|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| 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 | Ordinal |

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

Ans: If three coins are tossed the no of possible outcomes are

{HHH}{HHT}{HTH}{HTT}{TTT}{TTH}{THT}{THH}

From the above-mentioned outcomes, the probability of getting two heads and one tail is 3/8 or 0.375

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

Ans: If two dice are rolled then the total outcomes are 6\*6=36

1. Equal to 1

Ans: As the minimum sum is 2 for outcome (1,1) there is no possibility for the that sum is equal to 1, Hence probability that sum is equal to 1 is “0”.

1. Less than or equal to 4
2. Probability of getting sum 1 is “0”
3. Probability of getting number 2 is (1,1), so the no of favorable outcome is 1 thus 1/36 is the probability of rolling two dice and getting sum 2
4. Probability of getting number 3 is (1,2), (2,1) so the no of favorable outcome is 2 thus 2/36=1/18 is the probability of rolling two dice and getting sum 3
5. Probability of getting number 4 is (1,3), (2,2), (3,1) so the no of favorable outcome is 3 thus 3/36=1/12 is the probability of rolling two dice and getting sum 4
6. Sum is divisible by 2 and 3

Ans: Probability of sum which is divisible by 2 and 3 is (1,5), (2,4), (3,3), (4,2), (5,1) so the no of favorable outcome is 5 thus 5/36 is the probability of rolling two dice and getting the sum is divisible by 2 and 3

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 no of balls from the given data are 2+3+2=7

No of ways of drawing 2 balls out of 7 balls are 7C2 = (7\*6)/ (2\*1) = 42/2 = 21

No of balls other than blue are 5 balls, so the no of ways of drawing 2 balls out of 5 are 5C2 = (5\*4)/ (2\*1) =20/2=10

Therefore, the probability that none of the balls drawn is blue are 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: from the above data expected number of candies for a randomly selected child = 1\*0.015+4\*0.20+3\*0.65+5\*0.005\*6\*0.01+2\*0.120 = 3.090

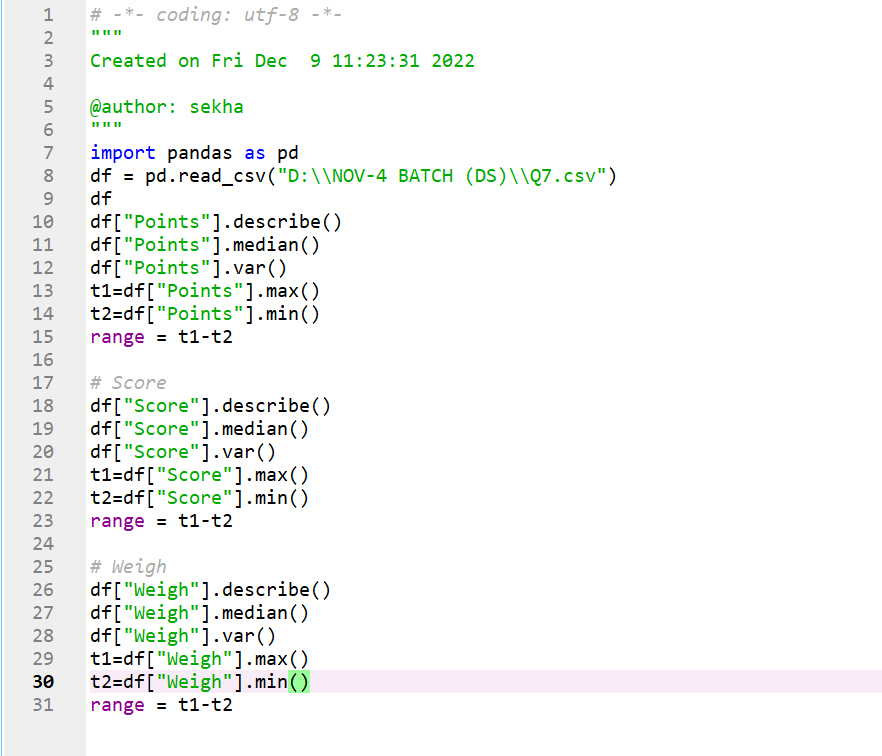
the Expected number of candies for a randomly selected child = 3.09

