**Basic statistics level -1**

Q1) Identify the Data type for the Following:

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

Q2) Identify the Data types, which were among the following Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| **Data** | **Data Type** |
| Gender | Nominal data |
| High School Class Ranking | Ordinal data |
| Celsius Temperature | Interval data |
| Weight | Ratio data |
| Hair Color | Nominal data |
| Socioeconomic Status | Ordinal data |
| Fahrenheit Temperature | Interval data |
| Height | Ratio data |
| Type of living accommodation | Nominal data |
| Level of Agreement | Ordinal data |
| IQ (Intelligence Scale) | Ratio data |
| Sales Figures | Interval data |
| Blood Group | Nominal data |
| Time Of Day | Ordinal data |
| Time on a Clock with Hands | Ratio data |
| Number of Children | Ordinal data |
| Religious Preference | Nominal data |
| Barometer Pressure | Ratio data |
| SAT Scores | Ratio data |
| Years of Education | Interval data |

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

**Ans:** When tossing three coins, there are a total of 8 possibilities.

Sample space = {(TTH)(HHH)(THH)(HTT)(HTH)(THT)(HHT)(TTH)}

Probability = favorable outcomes/total outcomes

n(s) = 8

Condition: two heads and one tail.

n(A) = {HHT, HTH, THH}

P(A) = n(A)/n(S) =3/8= 0.375

**The probability that two heads and one tail is 0.375**

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:**

Two dice are rolled,

S= {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6),

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6),

(5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

**n(S)=36**

1. A= Sum is Equal to 1

n(A)=0

P(A)=0

**Probability of get sum of 1 is 0**

1. The possible outcomes that give a sum less than or equal to 4 are:

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

Number = 6

Probability = number of favorable outcomes / total no of outcomes

= 6 / 36

= 1 / 6

**Probability of getting sum is Less than or equal to 4 is 1/6**

1. Sum is divisible by 2 and 3 are 6 and 12

For 6: (1,5), (5,1), (2,4), (4,2), (3,3)

Probability = number of favorable outcomes / total no of outcomes

= 5/36

For 12: (6,6)

Probability = number of favorable outcomes / total no of outcomes

= 1/6

Sum of 2 & 3 = 5/36 +1/6

= 6/36

1/6

**Probability of getting Sum is divisible by 2 and 3 is 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 balls are 2red + 3green + 2 blue = 7 balls

It can be written as 7c2

So, none of the balls drawn is blue = 5c2 or 5 balls from 7

So, 5c2 / 7c2 = 10/21

**The probability that none of the balls drawn is blue**

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 = candies count \* probability

= (1\*0.015) + (4\*0.20) + (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

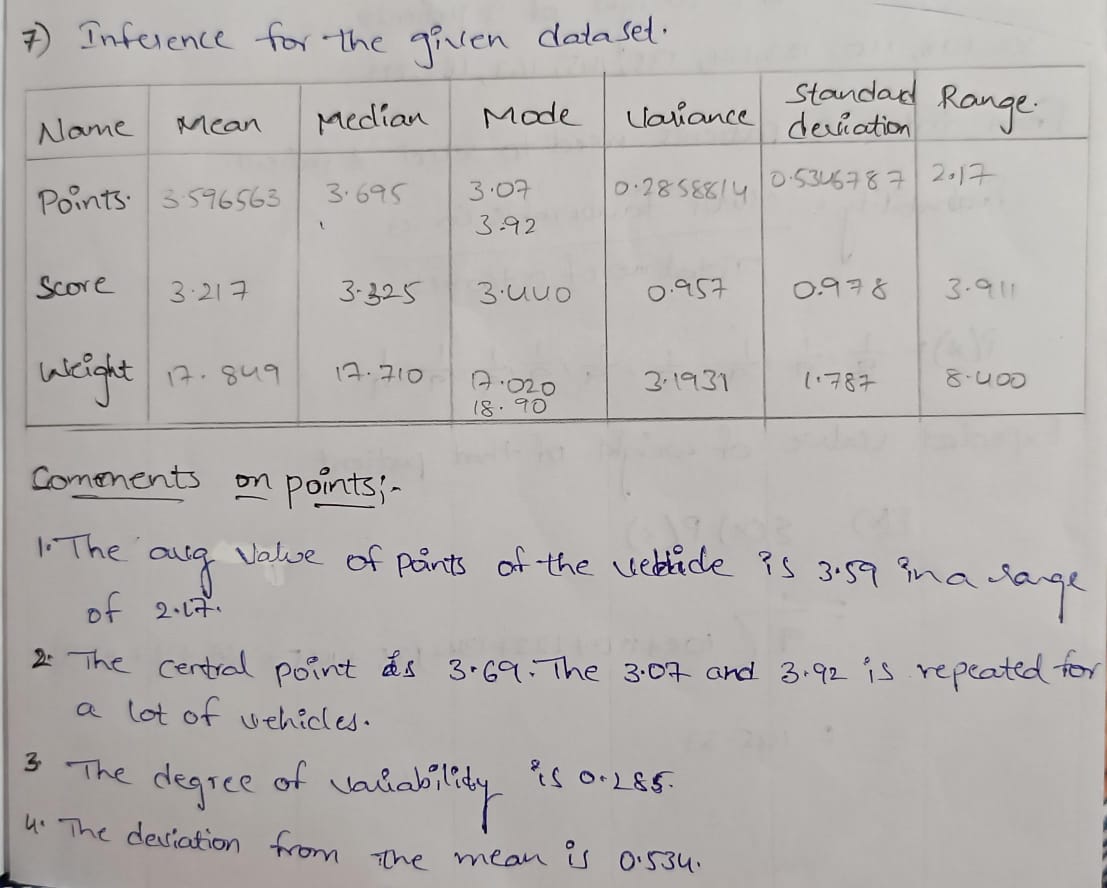
Expected number of candies for a randomly selected child = 3.09

