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

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 | Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Interval |
| 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 | Nominal |
| Religious Preference | Ratio |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

* Total possible outcomes ={HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}

n (s) = 8

{HHT,HTH,THH}=3/8

Probability of two heads and one tail 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

* Total possible outcomes={ (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. Probability =zero
2. n (s) = { (1,1),(1,2),(1,3),(2,1),(2,2),(3,1) }

Probability = 6/36 = 6

1. n(s) = {(1,1),(1,2),(1,3),(1,5),(2,1),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5),(3,6)

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

Probability = 24/36

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?

* Total numbers of balls = 2+3+2 = 7

Total Case = 7c2 = 7!/2!\*5! (! = Factorial)

= (7\*6\*5\*4\*3\*2\*1)/(2\*1)\*(5\*4\*3\*2\*1)

n (s) =21

Favourable Condition = 5c2 = 5!/2!\*3!

= (5\*4\*3\*2\*1)/(2\*1)\*(3\*2\*1)

= 10

Probability = 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 = summation (Candies Count\*Probability)

=summation ((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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.59 | 3.21 | 17.84 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.285 | 0.957 | 3.19 |
| Standard Deviation | 0.534 | 0.978 | 1.786 |
| Range | 2.17 | 3.911 | 8.4 |

|  |  |  |
| --- | --- | --- |
|  |  |  |

**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 = Sum{X\*Probability(X)}

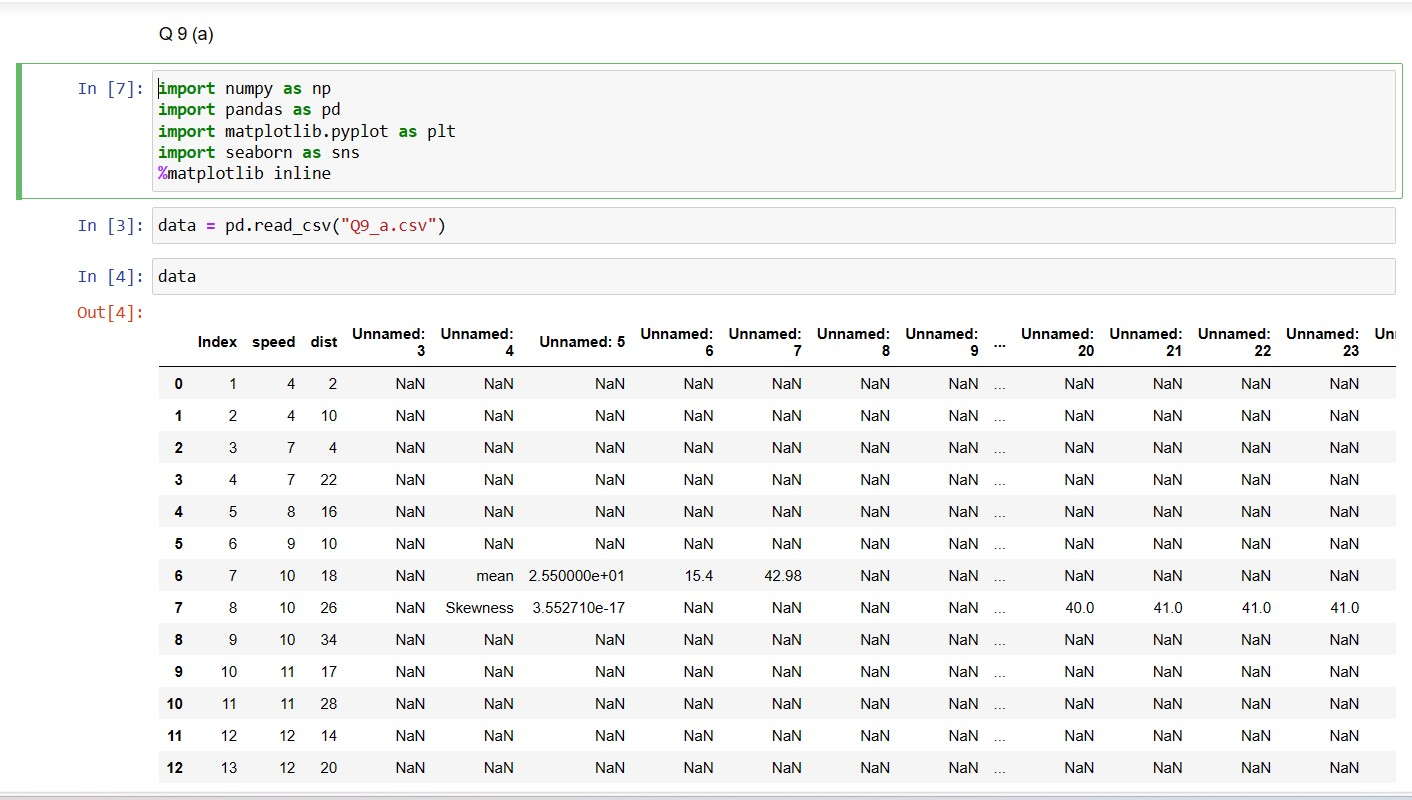
=Sum{(108\*1/9),(110\*1/9),(123\*1/9),(134\*1/9),(135\*1/9),(145\*1/9),(167\*1/9),(187\*1/9),(199\*1/9)}

