|  |  |
| --- | --- |
| 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-Countable |
| Number of times married | Discrete-Countable |
| Gender (Male or Female) | Discrete-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 | Interval |
| 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 | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Let's consider the three possible orders:

Head Head Tail: Probability = (1/2) \* (1/2) \* (1/2) = 1/8

Head.Tail.Head: Probability = (1/2) \* (1/2) \* (1/2) = 1/8

Tail.Head.Head: Probability = (1/2) \* (1/2) \* (1/2) = 1/8

Total Probability = (1/8) + (1/8) + (1/8) = 3/8

So, the probability of getting two heads and one tail when tossing three coins is 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

**a) Probability of a sum equal to 1: 1/36**

**b) Probability of a sum less than or equal to 4: 1/6**

**c) Probability of a sum divisible by 2 and 3: 1/36**

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?

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

**Expected Number of Candies = (Probability of A \* Candies Count of A) + (Probability of B \* Candies Count of B) + (Probability of C \* Candies Count of C) + (Probability of D \* Candies Count of D) + (Probability of E \* Candies Count of E) + (Probability of F \* Candies Count of F)**

**Expected Number of Candies = (0.015 \* 1) + (0.20 \* 4) + (0.65 \* 3) + (0.005 \* 5) + (0.01 \* 6) + (0.120 \* 2)**

**Expected Number of Candies = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24**

**Expected Number of Candies = 3.105**

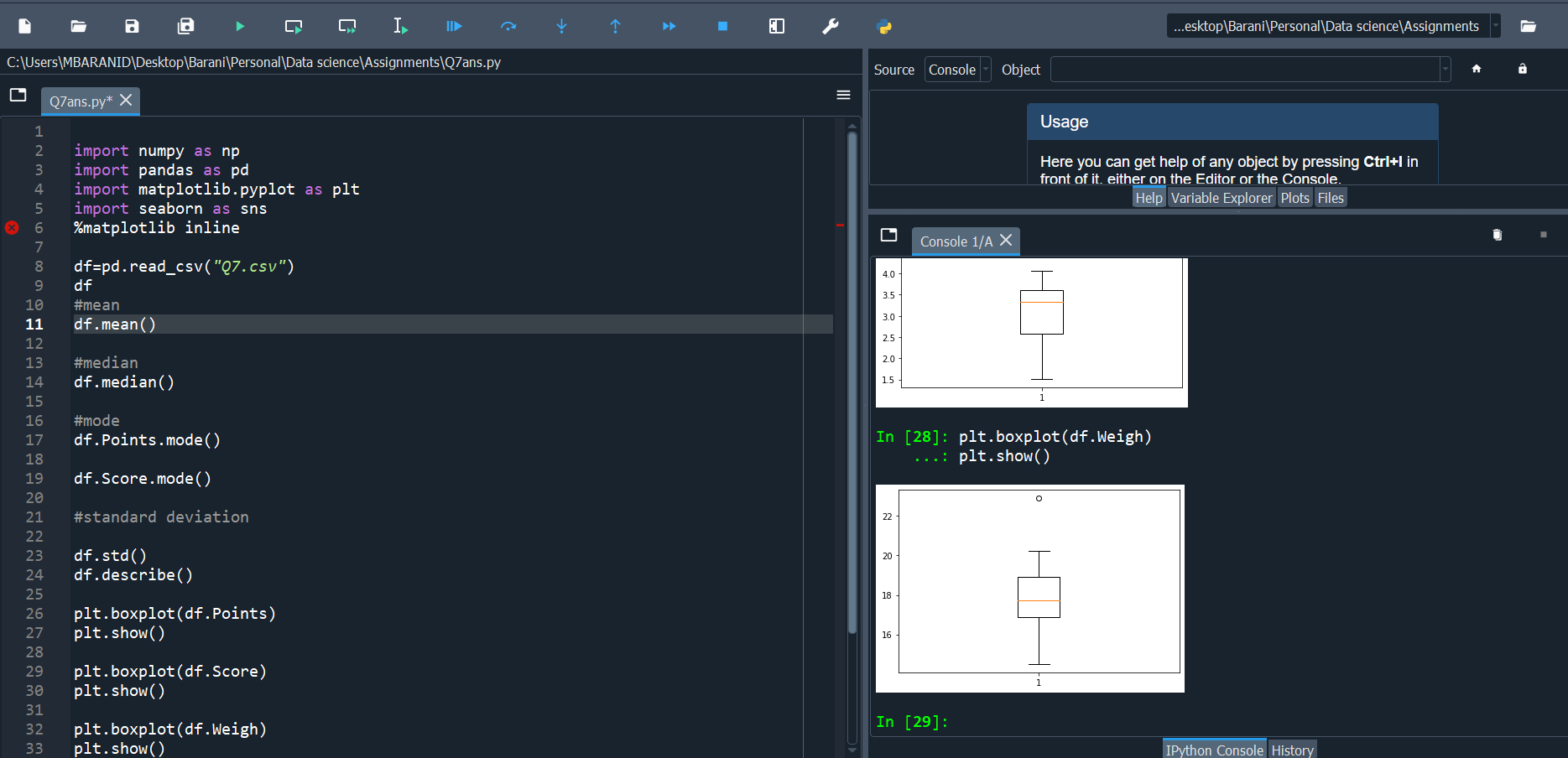
**So, the expected number of candies for a randomly selected child is 3.105.**