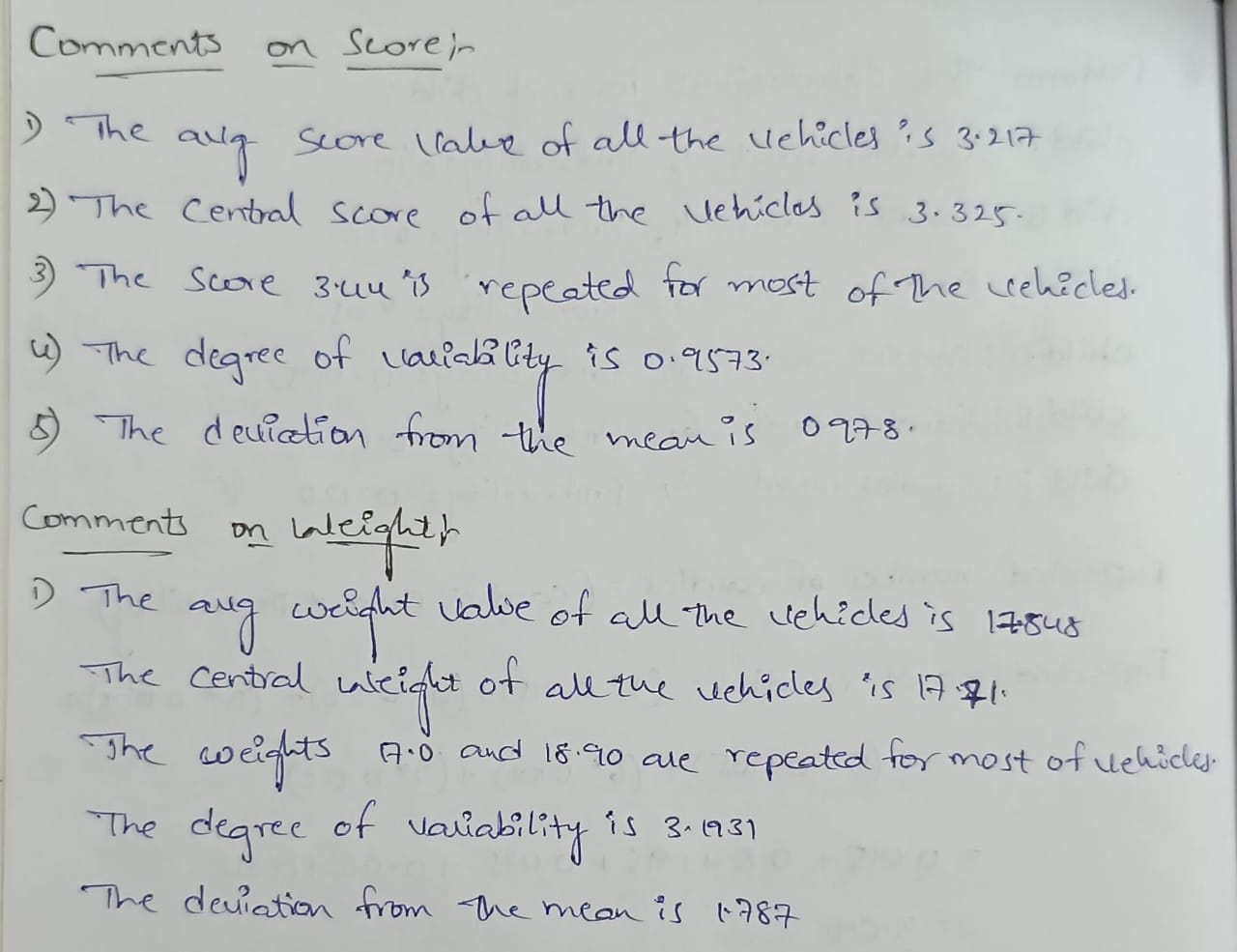
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Ans:**





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:** x represents the weight of a patients.

