

Q1) Identify the Data type for the Following:

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

e

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	Ratio
Blood Group	Nominal
Time Of Day	Interval
Time on a Clock with Hands	Interval
Number of Children	Ordinal
Religious Preference	Nominal

Barometer Pressure	Interval
SAT Scores	Ratio
Years of Education	Ordinal

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

Ans: $S = \{HHH, TTT, HTT, THT, TTH, THH, HTH, HHT\}$

$A = \{THH, HTH, HHT\}$

$P(A) = 3/8$

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

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

a) Equal to 1

Ans : 0

b) Less than or equal to 4

Ans : $6/36 = 1/6 = 0.166$

c) Sum is divisible by 2 and 3

Ans : $6/36 = 1/6 = 0.16$

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 = $2+3+2=7$

$n(S) = 7$

First, let's find the number of ways to draw 2 balls that are not blue:

Number of ways to choose 2 non-blue balls = (2 red + 3 green) choose 2
 $= (2 + 3) \text{ choose } 2$

$$\begin{aligned}
&= 5 \text{ choose } 2 \\
&= 5! / (2! * (5 - 2)!) \\
&= (5 * 4) / (2 * 1) \\
&= 10 \text{ ways}
\end{aligned}$$

find the total number of ways to draw 2 balls from the 7 balls in the bag:

$$\begin{aligned}
&\text{Total number of ways to choose 2 balls from 7} = 7 \text{ choose } 2 \\
&= 7! / (2! * (7 - 2)!) \\
&= (7 * 6) / (2 * 1) \\
&= 21 \text{ ways.}
\end{aligned}$$

The probability of drawing 2 balls that are not blue is the number of ways to choose 2 non-blue balls divided by the total number of ways to choose 2 balls:

$$\begin{aligned}
&\text{Probability} = (\text{Number of ways to choose 2 non-blue balls}) / (\text{Total number of ways to choose 2 balls}) \\
&= 10 / 21
\end{aligned}$$

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 : Child A: 1 candy * Probability = 1 * 0.015 = 0.015

Child B: 4 candies * Probability = 4 * 0.20 = 0.80

Child C: 3 candies * Probability = 3 * 0.65 = 1.95

Child D: 5 candies * Probability = $5 * 0.005 = 0.025$

Child E: 6 candies * Probability = $6 * 0.01 = 0.06$

Child F: 2 candies * Probability = $2 * 0.120 = 0.24$

Expected number of candies = $0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24 = \mathbf{3.09}$

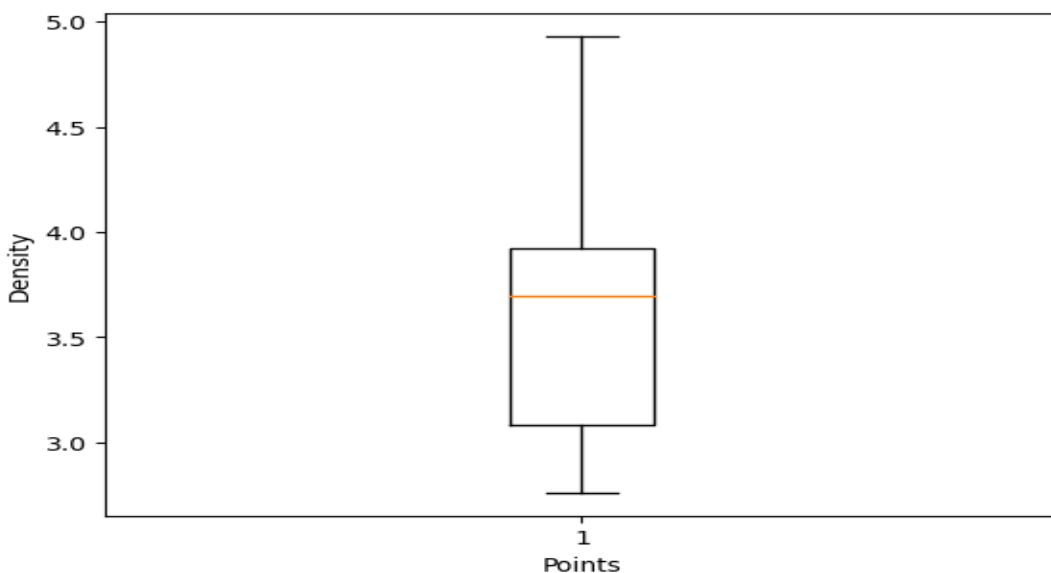
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