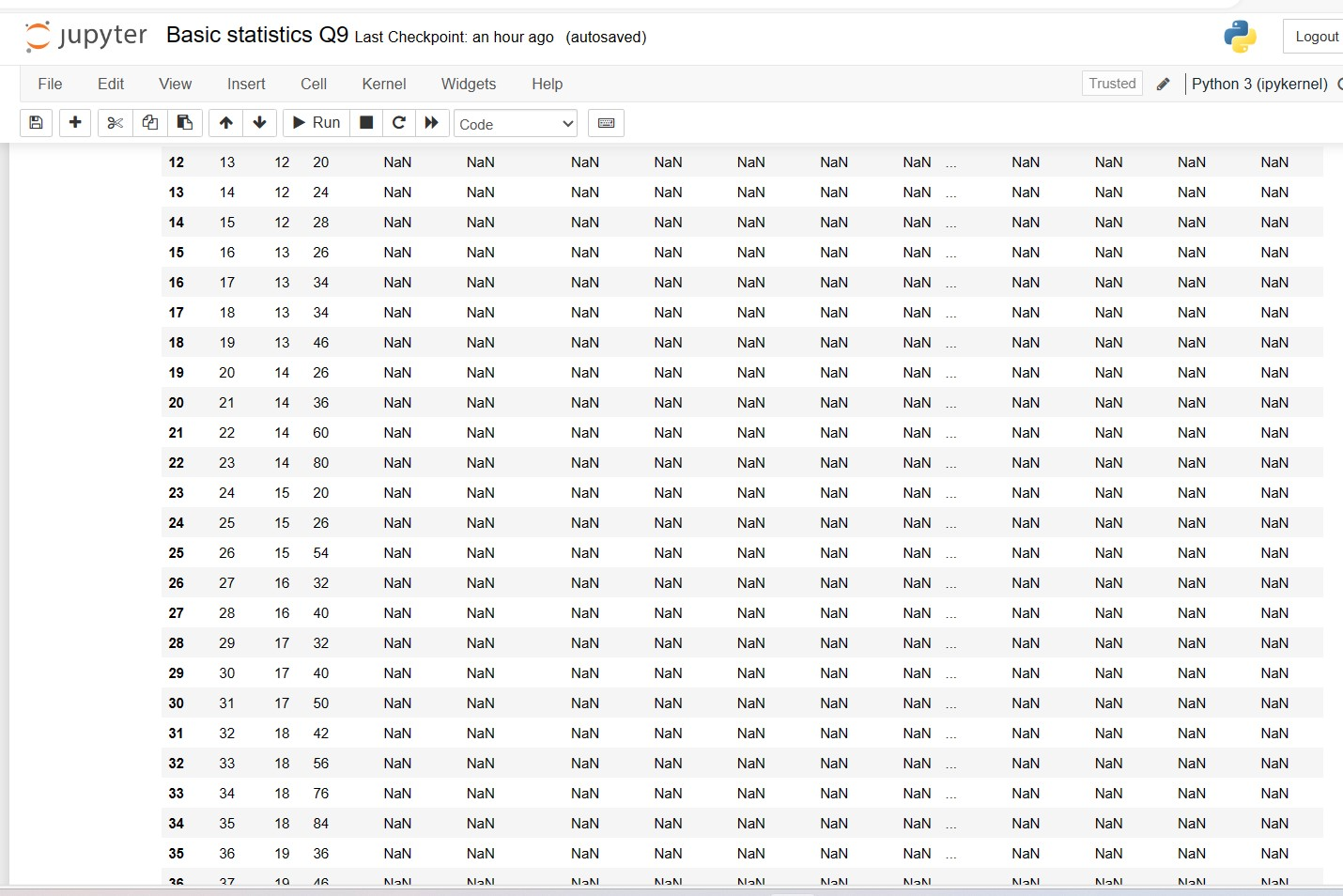
= 130.33

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

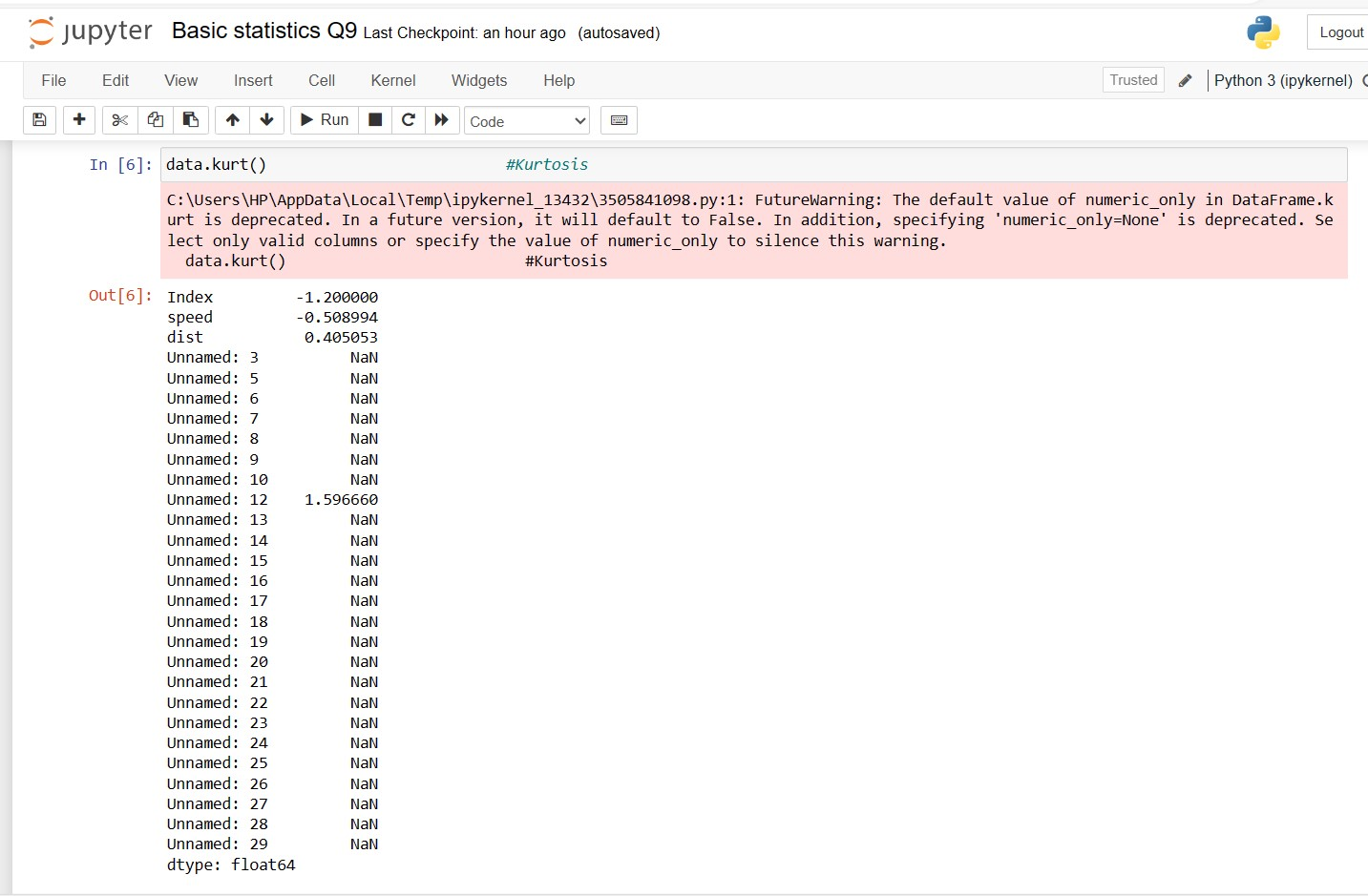
**Cars speed and distance**

**Use Q9\_a.csv**



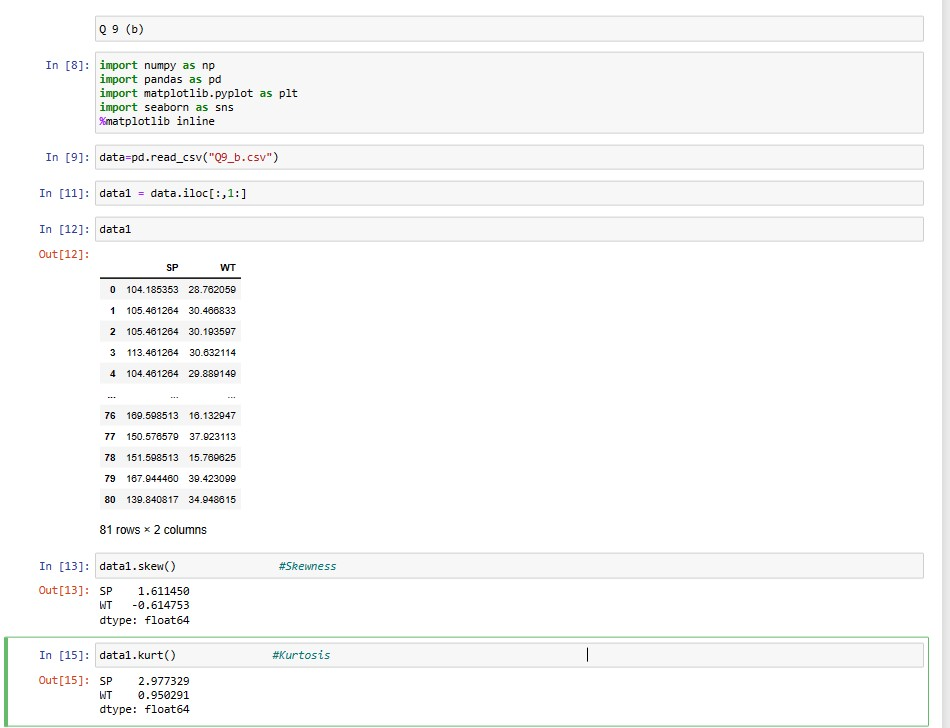




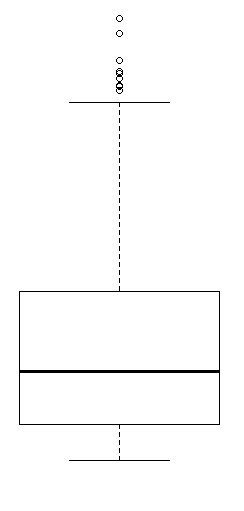


**SP and Weight(WT)**

**Use Q9\_b.csv**



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



* The histogram pick has right skew and tail is one right. Mean>Media .We have outlier on the higher side.

The boxplot has outliers on the maximum 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?



|  |  |  |  |
| --- | --- | --- | --- |