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.

**Code files attached-Q7ans.py**



**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  =  ∑ ( probability  \* Value )**

**∑ P(x).E(x)**

**there are 9 patients**

**Probability of selecting each patient = 1/9**

**Ex  108, 110, 123, 134, 135, 145, 167, 187, 199**

**P(x)  1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9**

**Expected Value  =  (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)**

**= 145.33**

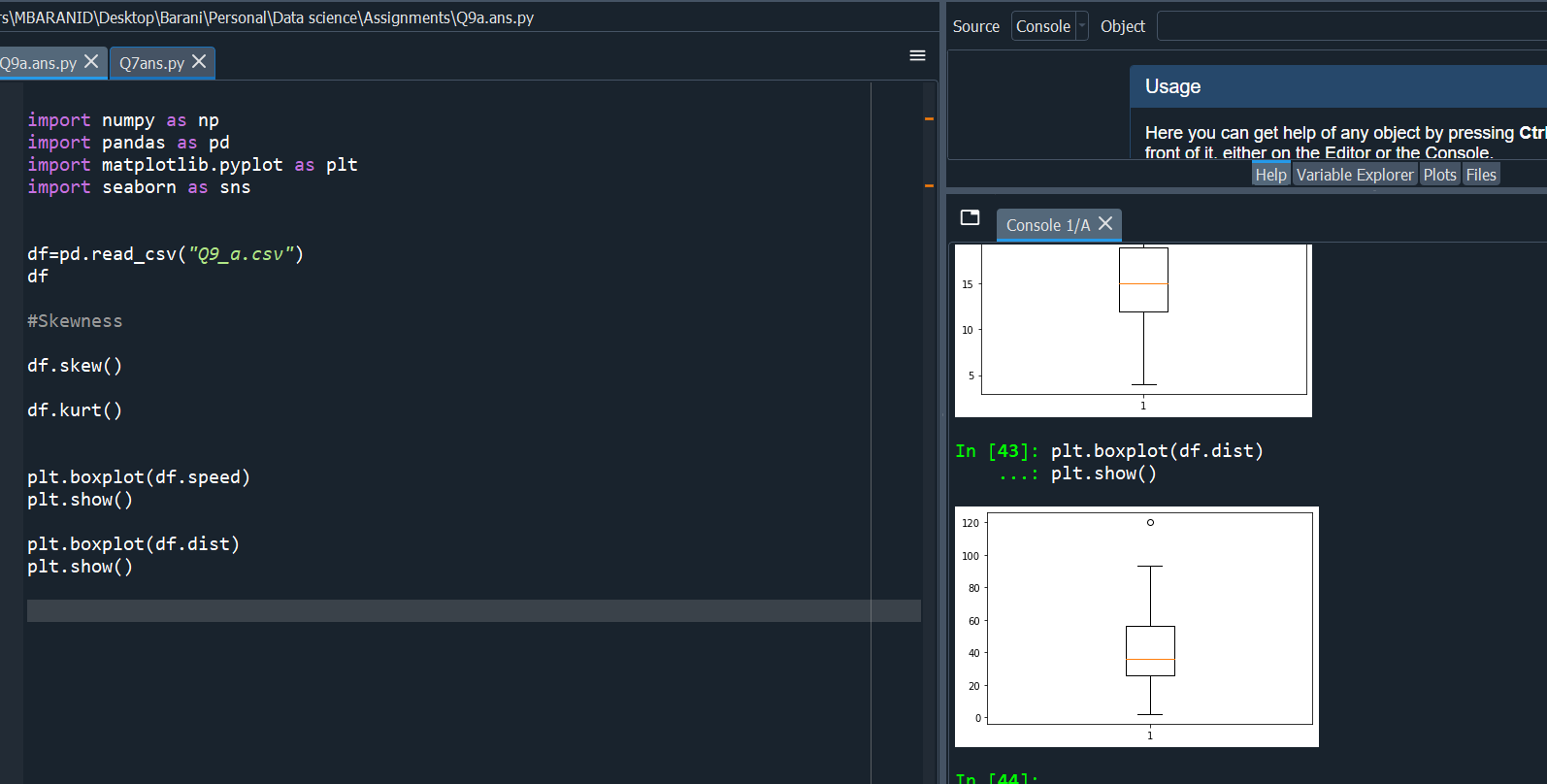
**Expected Value of the Weight of that patient = 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