X = 108, 110, 123, 134, 13, 145, 167, 187, 199

From total 9 patients need to select 1 patient

Probability = number of favorable outcomes / total no of outcomes

= 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9

Expected value of weight of that patient

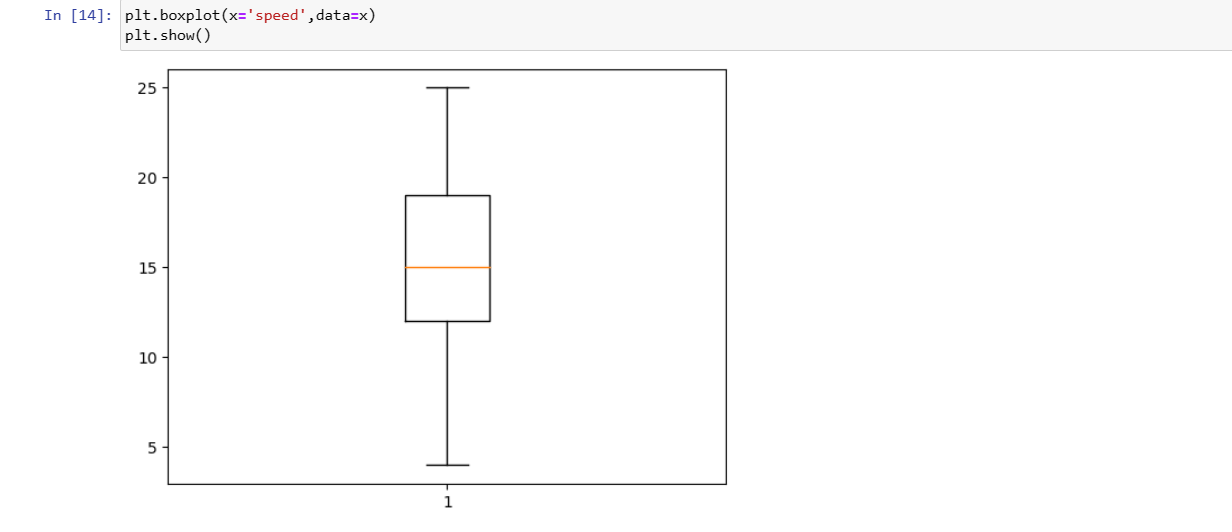
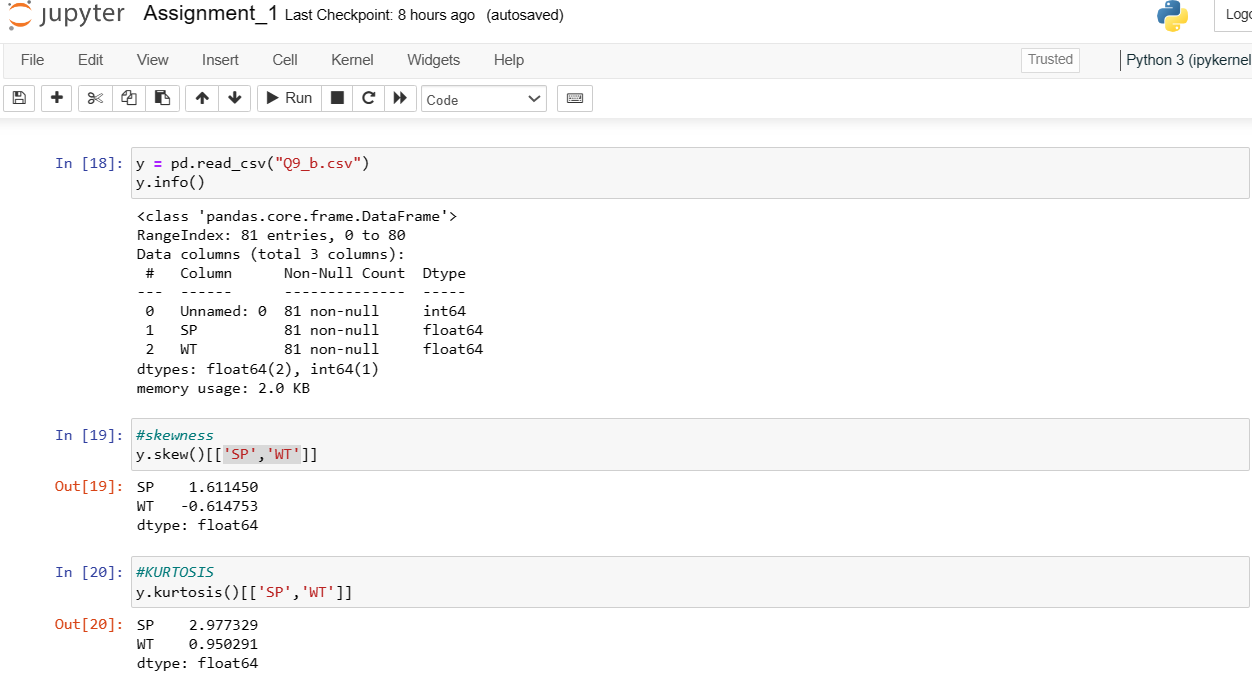
E(x) = ∑(x) P(x)