| Mean | Median | Variance | Standard Deviation |
| 41 | 40.5 | 25.52 | 5.052 |

* We don’t have an outliers & the data is slightly skewed towards right because mean is greater than median.

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

* No skewness is present .

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

* Skewness towards right .

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

* Skewness towards left.

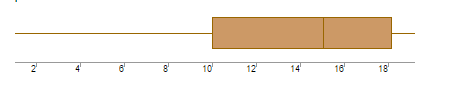
Q16) What does positive kurtosis value indicates for a data ?

* Positive kurtosis means peaked the curve is more peaked.

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

* Negative kurtosis means peaked the curve wide.

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



What can we say about the distribution of the data?

* The boxplot is not normally distributed , the median is skew towards the higher value.

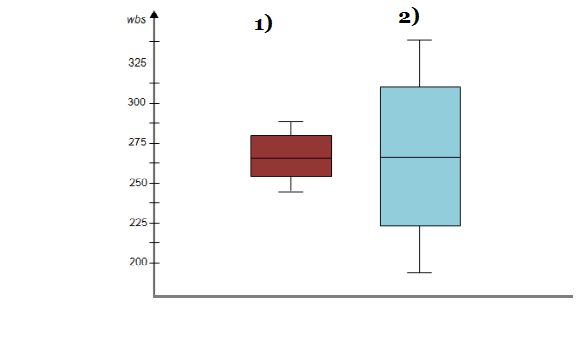
What is nature of skewness of the data?

* Left skewness

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

* IQR = (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.

* First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

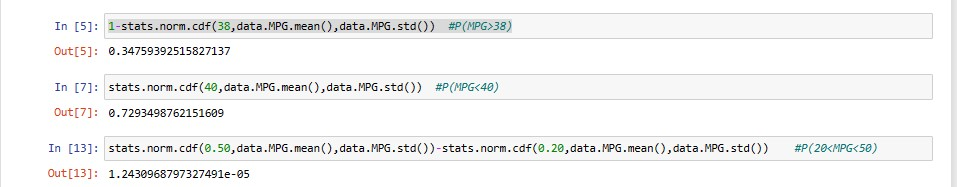
Data \_set: Cars.csv

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)