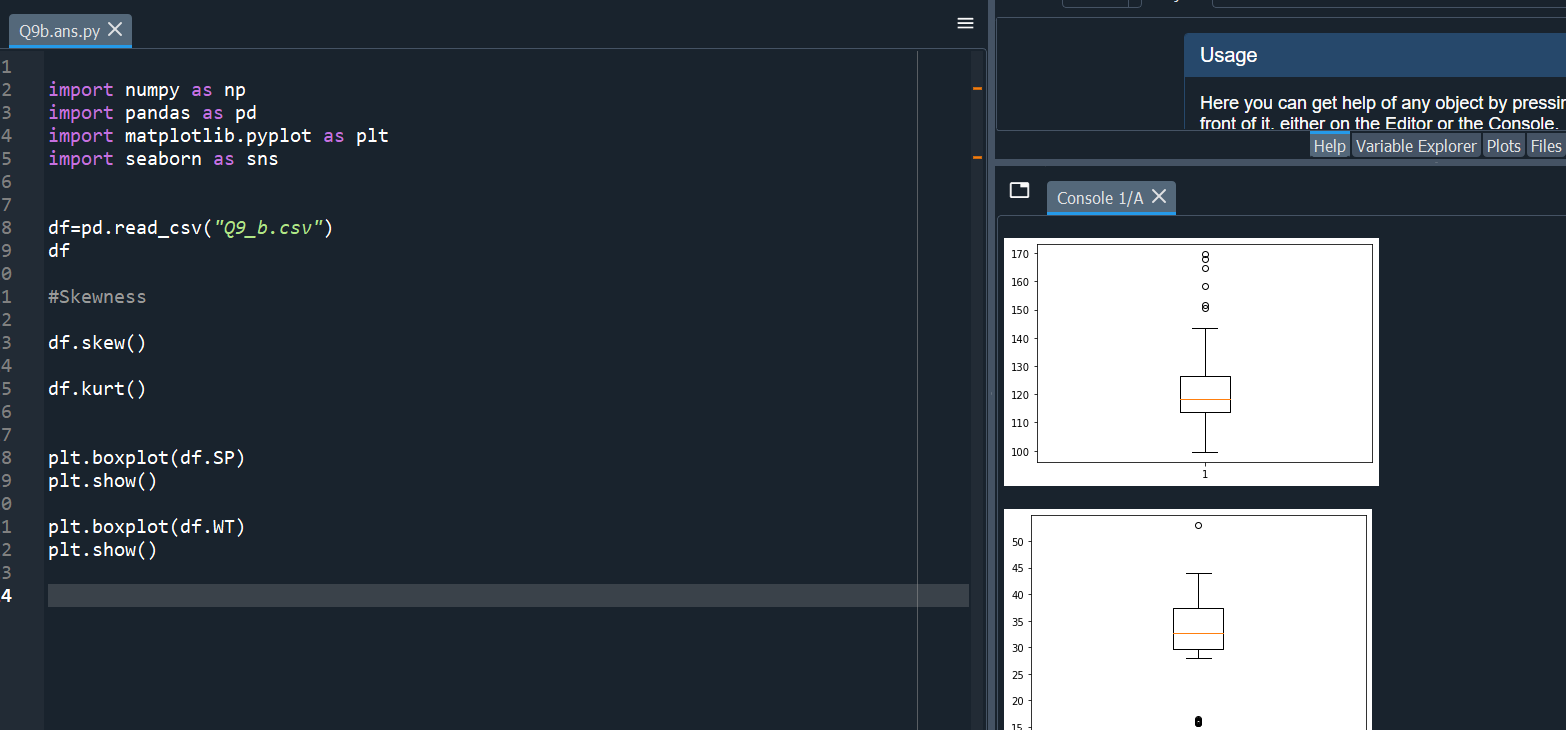
**Code file attached (Q9a.ans)**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Code file attached (Q9b.ans)**

