|  |  |
| --- | --- |
| 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 | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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?

* Total combinations = 8 (23)
* The combinations of getting 2 heads and 1 tail = 3 (HHT, THH, HTH)
* The probability that two heads and one tail are = 3/8

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

* Total combinations = 36 (62)
* Probability of the sum is equal to 1 = 0 (0/36)
* Probability of the sum is less than or equal to 4 = 6/36 => 1/6 [(1,1), (1,2),(1,3),(2,1),(2,2),(3,1)]
* Probability of the sum is divisible by 2 and 3 = 6/36=>1/6 [(1,5),(2,4),(3,3),(4,2)(5,1)(6,6)]

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 in a bag = 7
* Two balls are drawn at random = 7C2 = > 21
* none of the balls drawn is blue = 5C2  => 10
* Probability that none of the balls drawn is blue = 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)

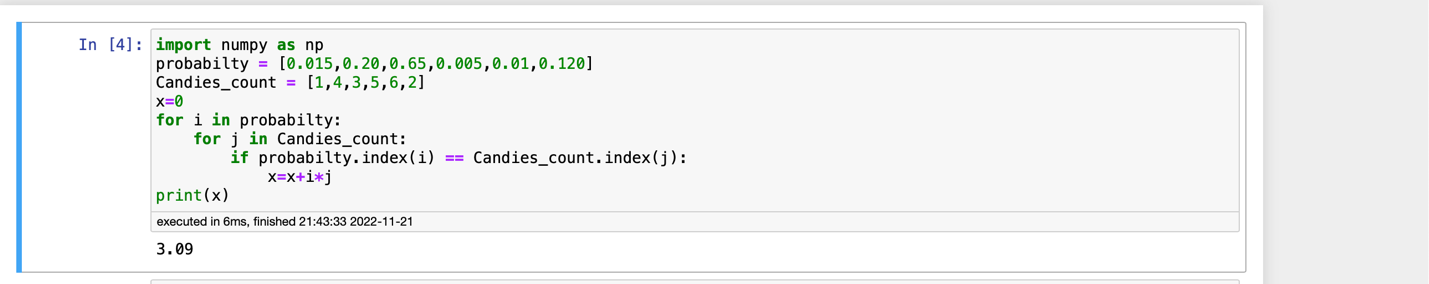
|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

* Sol :- number of candies for a randomly selected child =

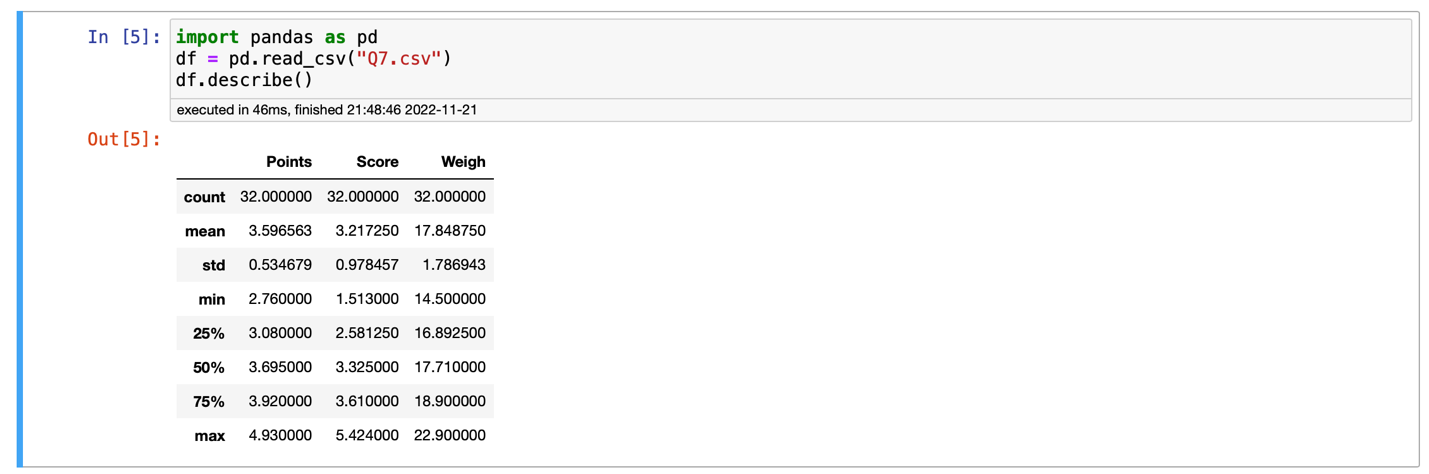
1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120 = 3.09