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

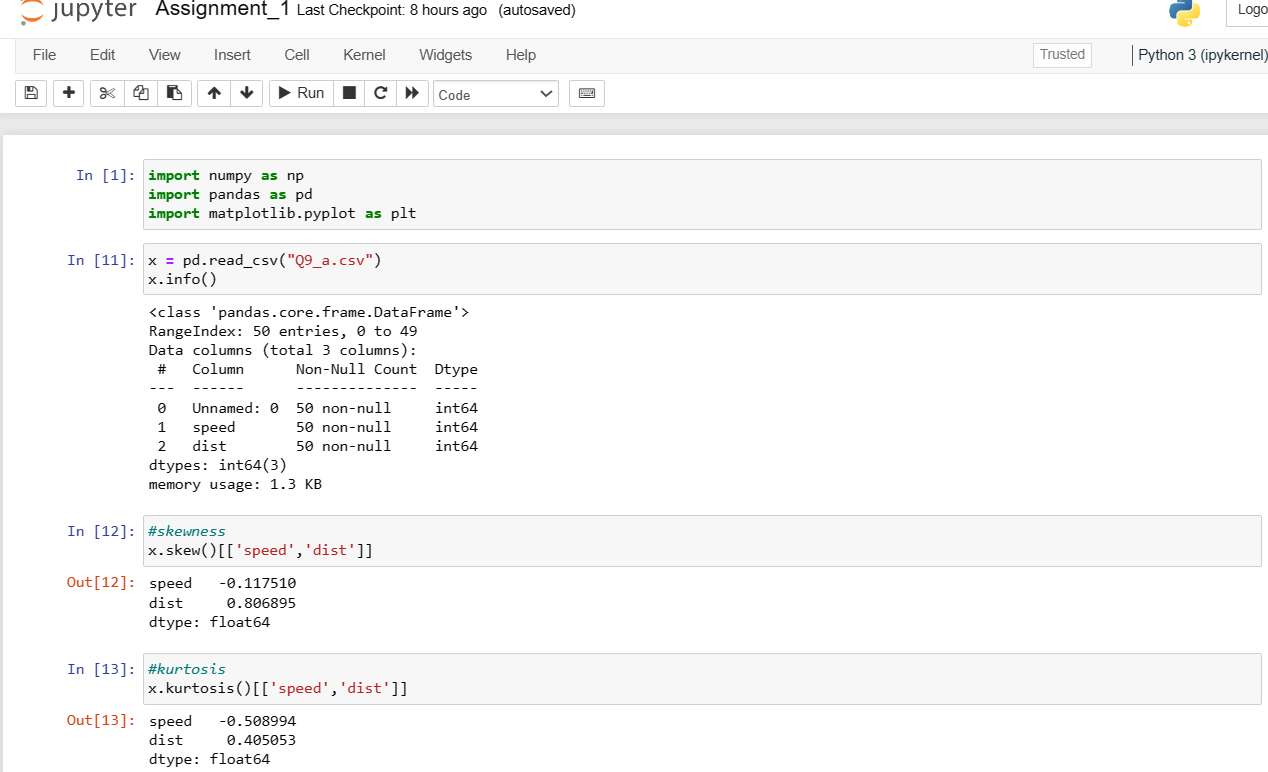
= 145.33

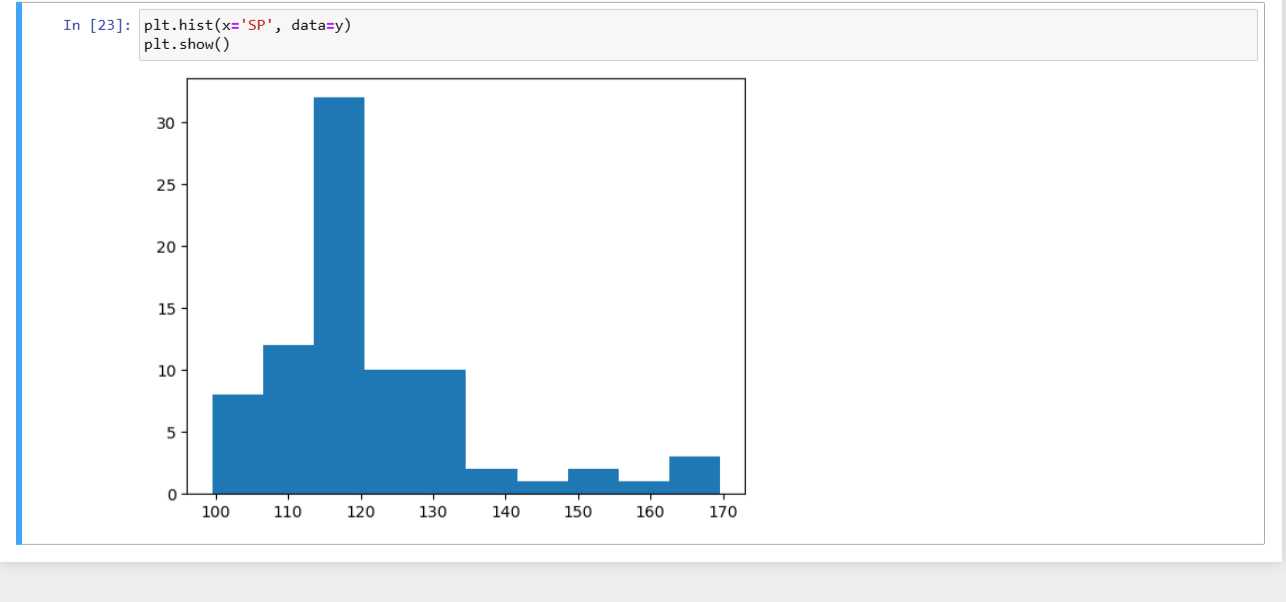
Q9) Calculate Skewness, Kurtosis & draw inferences on the following data

Car’s speed and distance



SP and Weight (WT)





Q10) Draw inferences about the following boxplot & histogram



**Ans:** I) The skewness is right skewed

II) That means the mean > medium

III) This is a positive skewed observation.



**Ans:** From the above box plot the observations are quietly scattered across th dataset. Specially

in the upper extreme and upper quartile. The above box plot tells that plot is left skewed

plot. The mean < median resulting in the negative skewness.

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:** Sample size (n) = 2,000

Sample mean (x̄) = 200 pounds

Sample standard deviation (s) = 30 pounds

Calculate the standard error (SE)

SE = s / √n

SE = 30 / √2000 ≈ 0.6708

Find the critical values (z) for each confidence interval

For 94% confidence interval, z = 1.88 (from z-table)

For 98% confidence interval, z = 2.33 (from z-table)

For 96% confidence interval, z = 2.05 (from z-table)

Calculate the margin of error (ME) for each confidence interval

ME = z \* SE

For 94%: ME = 1.88 \* 0.6708 ≈ 1.2607

For 98%: ME = 2.33 \* 0.6708 ≈ 1.5627

For 96%: ME = 2.05 \* 0.6708 ≈ 1.3751

Calculate the confidence intervals

Confidence interval = x̄ ± ME

For 94%: 200 ± 1.2607 = (198.7393, 201.2607)

For 98%: 200 ± 1.5627 = (198.4373, 201.5627)

For 96%: 200 ± 1.3751 = (198.6249, 201.3751)

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 (Average)\*\*:

To find the mean, sum up all the scores and divide by the total number of scores.