****

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



Histogram: - Chick weight data is right skewed or positively skewed.---- Yes –

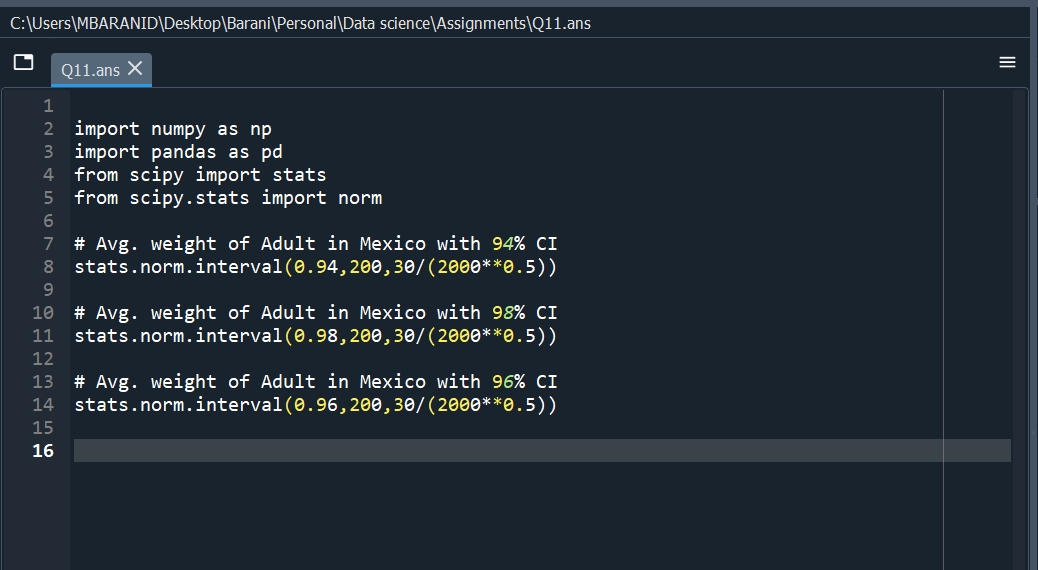
More than 50% Chick Weight is between 50 to 150. ---- Yes

Most of the chick weight is between 50 to 100. --- Yes

The data is right skewed.