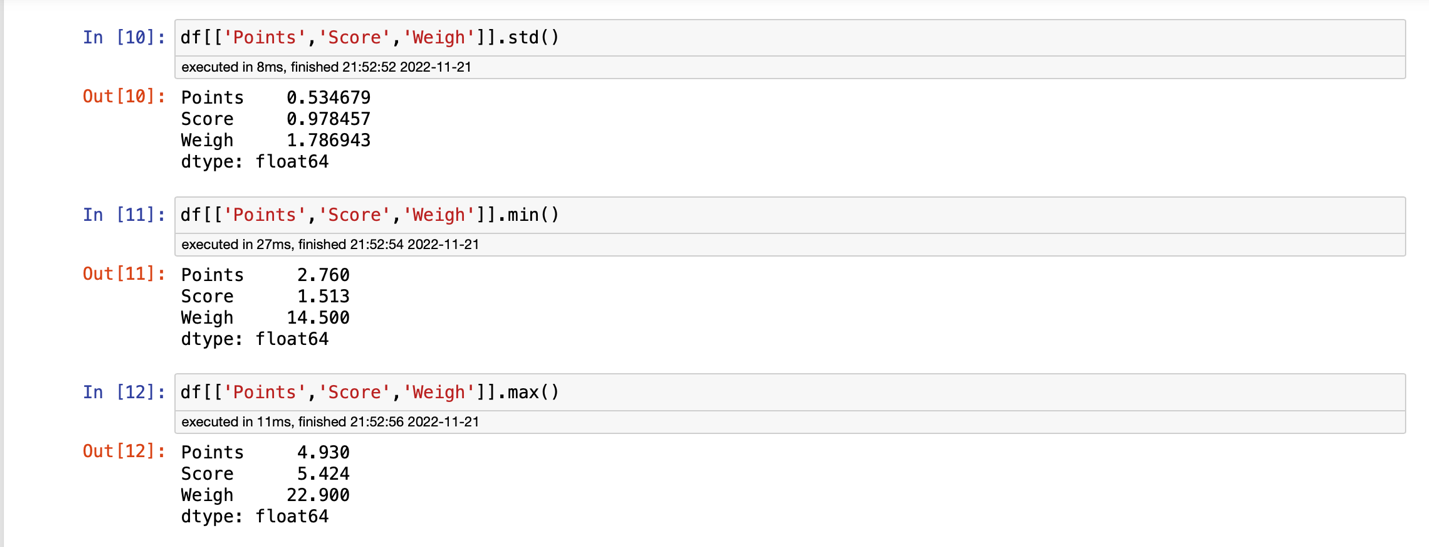
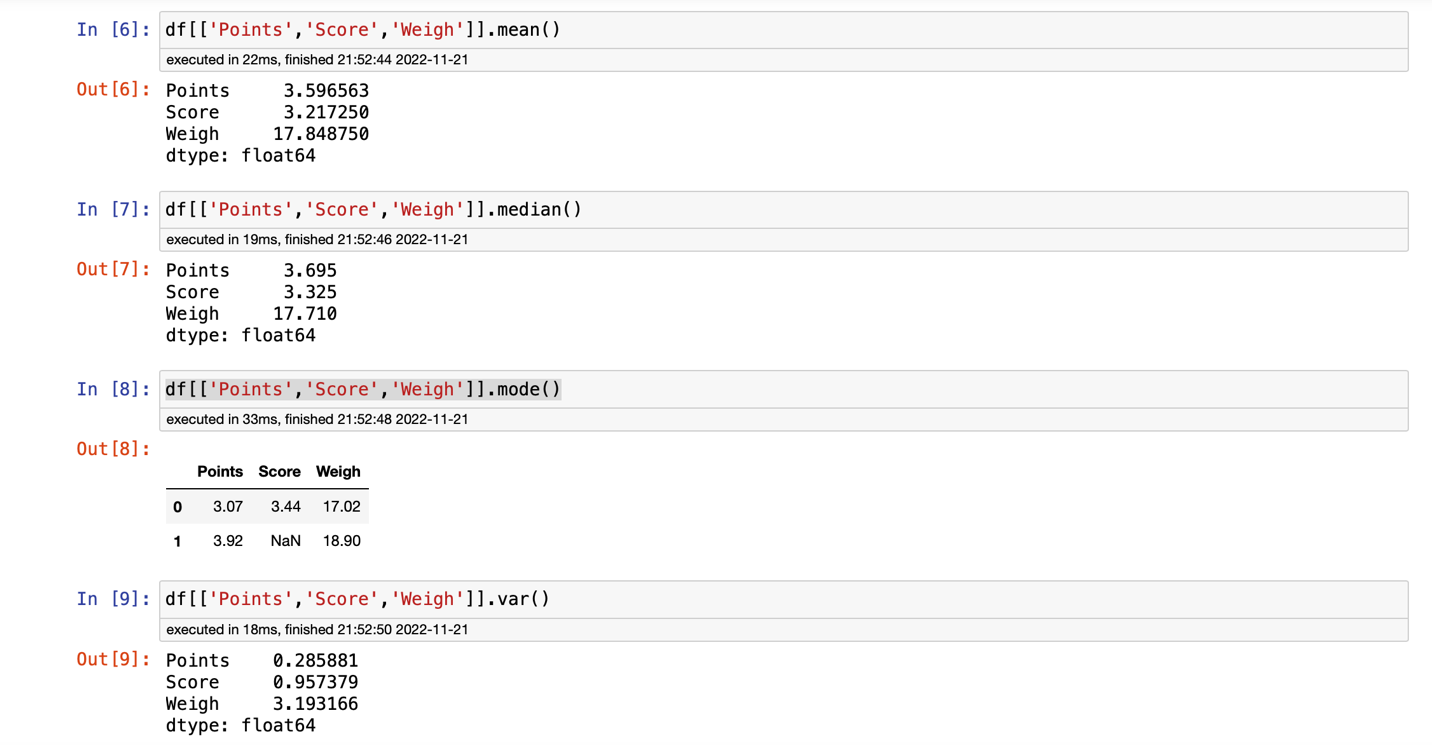


Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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 **

****

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

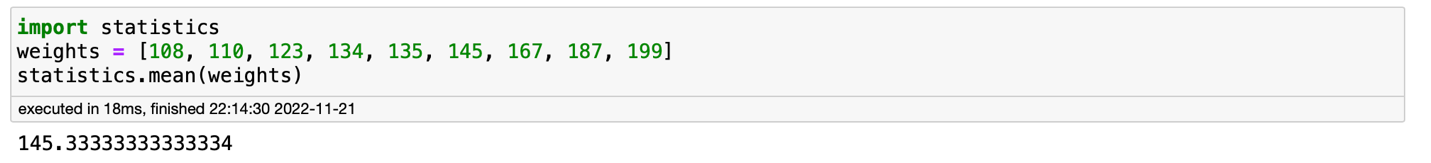
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Sol:- Expected Value = ∑ (Probability\*Value)

Probability to choose one patient = 1/9

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

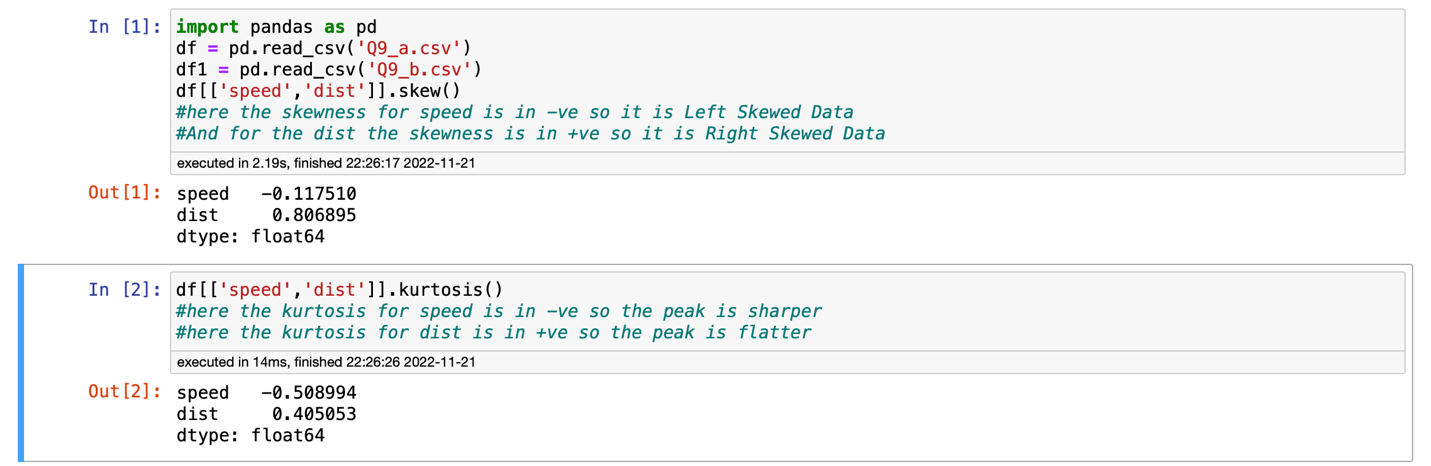
= 145.33



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

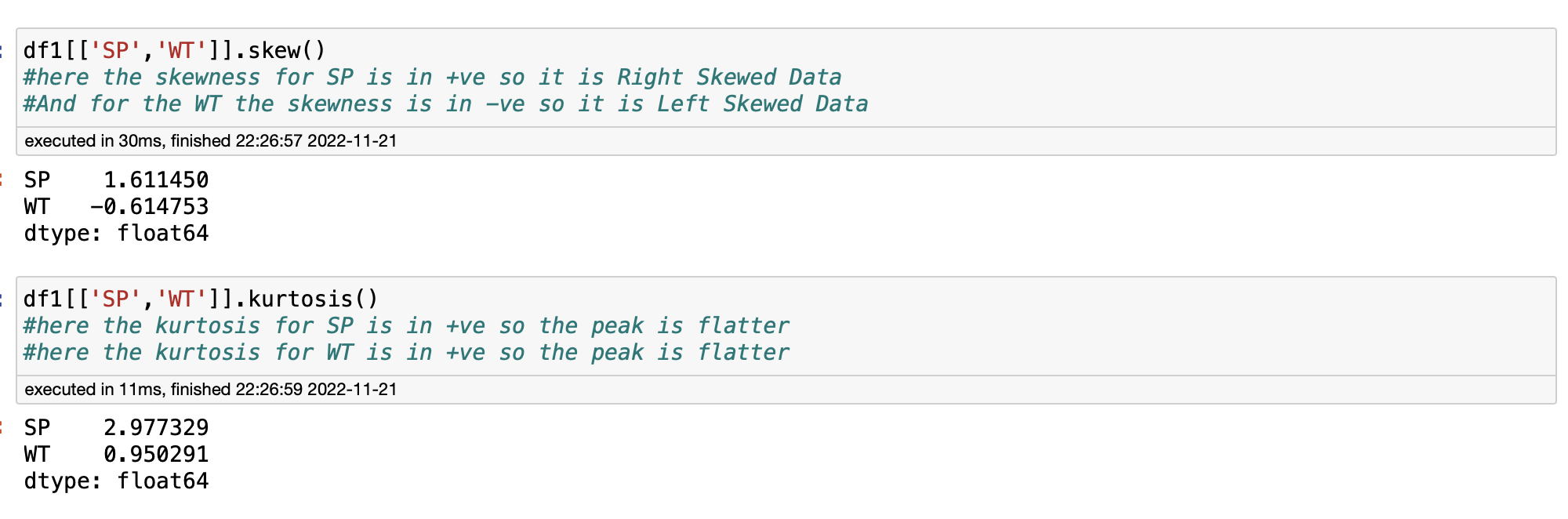
**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