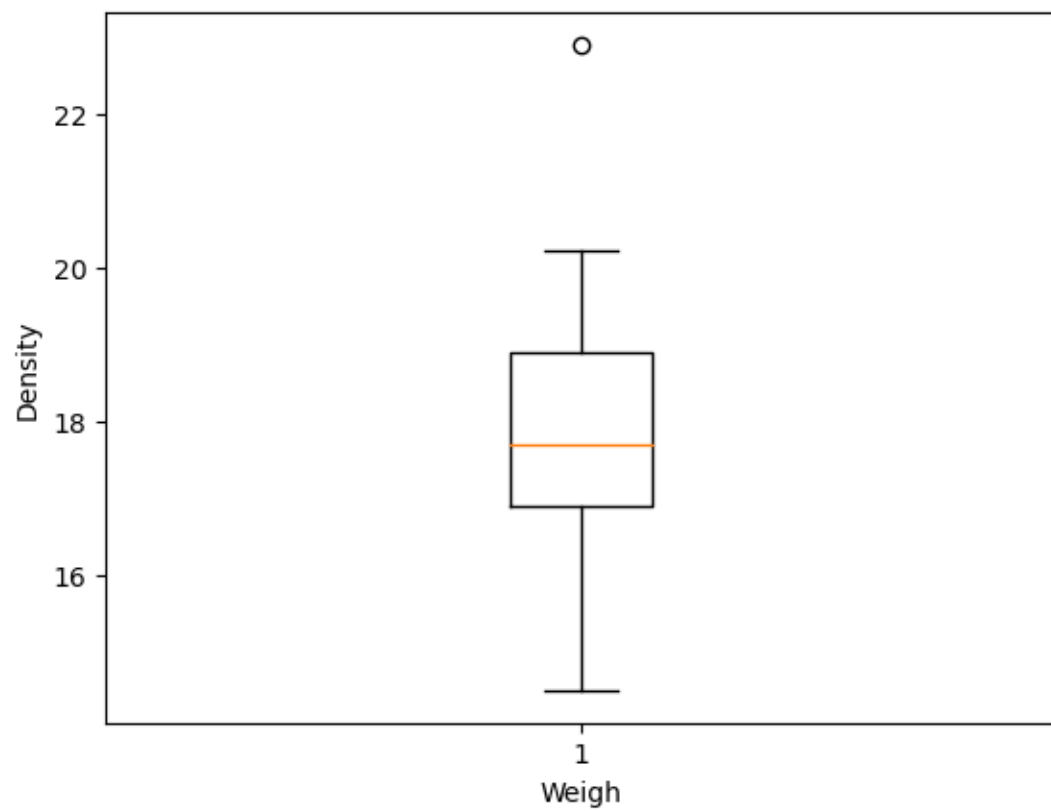
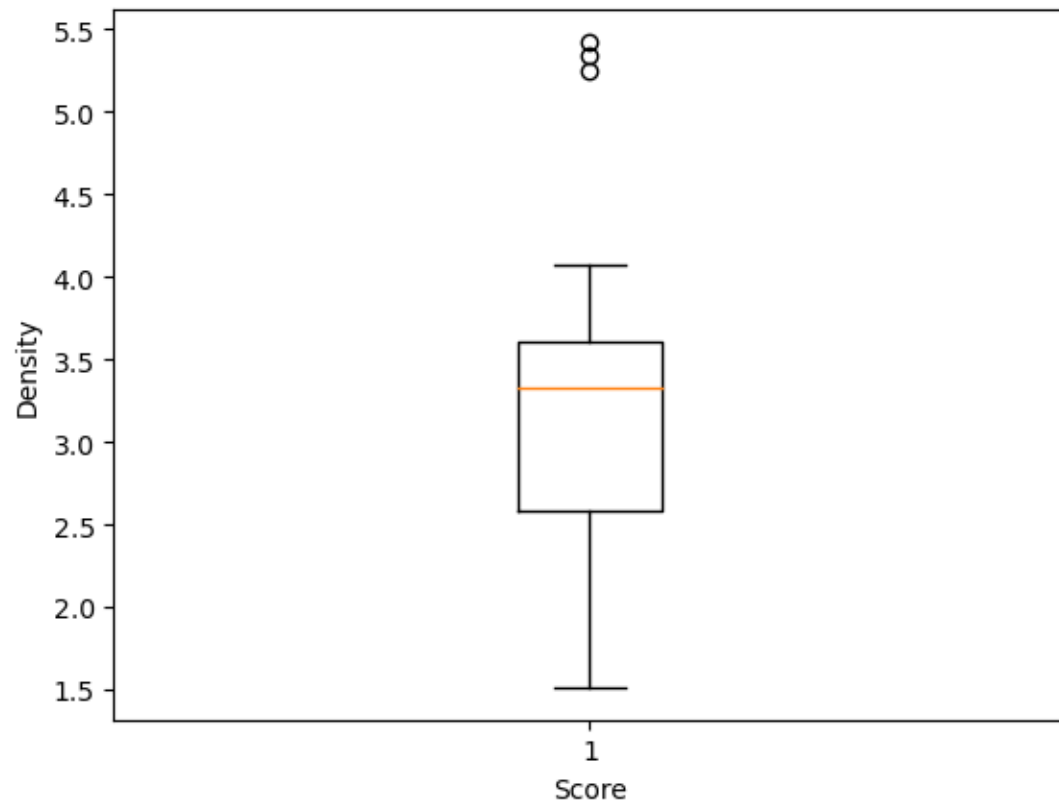
- For Points, Score, Weigh>
Find Mean, Median, Mode, Variance, Standard Deviation, and Range
and also Comment about the values/ Draw some inferences.