There are outliers at upper side

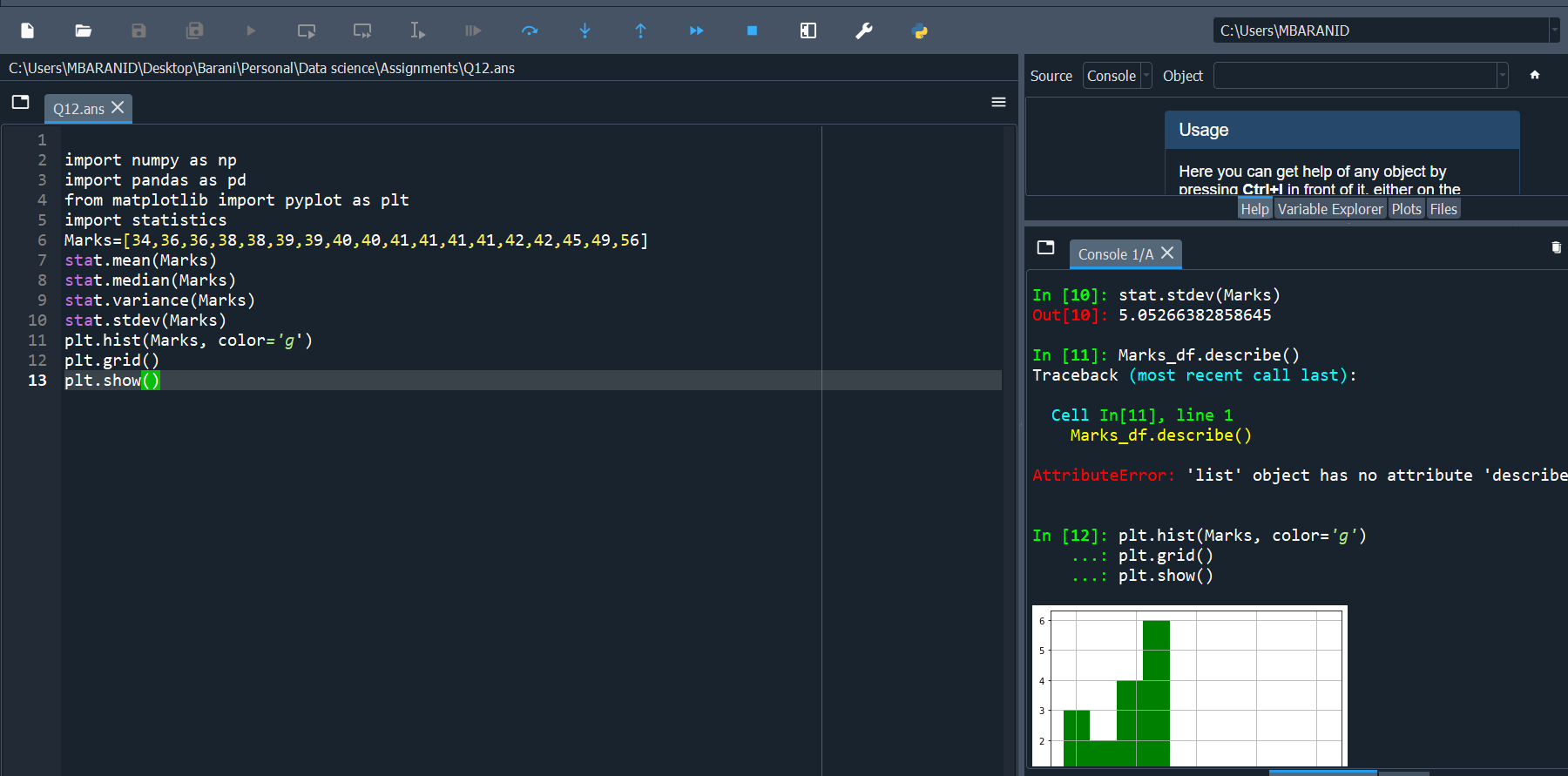
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?



**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?



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

**It is called Zero Skewed**

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

**Ans: Right skewed distribution**

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

**Ans: Left skewed distribution**

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

**Ans: A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.**

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

**Ans: : A distribution with a negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.**

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



What can we say about the distribution of the data?

**Answer: Asymmetrical distribution**

What is nature of skewness of the data?

**Ans: Left skewed**

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

The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 18 – 10   
  
**Ans: IQR = 8**

Q19) Comment on the below Boxplot visualizations?



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

Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots are equal. And the data spread in both of them are symmetrical