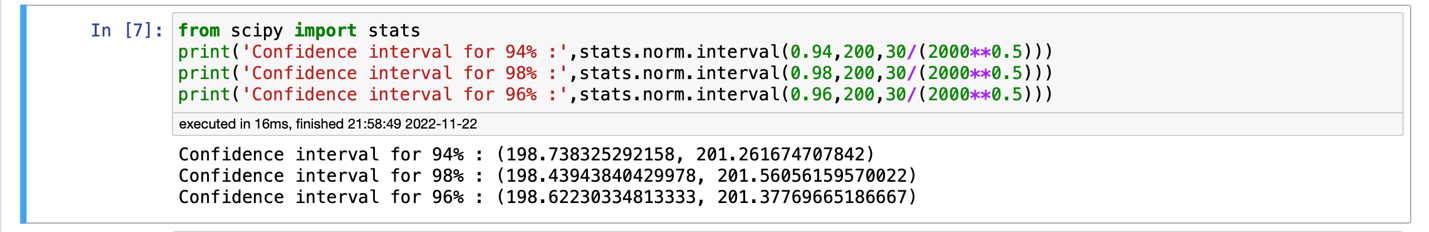
Sol :- Skewness :- Here it is clearly Right Skewed Data So Mean > Median and Skewness will be Positive



Sol:- There are outliers on the upper side of the Box

There are less data points between Q1(Lower Quartile) and Lower Extreme

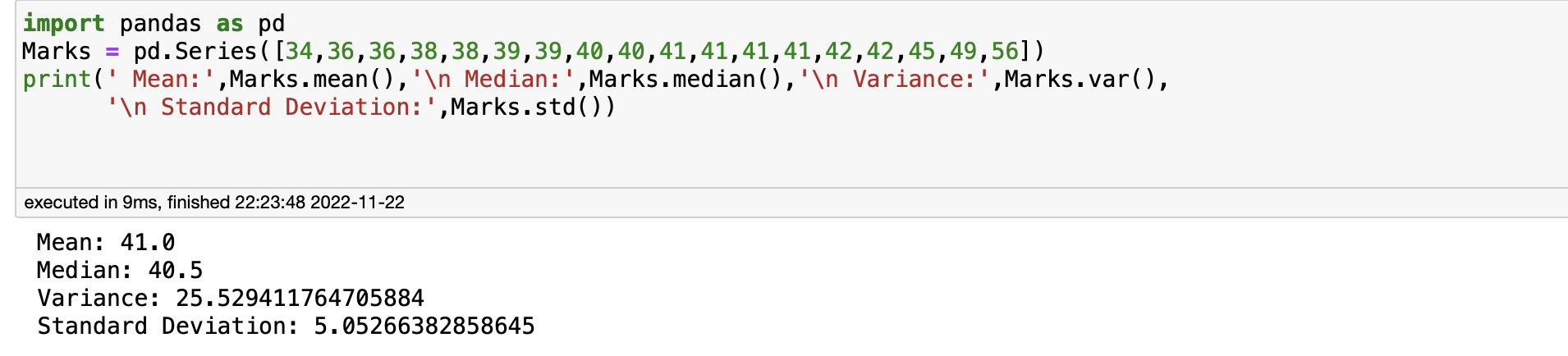
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 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.



1. What can we say about the student marks?

Sol : The most marks scored by the students is : 41

There are two outliers : 49 and 56

The mean is greater than median which is 41



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

* Then it is perfectly Symmetrical
* Skewness = 0 (Normal Distribution)

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

* Then data is Right Skewed
* Skewness will be +ve

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

* Then data is Left Skewed
* Skewness will be -ve

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

* If the kurtosis value is positive then the peak is Sharper
* The data is normally distributed and kurtosis value is o

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

* If the kurtosis value is negative then the peak is flatter and the data has lighter tails

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



What can we say about the distribution of the data?

Sol :- The above boxplot is not normally distributed the median is towards higher value

What is nature of skewness of the data?

Sol:- The skewness of the data is left skewed as median > mean

What will be the IQR of the data (approximately)?   
Sol:- IQR = 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.