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

* For Points, Score, Weigh>

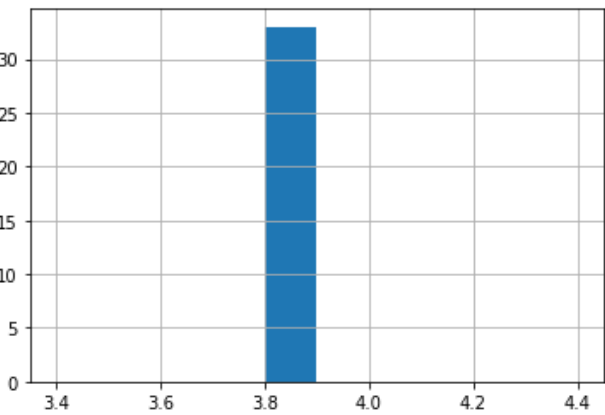
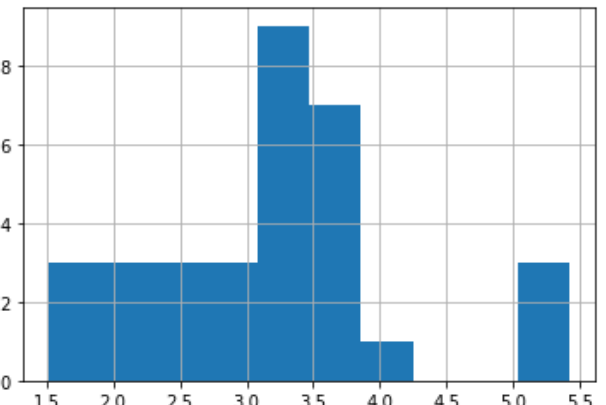
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

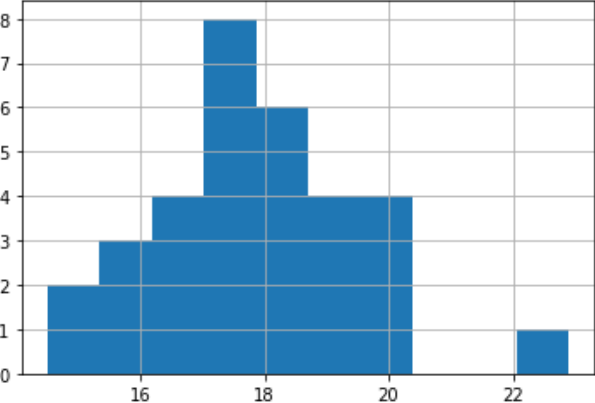
**Use Q7.csv file**

Ans: 

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | weigh |
| Mean | 3.90 | 3.21 | 17.84 |
| Median | 3.9 | 3.32 | 17.71 |
| Mode | 3.9 | 3.44 | 18.90 |
| variance | 5.08 | 0.95 | 3.19 |
| Standard deviation | 2.25 | 0.97 | 1.78 |
| Range | 0 | 3.91 | 8.39 |

Inferences: Histograms

Points: Score:

Weigh:

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: There are 9 patients so the probability of selecting each patient P(X)=1/9Weight of patients at a clinic are E(X)= 108,110,123,134,135,145,167,187,199

Expected value = P(X). E(X)

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

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

= (1/9)(1308)