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

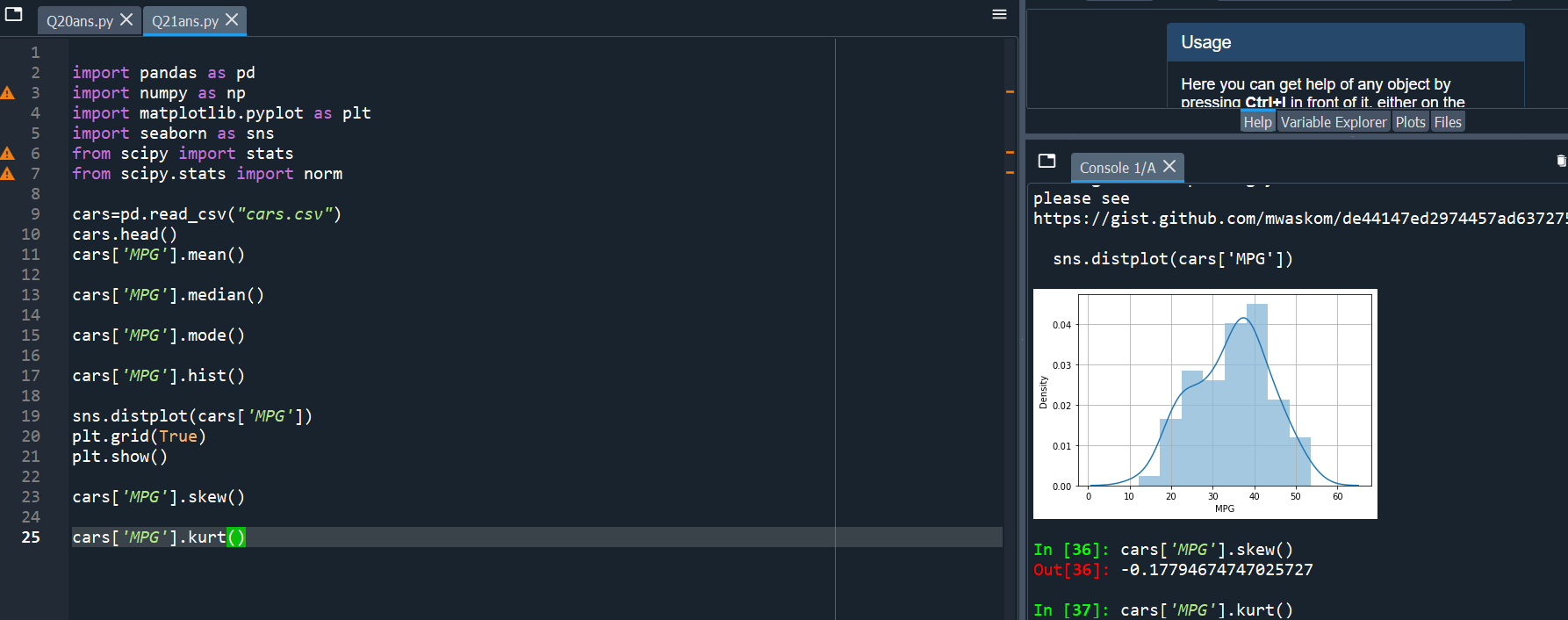
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