Sol:- There are no Outliers

Whiskers level is high in Boxplot 2

Mean and Median are equal so distribution is Symetrical

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

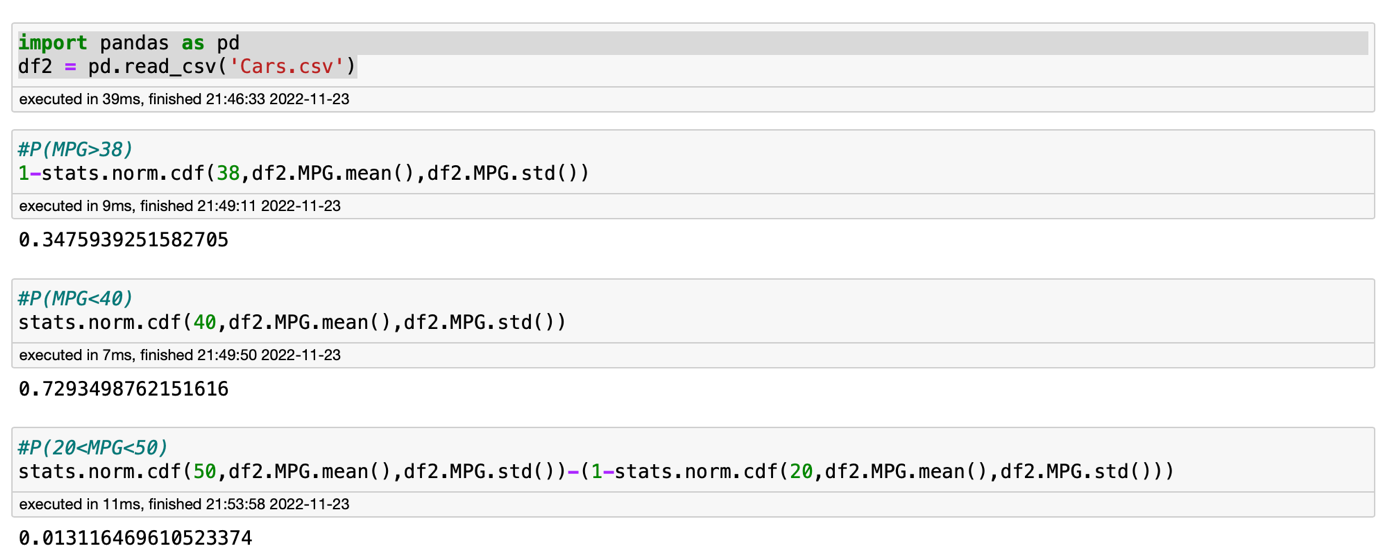
Data \_set: Cars.csv

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)