Use Q7.csv file

Ans: 1. Mean of Points: 3.596 Score: 3.217 Weigh: 17.848
2. Median of Points: 3.695 Score: 3.325 Weigh: 17.710
3. Mode of Points : 3.07 Score: 3.44 Weigh: 17.02
4. Variance of Points: 0.285881 Score: 0.957379 Weigh: 3.193166
5. Standard Deviation of Points : 0.534679 Score: 0.978457 Weigh: 1.7869
6. Range of Points(Min-Max) : (2.760-4.930), Score: (1.513 – 5.424) ,
Weigh: (14.500-22.900)

Draw inferences:





Q8) Calculate Expected Value for the problem below

a) 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 : Given the weights:

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

Sum of all values = $108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 = 1328$

Number of values = 9

Expected Value = $1328 / 9 \approx 145.33$

So, the expected value of the weight of a patient is approximately 145.33 pounds.

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

Cars speed and distance

Use Q9_a.csv

Ans : `q9= pd.read_csv('Q9_a.csv')`
 `q9`
 `print('For Cars Speed', 'Skewness value', np.round(q9.speed.skew(),2),`
`'and' , 'Kurtosis value =', np.round(q9.speed.kurt(),2))`

For Cars Speed Skewness value -0.12 and Kurtosis value = -0.51

`print('For Cars Distance', 'Skewness value', np.round(q9.dist.skew(),2), 'and' , 'Kurtosis value =', np.round(q9.dist.kurt(),2))`

For Cars Distance Skewness value 0.81 and Kurtosis value = 0.41

SP and Weight(WT)

Use Q9_b.csv

```
q9b = pd.read_csv('Q9_b.csv')
```

```
q9b
```

```
q9b.rename(columns ={'Unnamed: 0': 'Index'}, inplace = True)
```

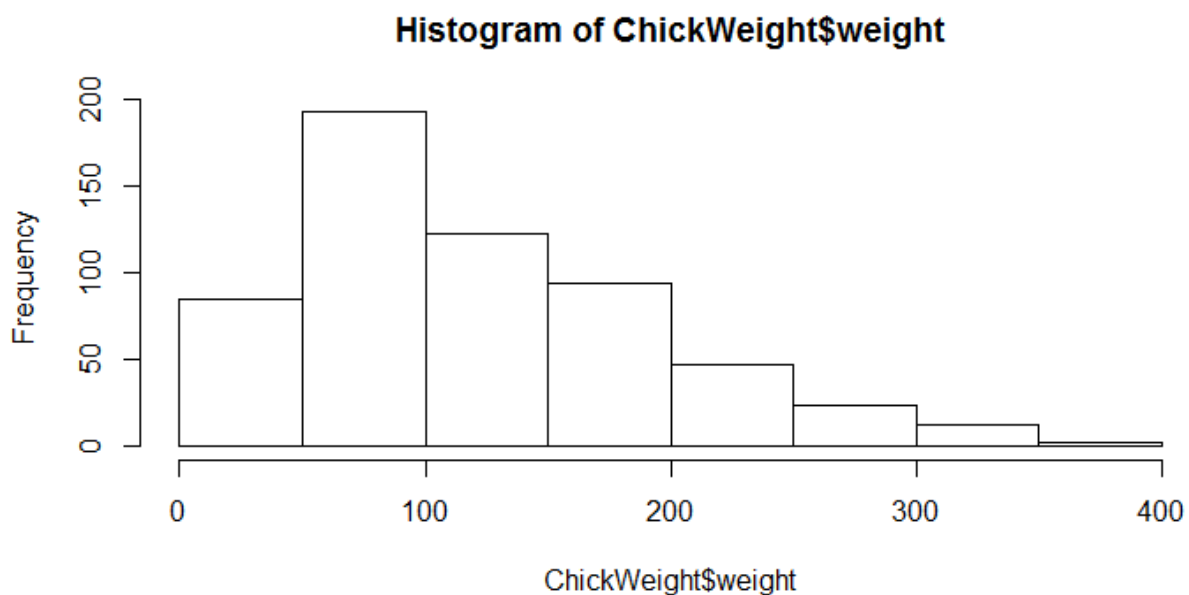
```
print('For SP Skewness= ', np.round(q9b.SP.skew(),2), 'Kurtosis =',  
np.round(q9b.SP.kurt(),2))
```

For SP Skewness= 1.61 Kurtosis = 2.98

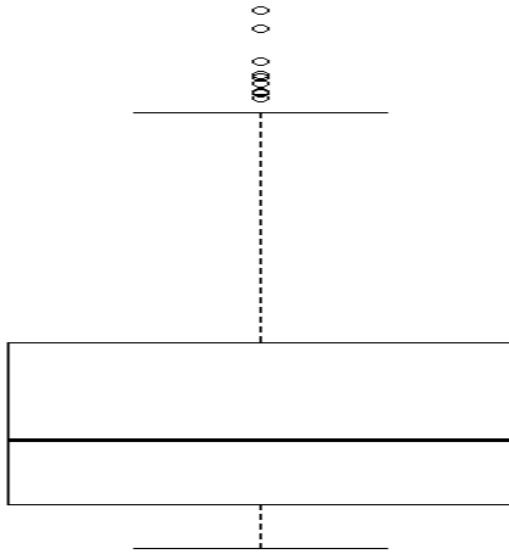
```
print('For WT Skewness= ', np.round(q9b.WT.skew(),2), 'Kurtosis =', np.round(q9  
b.WT.kurt(),2))
```

For WT Skewness= -0.61 Kurtosis = 0.95

Q10) Draw inferences about the following boxplot & histogram



Answer: The histogram peak has right skew and tail is on right . Mean > Median
We have an outlier on higher side.