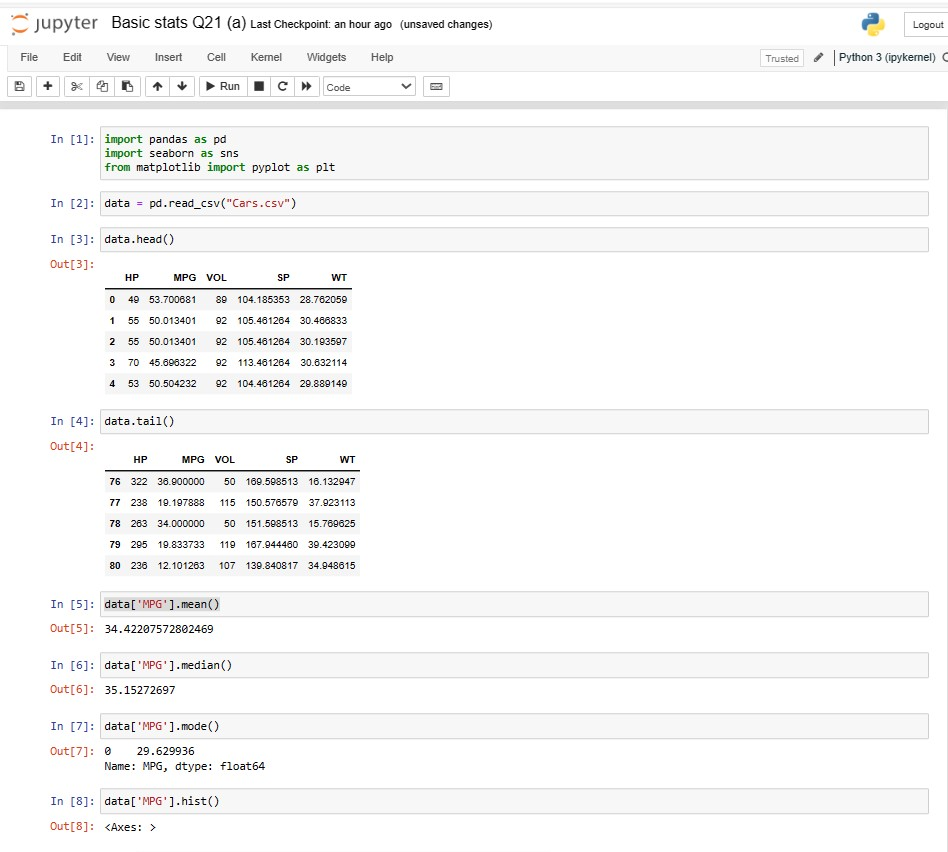


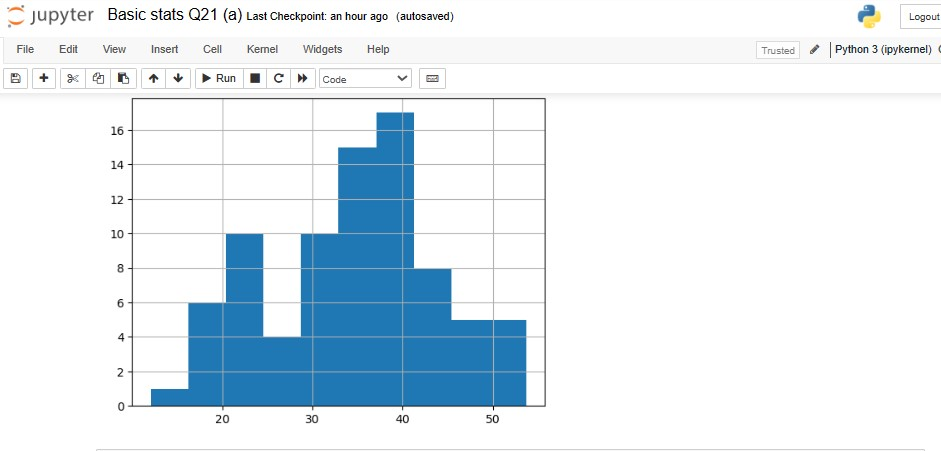


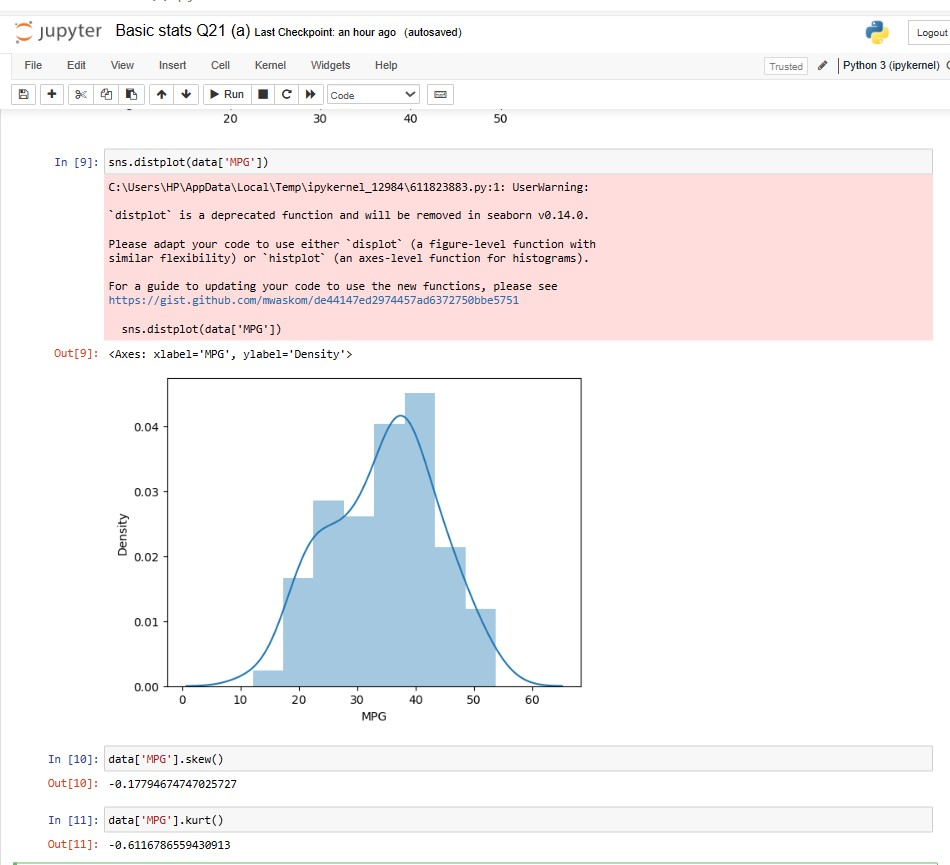
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