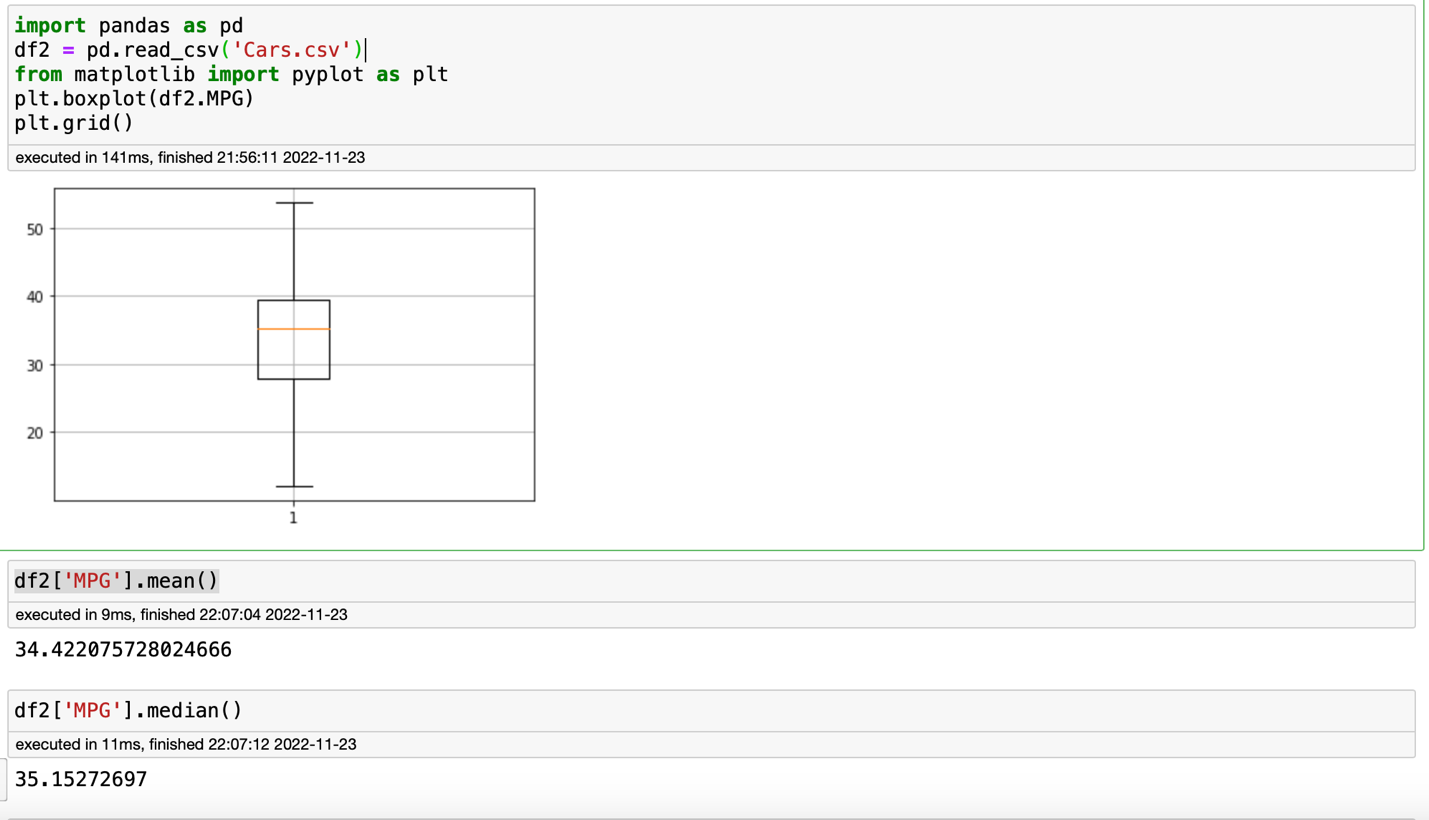


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol:- Here the mean and median almost same so the distribution is fairly Symmetrical.