= 145.3

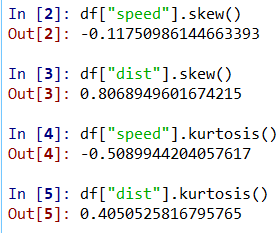
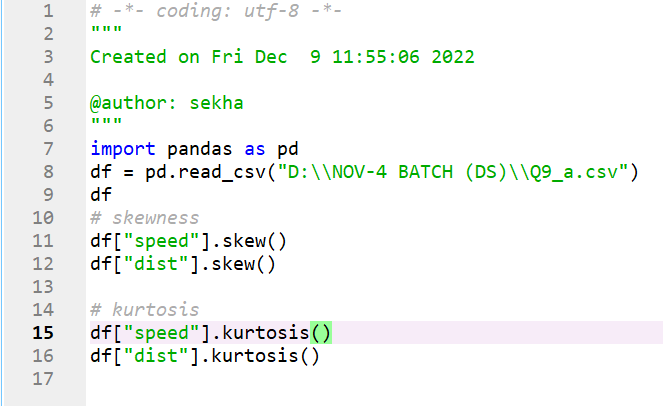
Expected Value of the Weight of that patient is 145.3

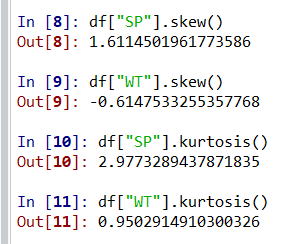
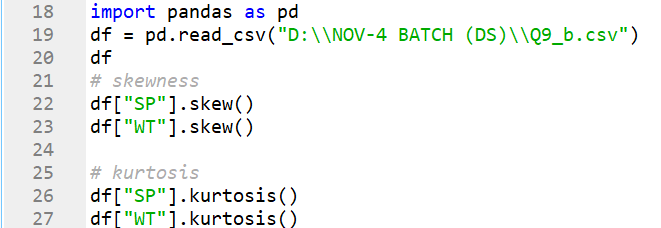
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Use Q9\_a.csv**