Mean = (34 + 36 + 36 + 38 + 38 + 39 + 39 + 40 + 40 + 41 + 41 + 41 + 41 + 42 + 42 + 45 + 49 + 56) / 18

Mean = 738 / 18

Mean = 41

\*\*Median\*\*:

In this case, after sorting the scores:

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

The median is the middle value (or the average of the two middle values if there's an even number):

Median = (40 + 41) / 2

Median = 40.5

\*\*Variance\*\*:

To calculate the variance, we need to find the average of the squared differences between each score and the mean.

Variance = [(34 - 41) ^2 + (36 - 41) ^2 + (36 - 41) ^2 + (38 - 41) ^2 + (38 - 41) ^2 + (39 - 41) ^2 + (39 - 41) ^2 + (40 - 41) ^2 + (40 - 41) ^2 + (41 - 41) ^2 + (41 - 41) ^2 + (41 - 41) ^2 + (41 - 41) ^2 + (42 - 41) ^2 + (42 - 41) ^2 + (45 - 41) ^2 + (49 - 41) ^2 + (56 - 41) ^2 )] / 18

Calculating this step by step for each score:

Variance = (49 + 25 + 25 + 9 + 9 + 4 + 4 + 1 + 1 + 0 + 0 + 0 + 0 + 1 +1 + 16 + 64 + 225) / 18

Variance = 24.1111111111

\*\*Standard Deviation\*\*:

The standard deviation is the square root of the variance.

Standard Deviation = √24.1111111111

Standard Deviation = 4.91030

2)

The mean (average) score is approximately 41.

The median score is 40.5.

The variance is approximately 24.1111111111

The standard deviation is approximately 4.91030.

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

**Ans:** Zero

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

**Ans:** positively skewed

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

**Ans:** Negatively skewed

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

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

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

**Ans:** Negative skew refers to a longer or fatter tail on the left side of the distribution.

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



What can we say about the distribution of the data?

**Ans:** The distribution is concentrated towards the right and is not normally distributed.

What is nature of skewness of the data?

**Ans:** The nature of the skew is left and is negative skewed.

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

**Ans:** Q3-Q1 = 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:**

1. Both the boxplots are normal distributions.

2. Medians of both 1 and 2 boxplots are at the same level. It means that the data has the

same median (260).

3. The dispersion of the data is more in the boxplot 2 than the boxplot 1. 4. The

interquartile range of boxplot 2 is more than the boxplot 1 shows the distribution.

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)

**Ans:** values are found by the formula stats.norm.cdf (x=38, mean=34, SD = 434.42,9.13) = 34.74%

* 1. P(MPG<40)

**Ans:** values are found by the formula stats.norm.cdf (x=38, mean=34, SD = 434.42,9.13) = 72.94%

c. P (20<MPG<50)

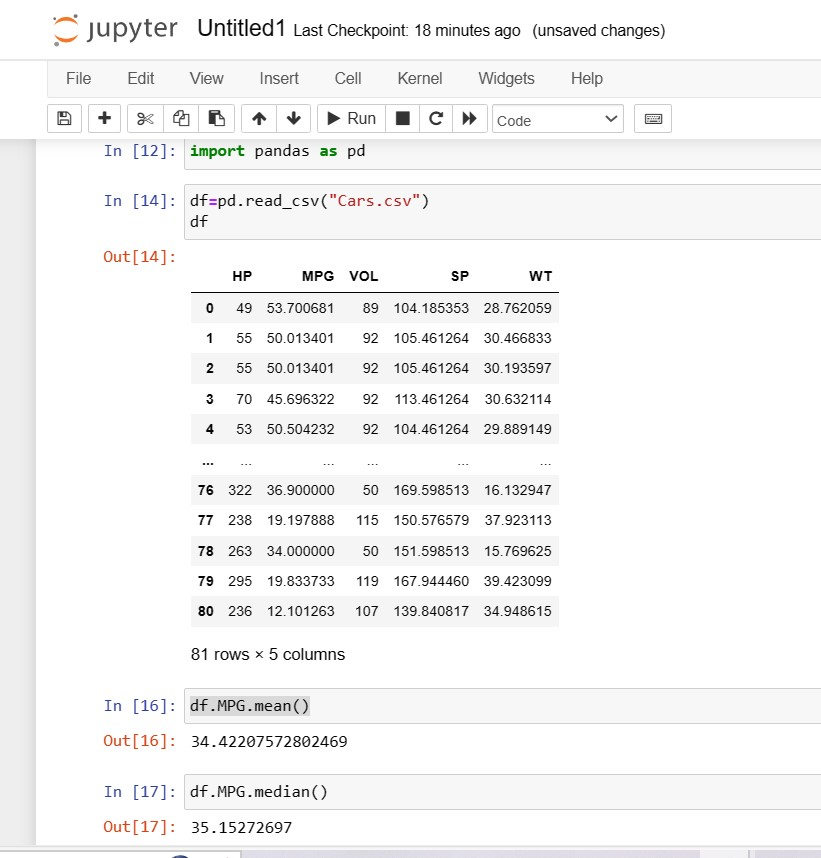
**Ans:** Dividing it into (MPG 20) = stats.norm.cdf (x=38, mean=34, SD = 434.42,9.13) – 1 + stats.norm.cdf(x=38, mean=34, SD = 434.42,9.13) = 0.0129 = 1.3%