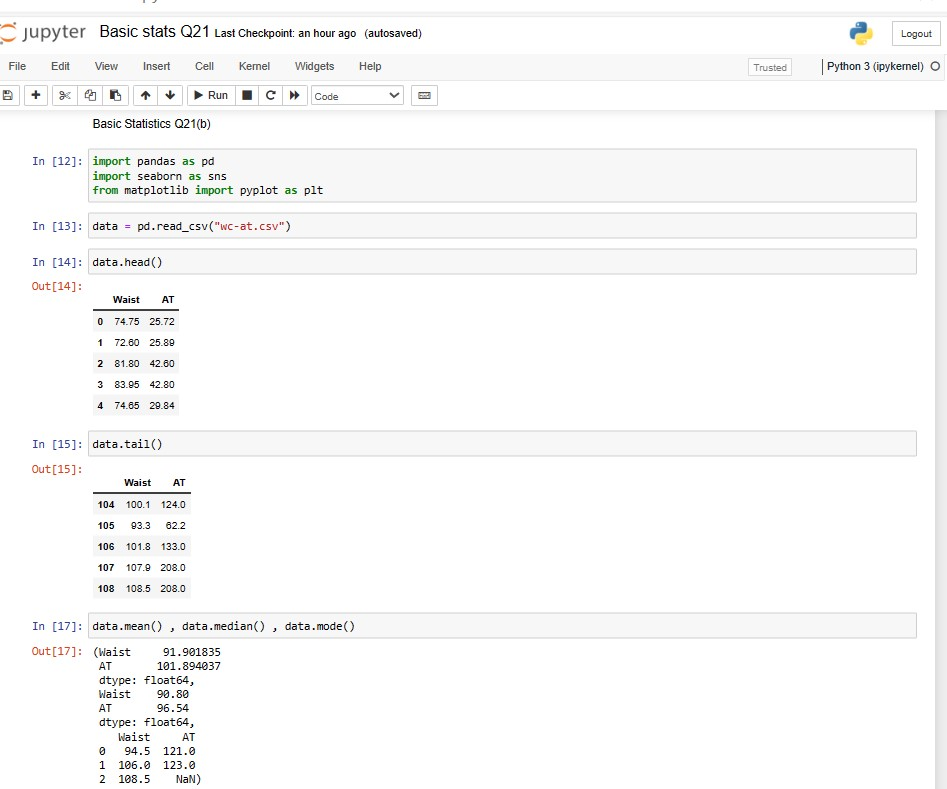


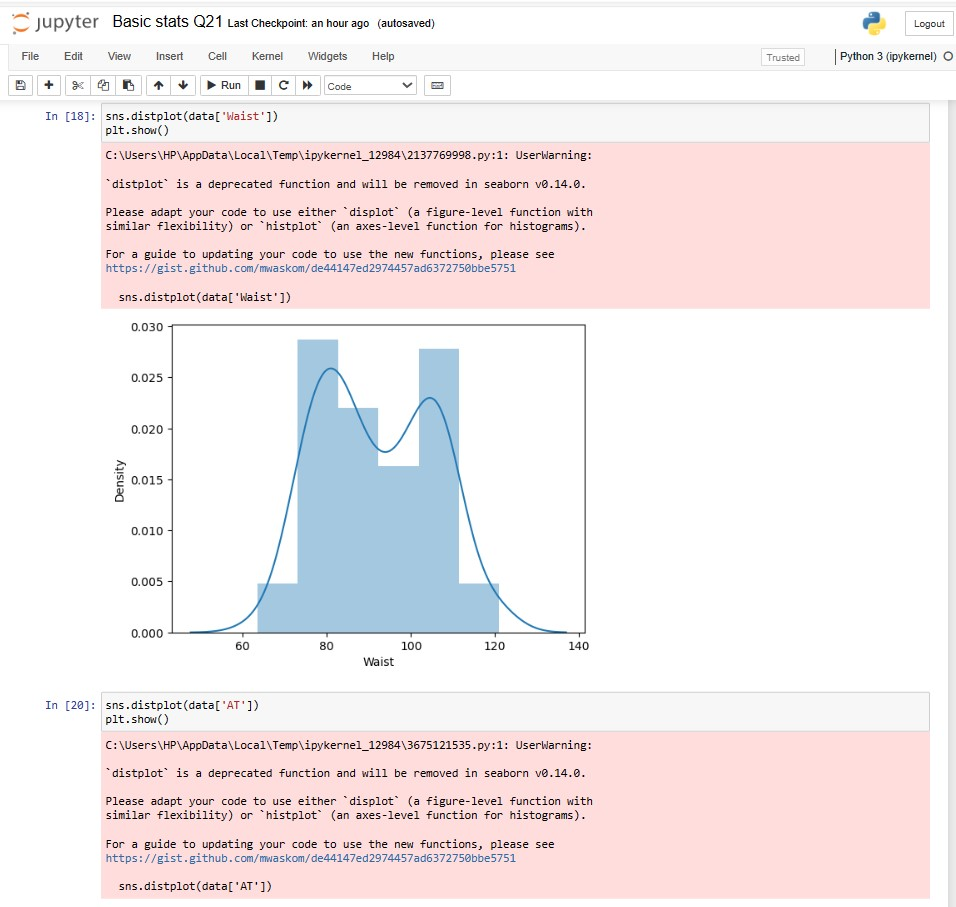