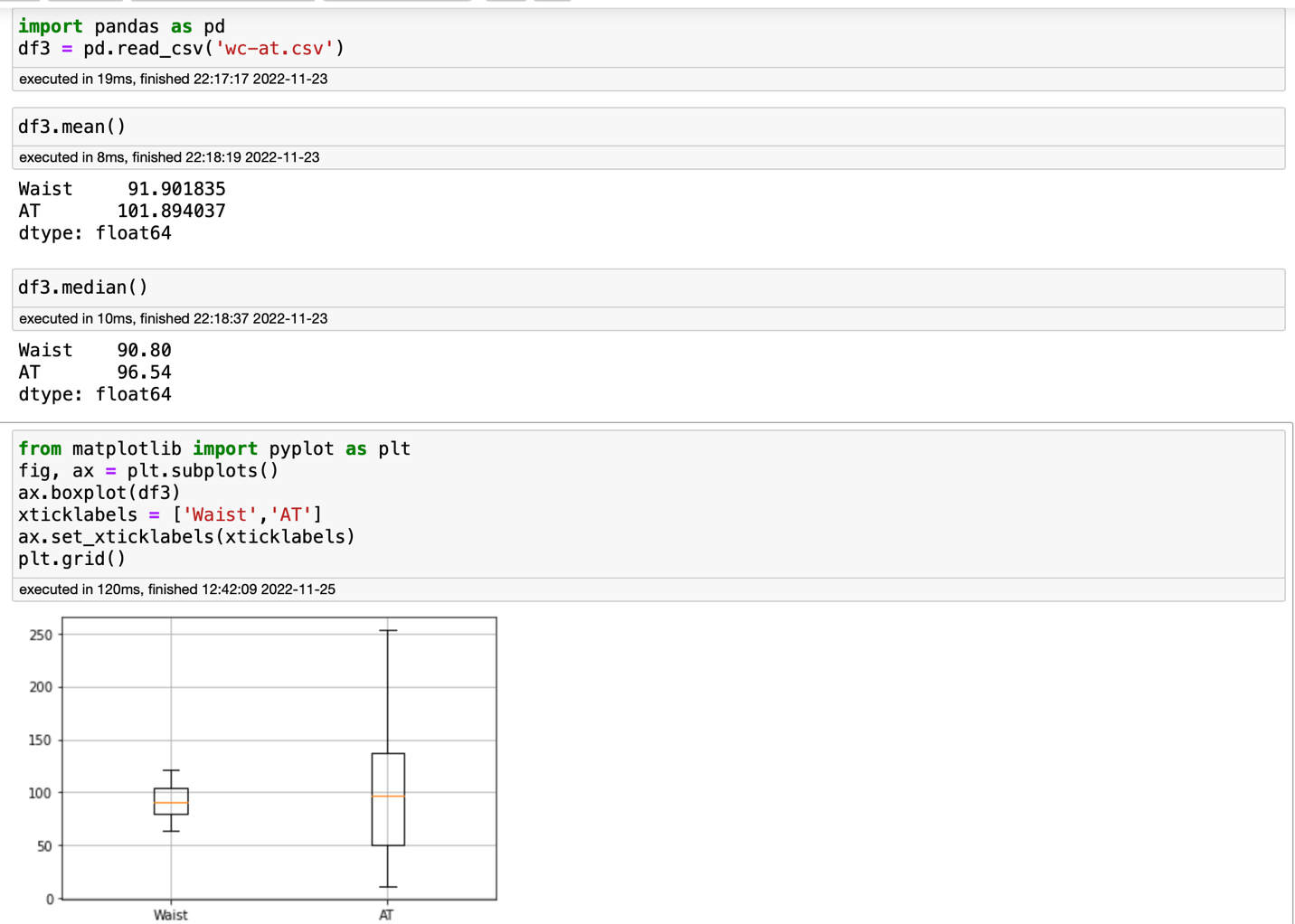


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

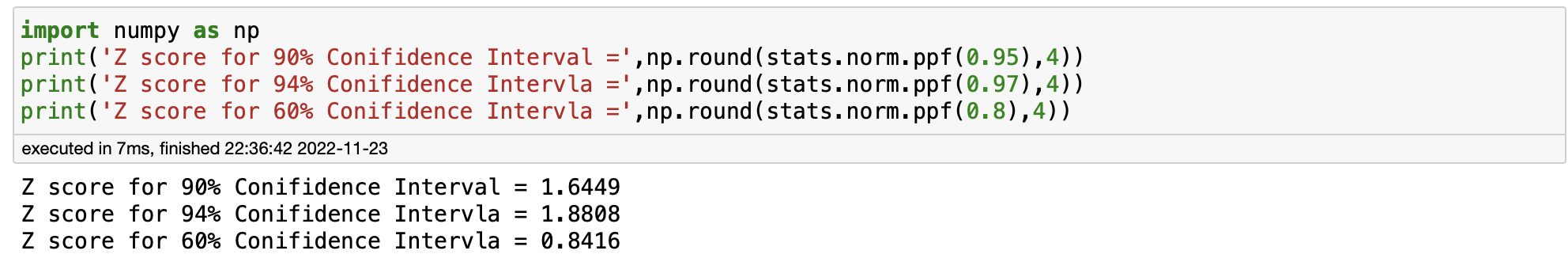
Dataset: wc-at.csv

Sol:- No Adipose Tissue does not follow Normal distribution as mean>median and and the lower whisker is larger