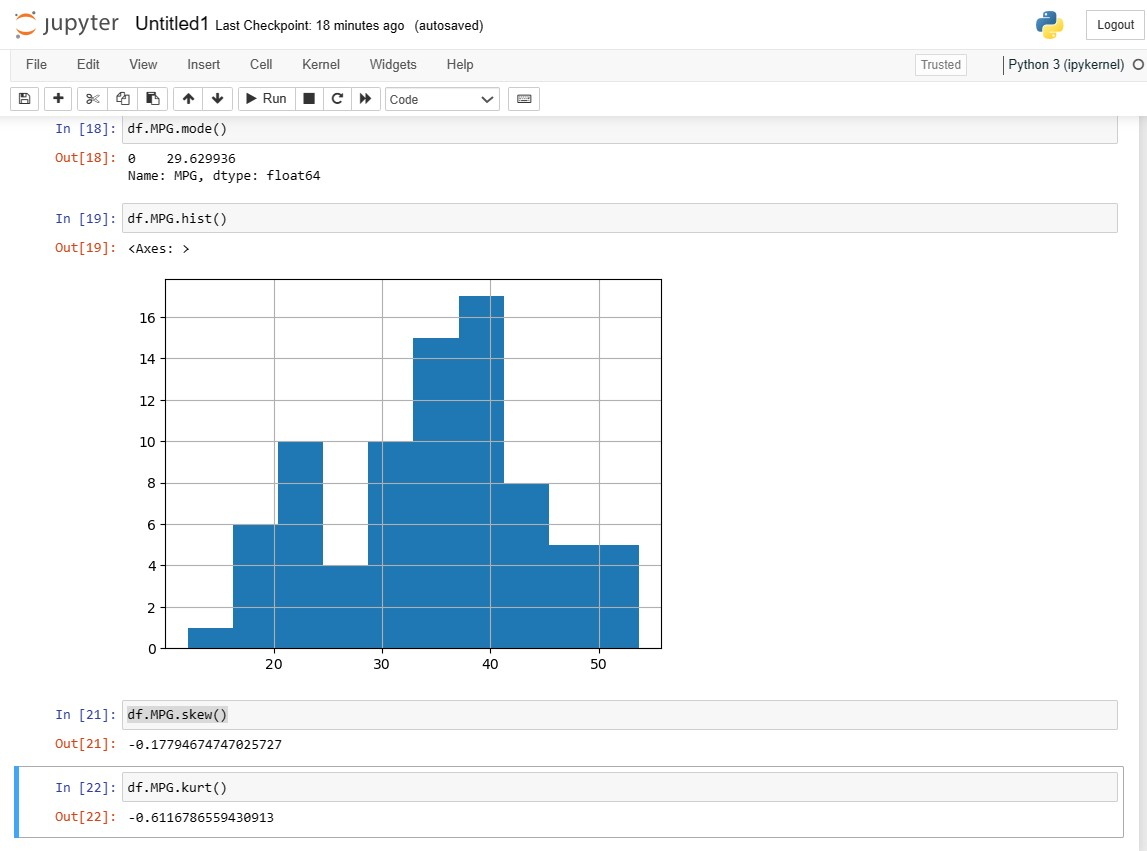
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:**