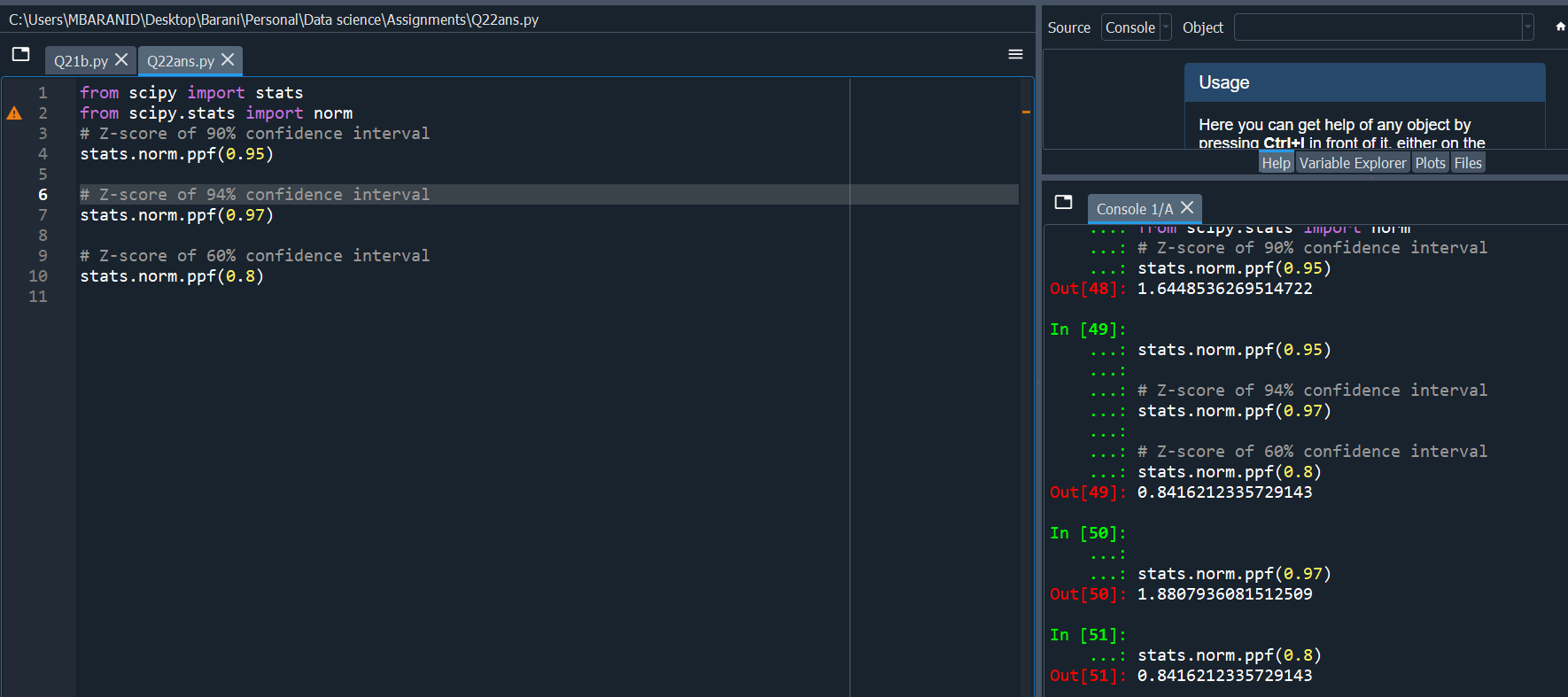


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

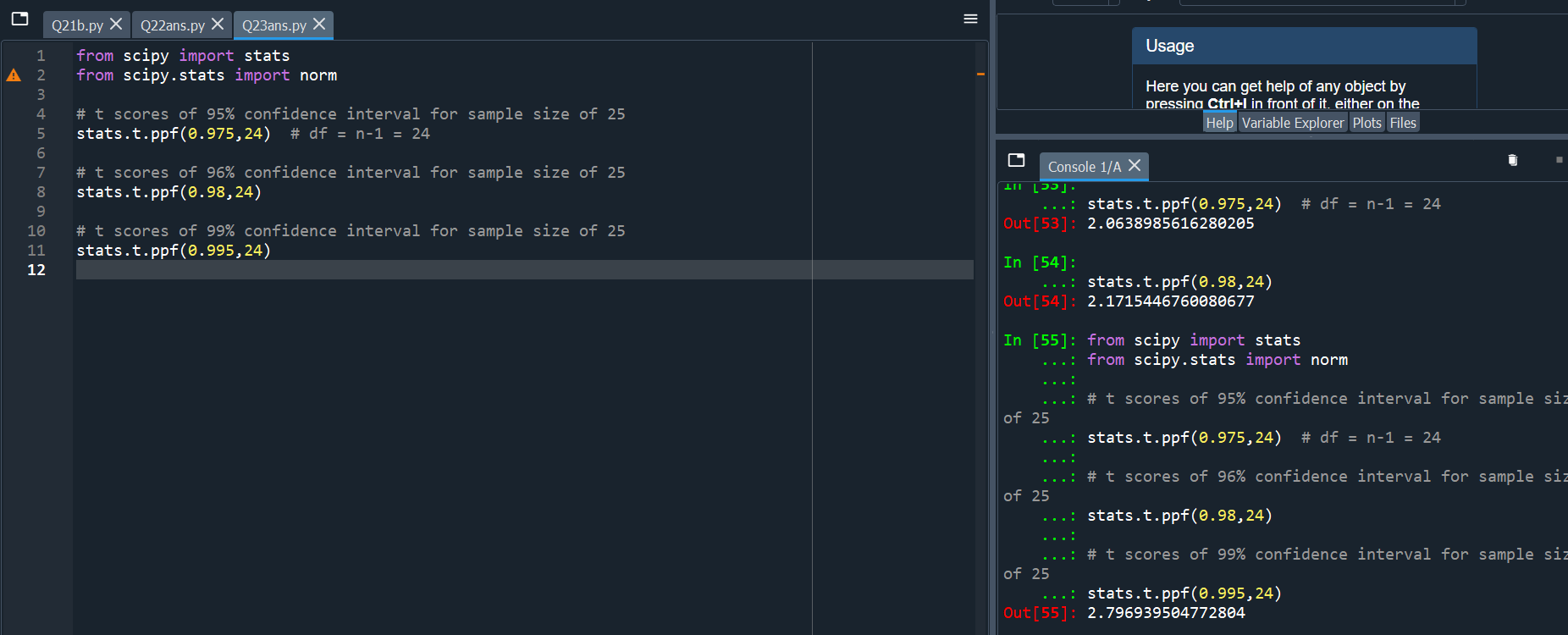
Dataset: wc-at.csv



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 degrees of freedom are equal to 18 - 1 = 17. pt(0.95,17) = 1.73 Probability using T distribution = (1.73)\*90/sqrt(18)

=(1.73)\*90/4.242

= 36.69

=36%