**Cars speed and distance**

Ans**: output**



**SP and Weight (WT) output**

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

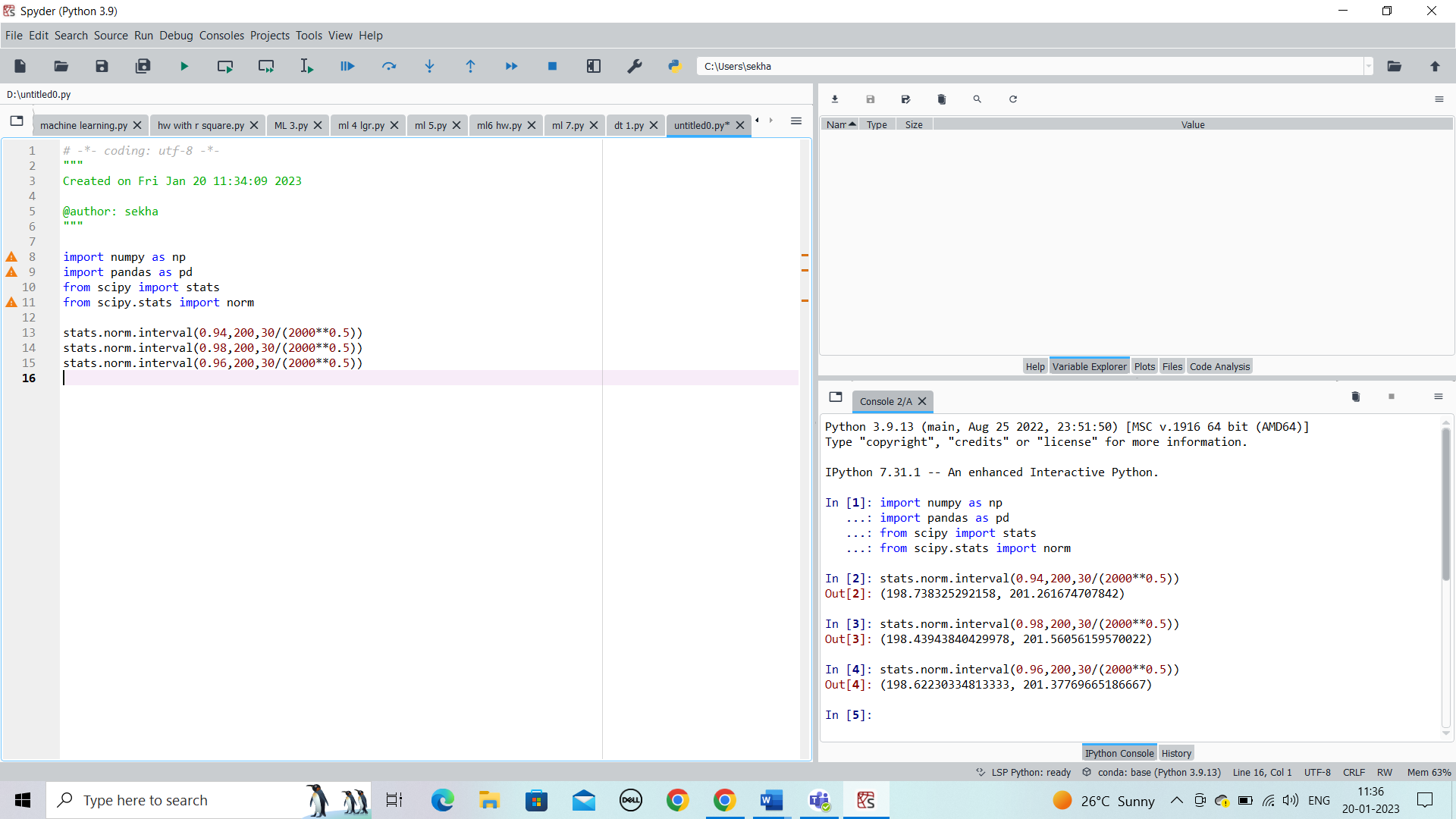


Ans: From the above histogram the peak is on the right side, so it is a positive skewness which means mean > median. We have outliers on higher side



Ans: From the above boxplot the outliers on the maximum side which means on the upper 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?

Ans: 

94% confidence interval is (198.73, 201.26)