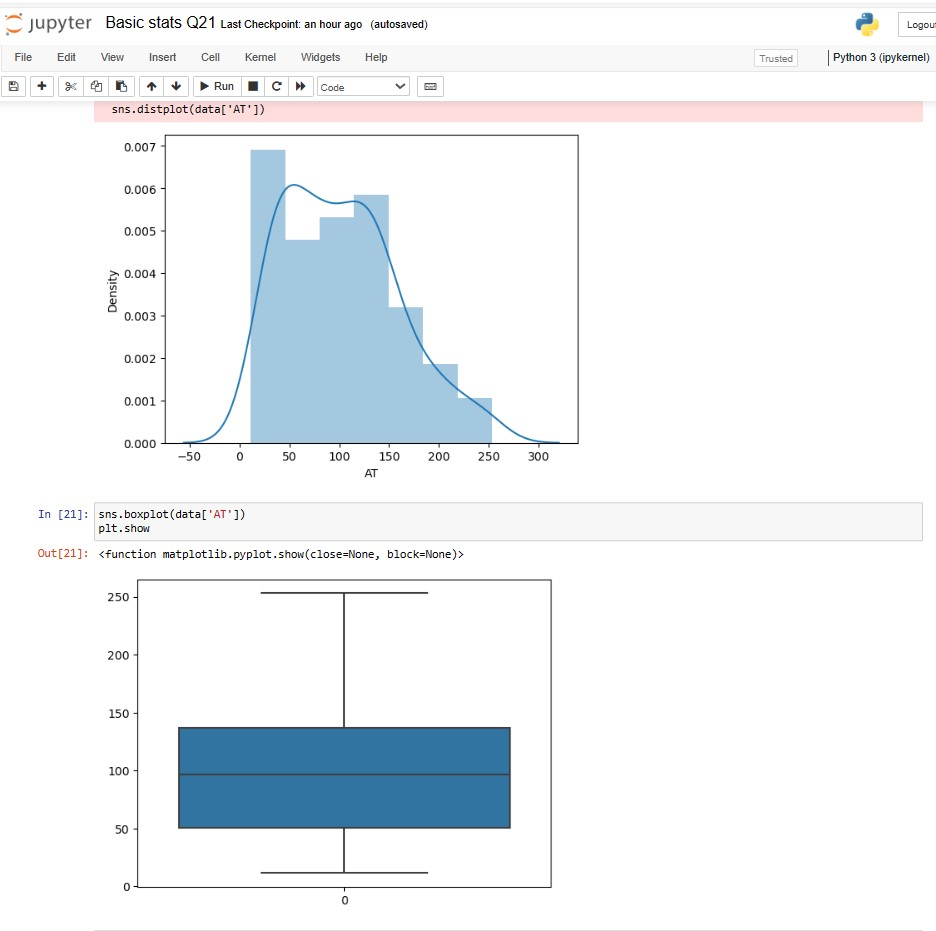
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