than higher whisker so the data is positively skewed.

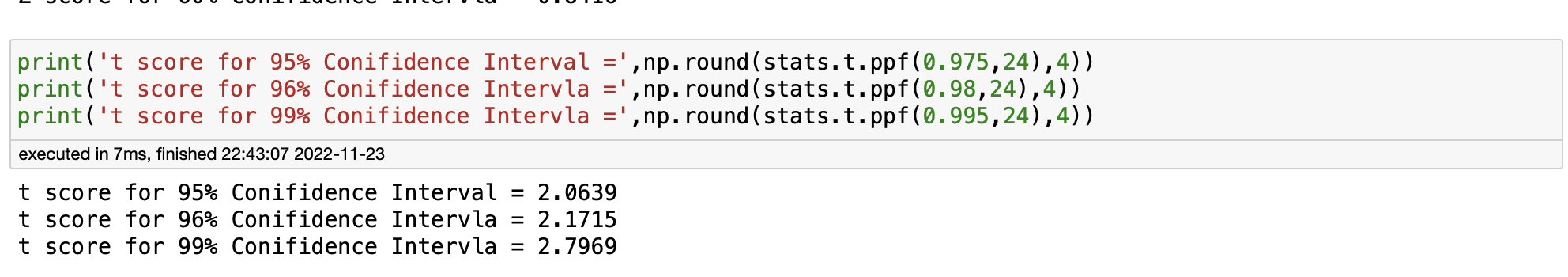
For Waist Mean and median are almost same and also both whiskers are in equal length it maybe fairly symmetrical



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