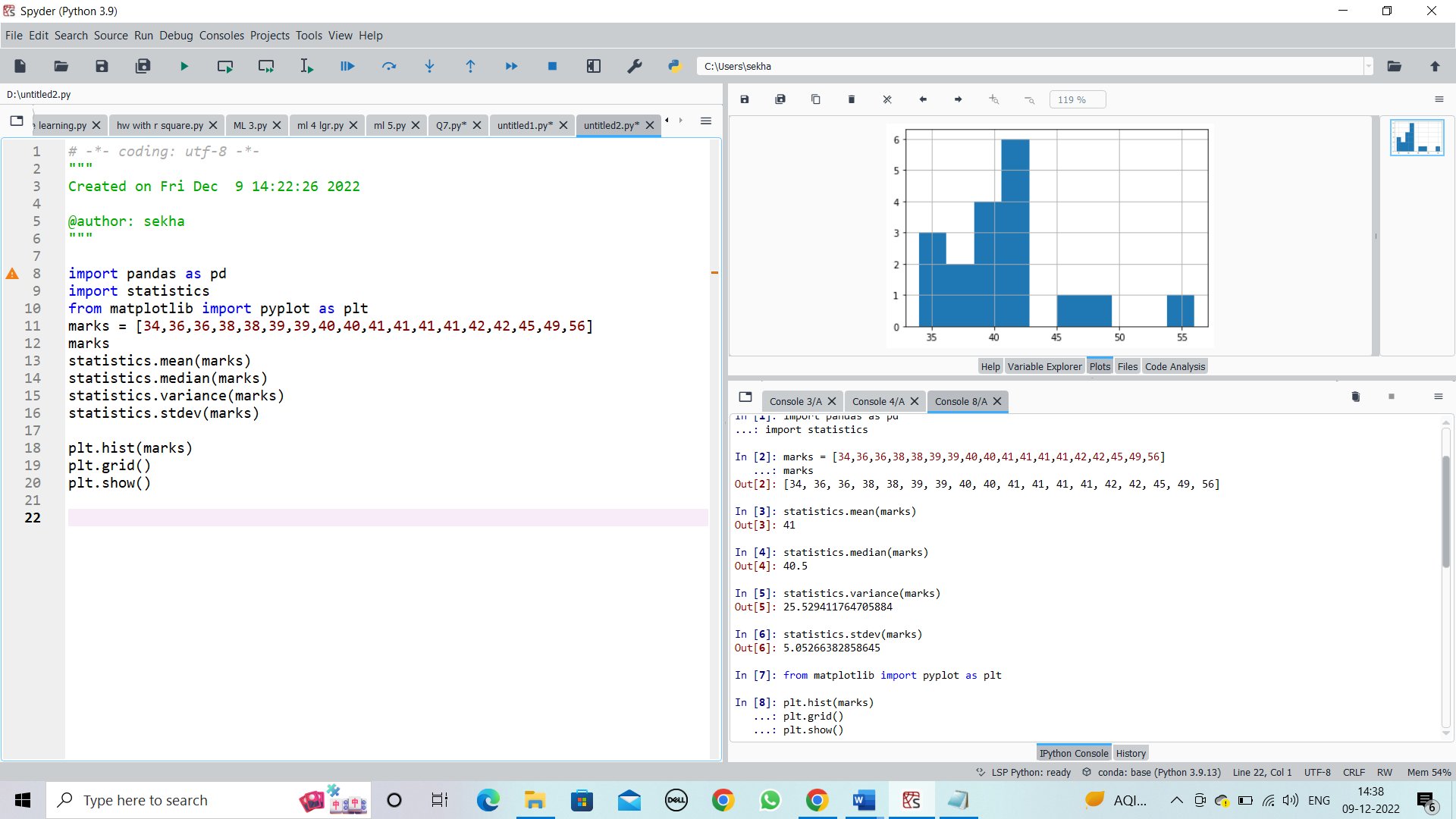
96% confidence interval is (198.43, 201.56)

98% confidence interval is (198.62, 201.37)

**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.

Ans:

\*Mean = 41, Median = 40.5, Variance = 25.52, standard deviation = 5.05

1. What can we say about the student marks?

Ans: from the above histogram most of the students have scored marks between 41-43

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

Ans: There will be no skewness ,it will be symmetric

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

Ans: Positive skewness

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

Ans: Negative skewness

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

Ans: Positive kurtosis means the curve is more peaked

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

Ans: Negative Kurtosis means the curve will be flatter

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



What can we say about the distribution of the data?

Ans: from the above boxplot it is not normally distributed the median is towards the higher value which is on the right side

What is nature of skewness of the data?

Ans: The data is on left side, whisker range of minimum value is more than the maximum, therefore it is negative skewness.

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

Ans: As we know IQR = Q3 upper quartile – Q1 lower quartile = 18-18= “8”

Q19) Comment on the below Boxplot visualizations?



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

Ans: From the above boxplots both are having zero errors and both the boxplots median is also in the same range and they are normally distributed.