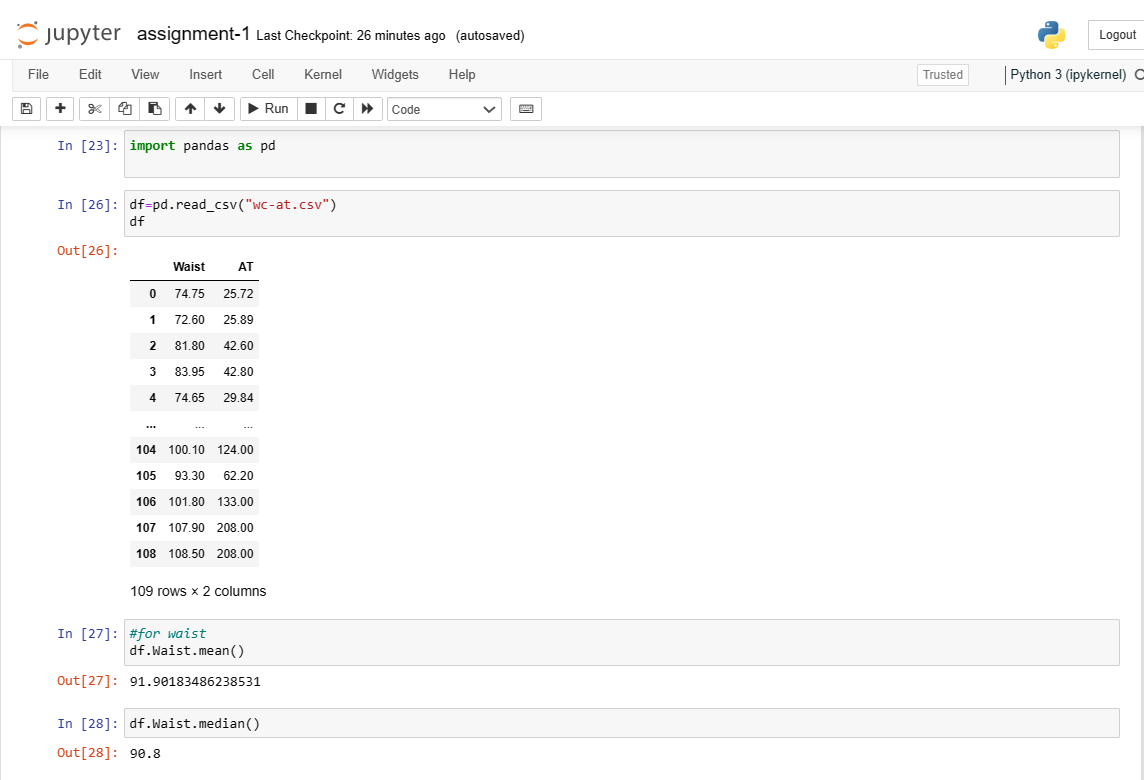


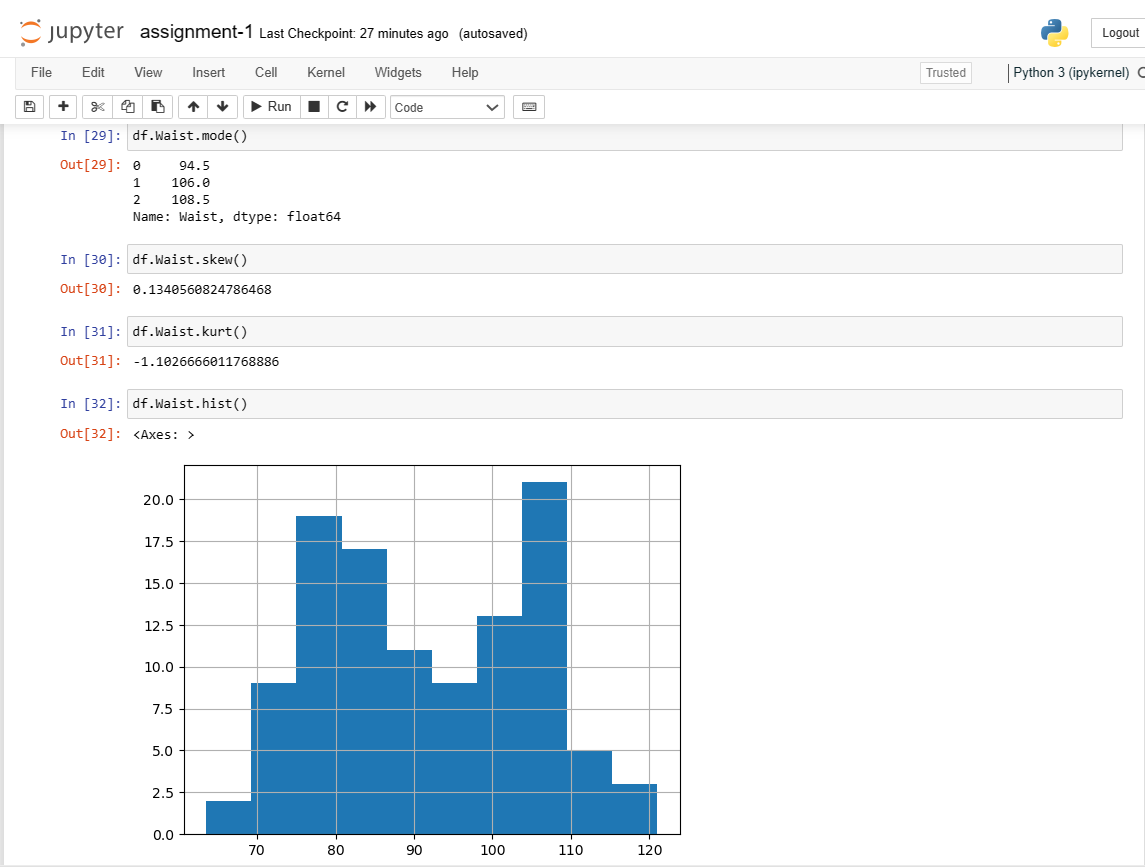


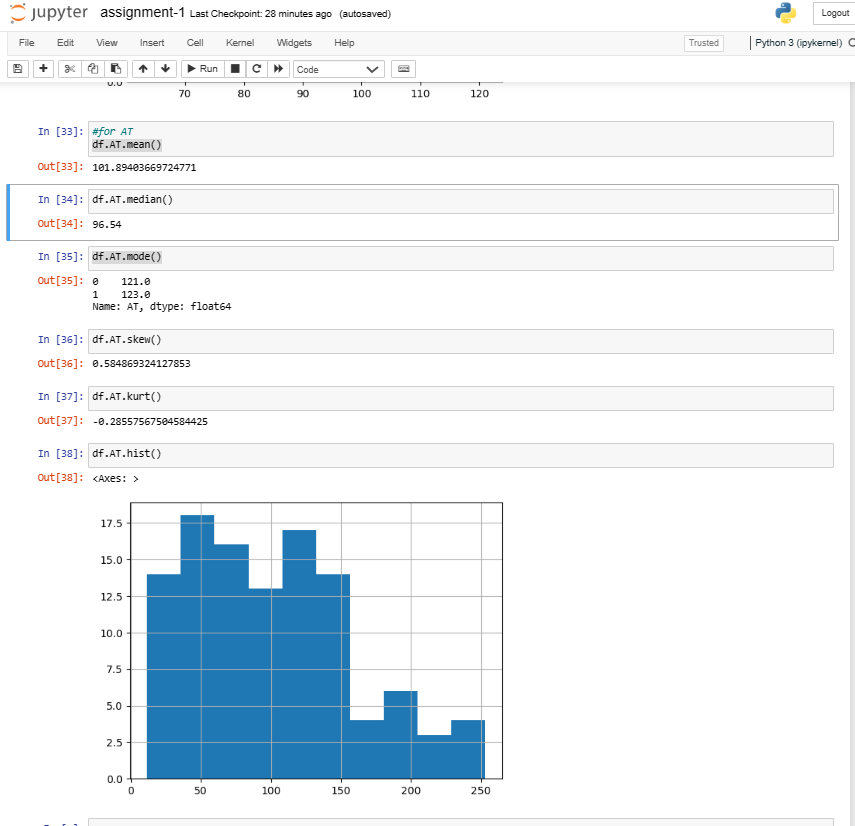
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