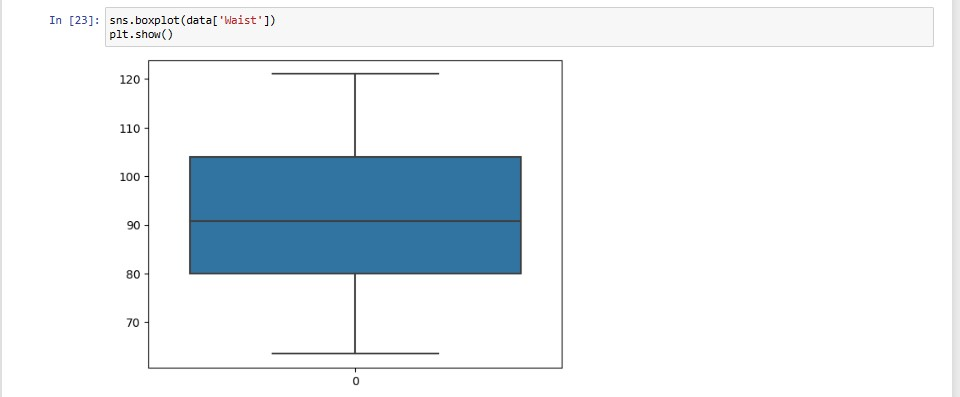
Dataset: wc-at.csv





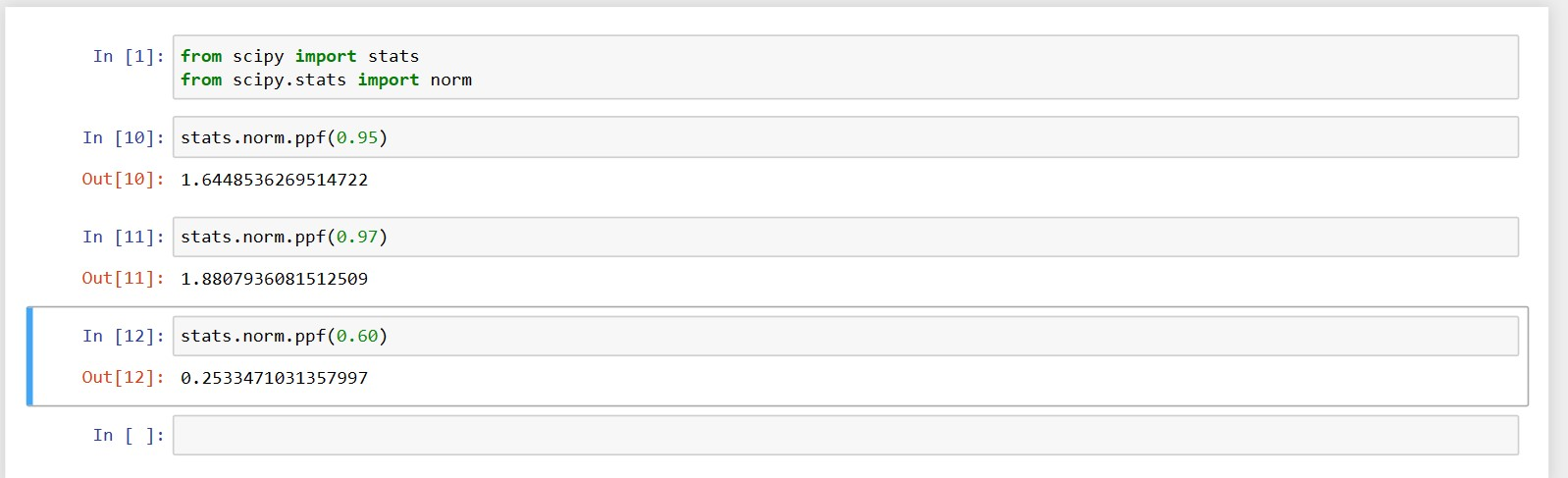






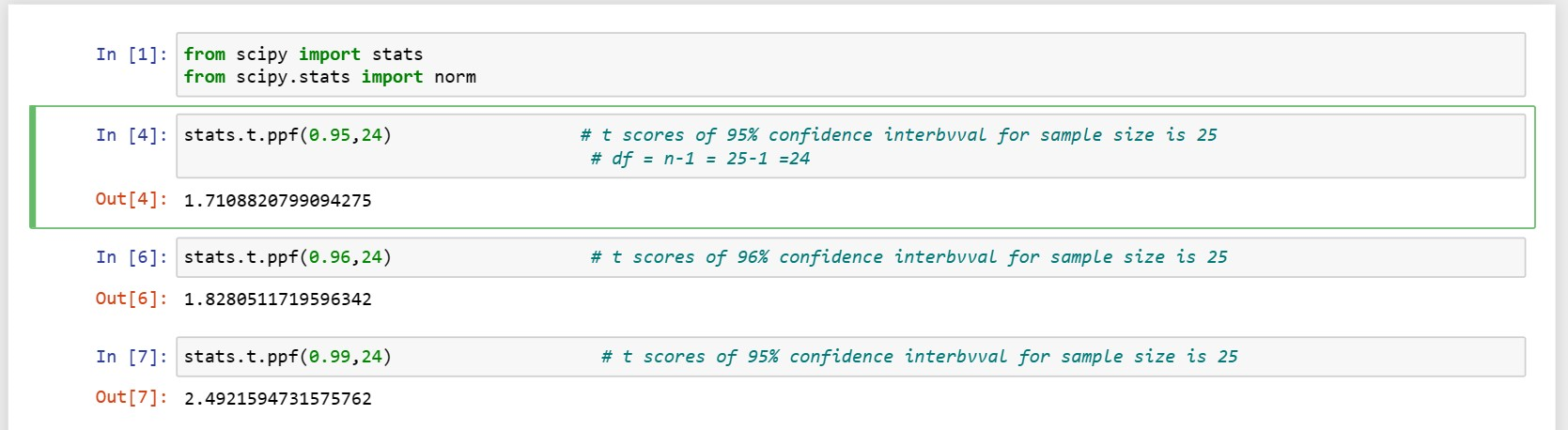
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval





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

* n = 18 , x bar = 260 , s = 90 , µ = 270

df = n-1 = 17

t = x bar - µ /(s/ root n) = (260-270)/(90/root18) = -0.471

P(t≤-0.471) = 0.321 (Using t distribution calculator and python)

There is a 32.1% chance that 18 randomly selected bulbs would have an average life of no more than 260 days .