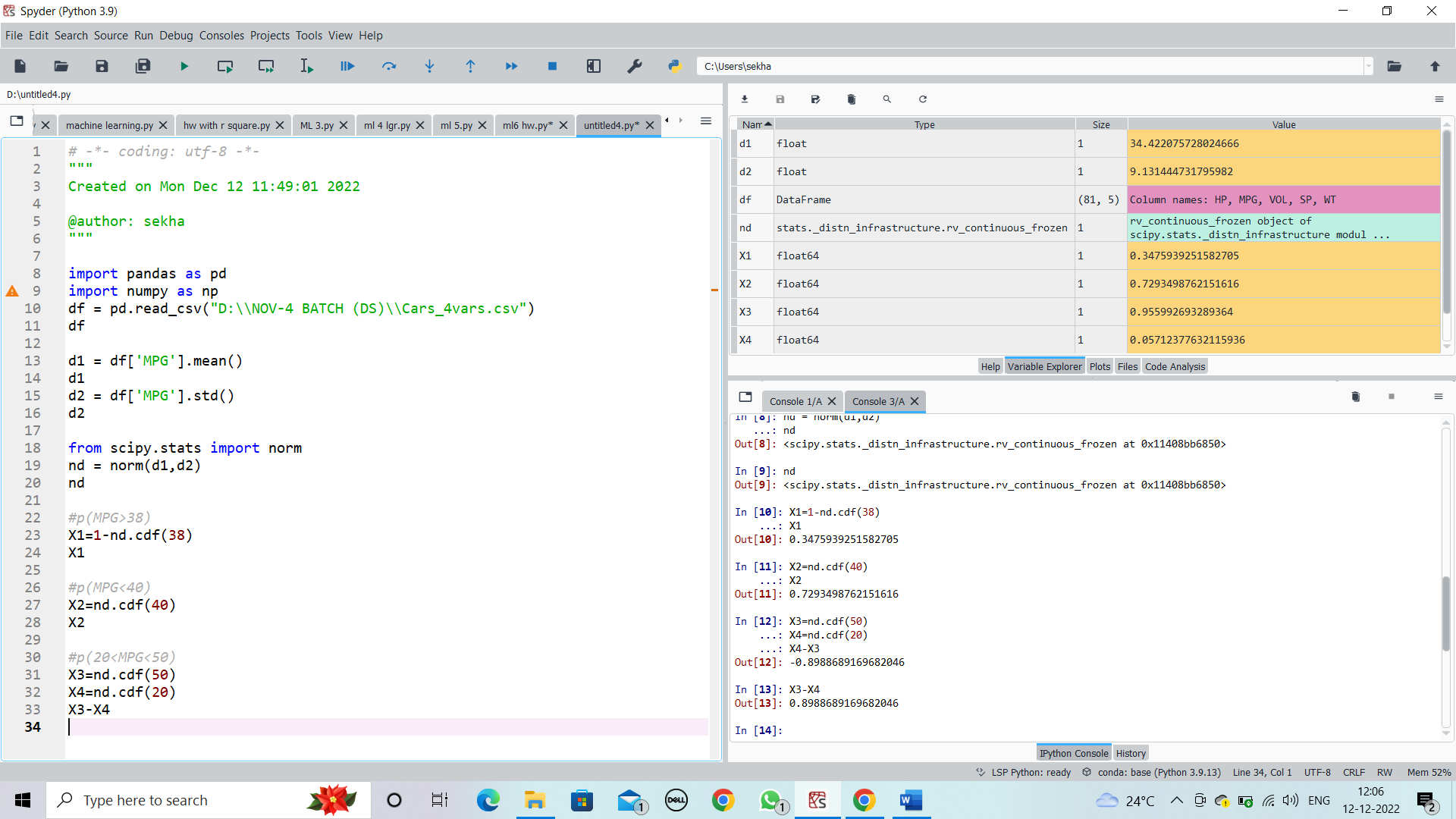
Q 20) Calculate probability from the given dataset for the below cases

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)
  3. P (20<MPG<50)

Ans : 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: The data 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

Ans: Adipose tissue and waist circumference does not follow normal distribution

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

Ans Z scores of 90% confidence interval is 1.64

Z scores of 94% confidence interval is 1.88

Z scores of 60% confidence interval is 0.84

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

Ans: t scores of 95% confidence interval is 2.06

t scores of 96% confidence interval is 2.17

t scores of 99% confidence interval is 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: The p value is 0.52