**Ans:**

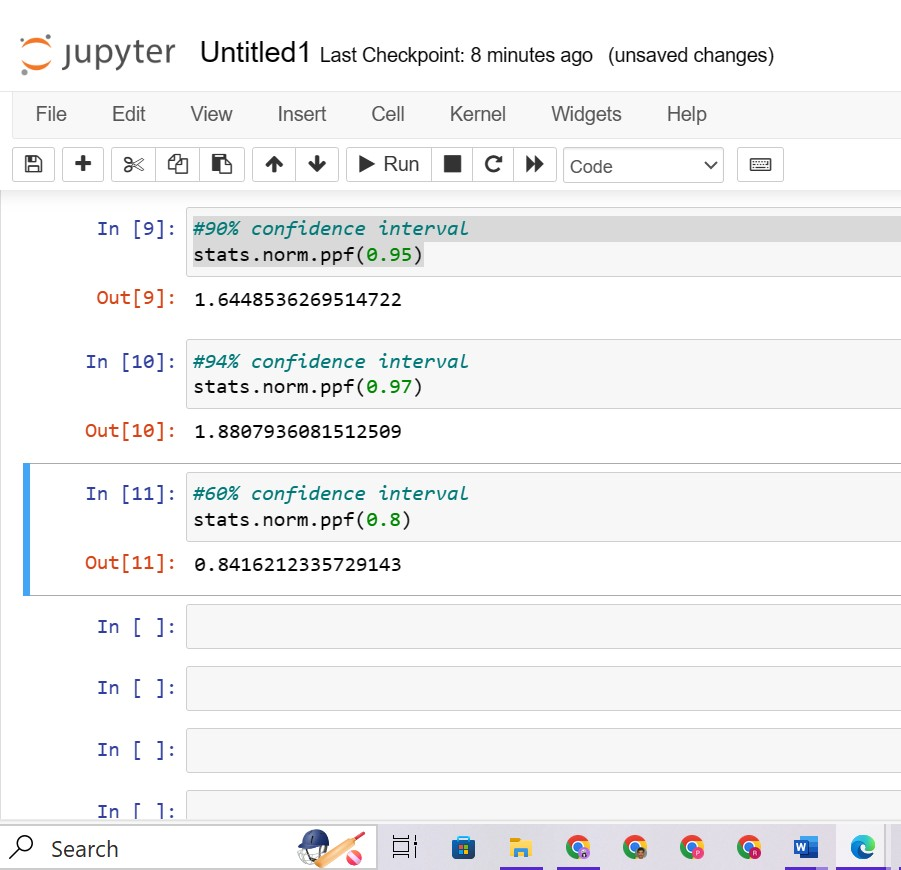






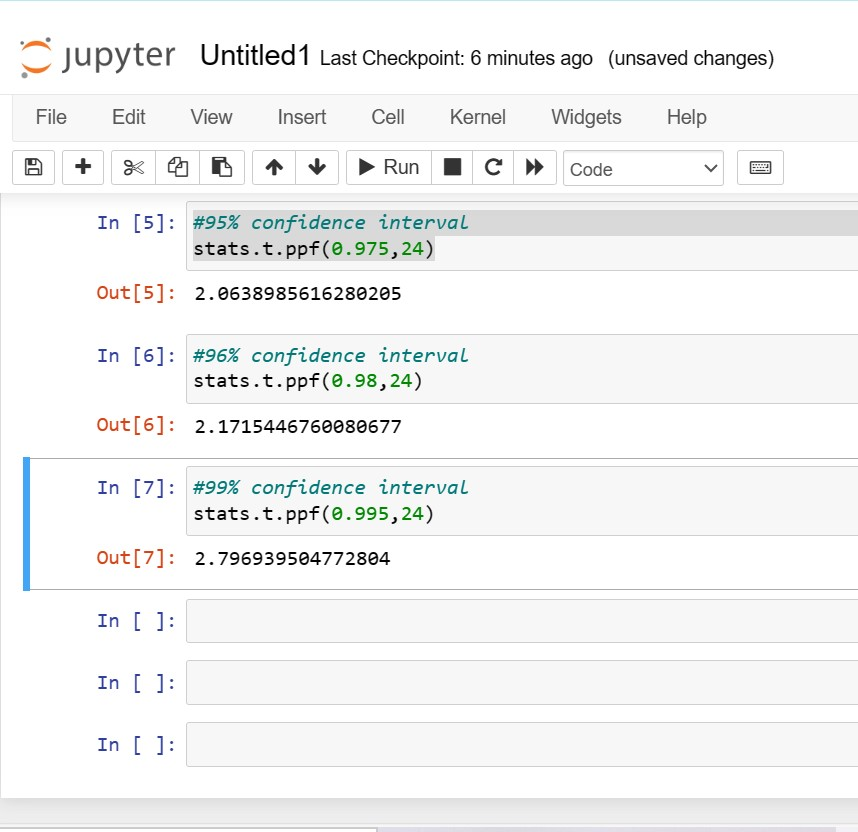
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**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:**

µ = 270

n = 18

x̄ = 260

s = 90

The standard error (SE) is calculated as follows:

SE = s / √n

SE = 90 / √18

SE ≈ 21.21

The z-score is calculated as follows:

z = (x̄ - µ) / SE

z = (260 - 270) / 21.21

z ≈ -0.47

we need to find the probability that the sample mean is no more than 260 days if the true mean is 270 days.

P (Z ≤ -0.47) ≈ 0.3216

So, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the true mean is 270 days is approximately 0.3216 or 32.16%.

