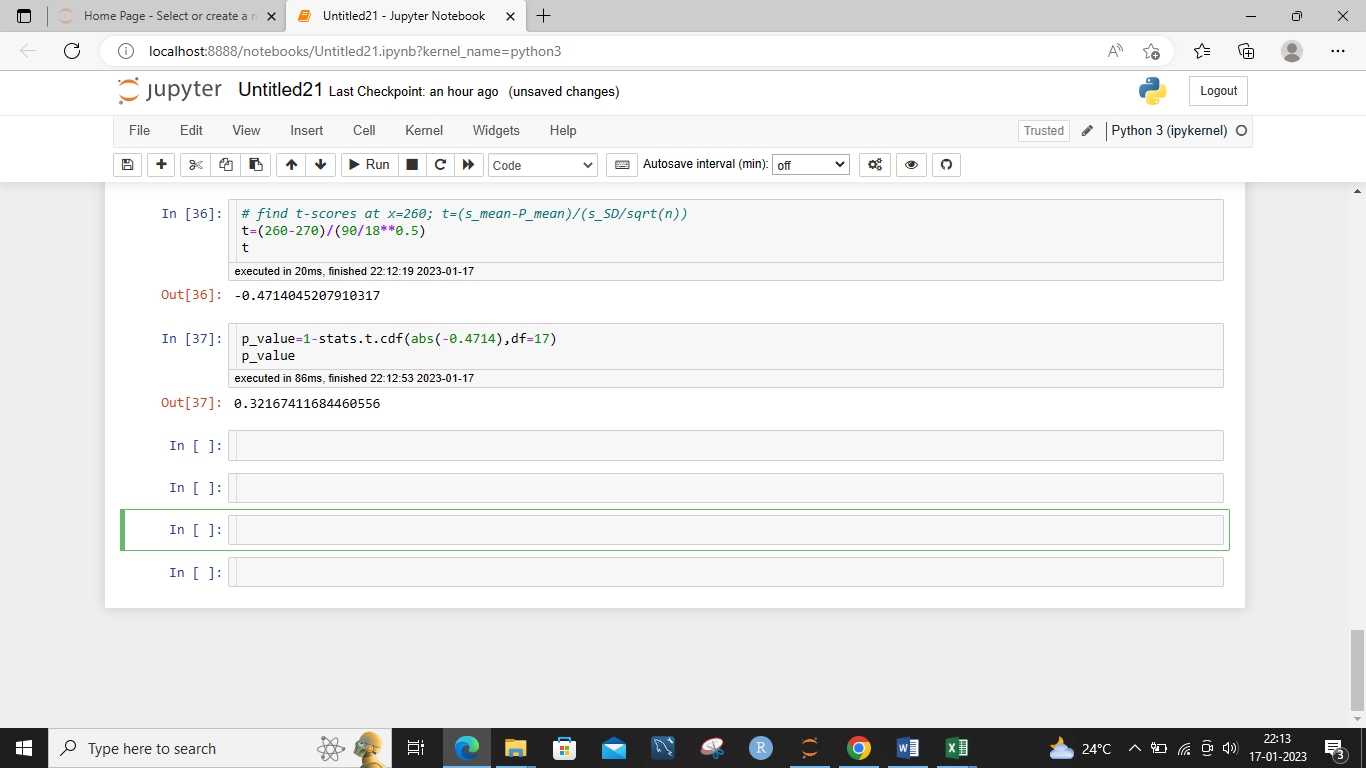


Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days?

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom



**The p-value will tell you the probability that the sample mean of 260 days or lower could have occurred by chance, assuming the population mean is 270 days.**