Answer: The boxplot has outlier 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?

Answer: `conf_94 = stats.t.interval(alpha = 0.94, df=1999, loc=200, scale=30/np.sqrt(2000))`
`print(np.round(conf_94,0))`
`print(conf_94)`

For 94% confidence interval Range is [198.73 – 201.26]

`conf_98 = stats.t.interval(alpha = 0.98, df=1999, loc=200, scale=30/np.sqrt(2000))`
`print(np.round(conf_98,0))`
`print(conf_98)`

For 98% confidence interval Range Is [198 . 202 .]

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.

Answer: mean : 41 median : 40.5 variance : 25.529411764705884
standard deviation : 5.05266382858645

2) What can we say about the student marks?

Answer: we don't have any outliers in the data and 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?

Ans: When the mean median of dataset are equal then there will be a symmetric distribution with minimal or no skewness

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

Ans: When the mean is greater than the median, the distribution is positively skewed with a longer tail on the right.

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

Ans: when the median is greater than the mean, it suggests a negatively skewed distribution with a longer tail on the left-hand side.

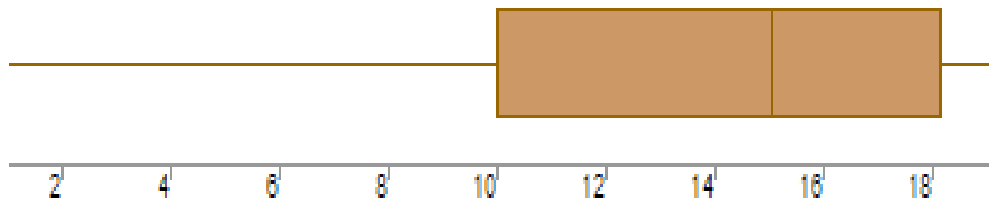
Q16) What does positive kurtosis value indicates for a data ?

Ans: A positive kurtosis value indicates that the peak of the data distribution is more pronounced or sharper than that of a normal distribution.

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

Ans: A negative kurtosis value indicates that the peak of the data distribution is less pronounced or flatter than that of a normal distribution.

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



What can we say about the distribution of the data?

Answer: The above boxplot is not normally distributed the median towards the higher value .

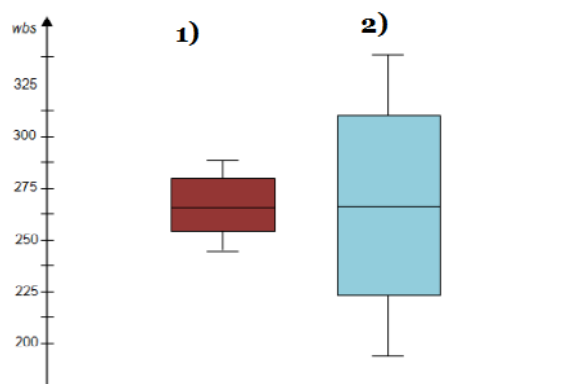
What is nature of skewness of the data?

Answer: data is skewed towards left and the whiskers range of minimum value is greater than maximum value.

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

Answer: The Inter Quantile Range = $Q3 - Q1$ Upper quartile = 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.

Answer: firstly there is an no outliers in this data and secondly both the boxplot shares a same median that is approximately in the range of between 275 to 250

and they normally distributed with zero to no skewness neither at the minimum or maximum whiskers 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.

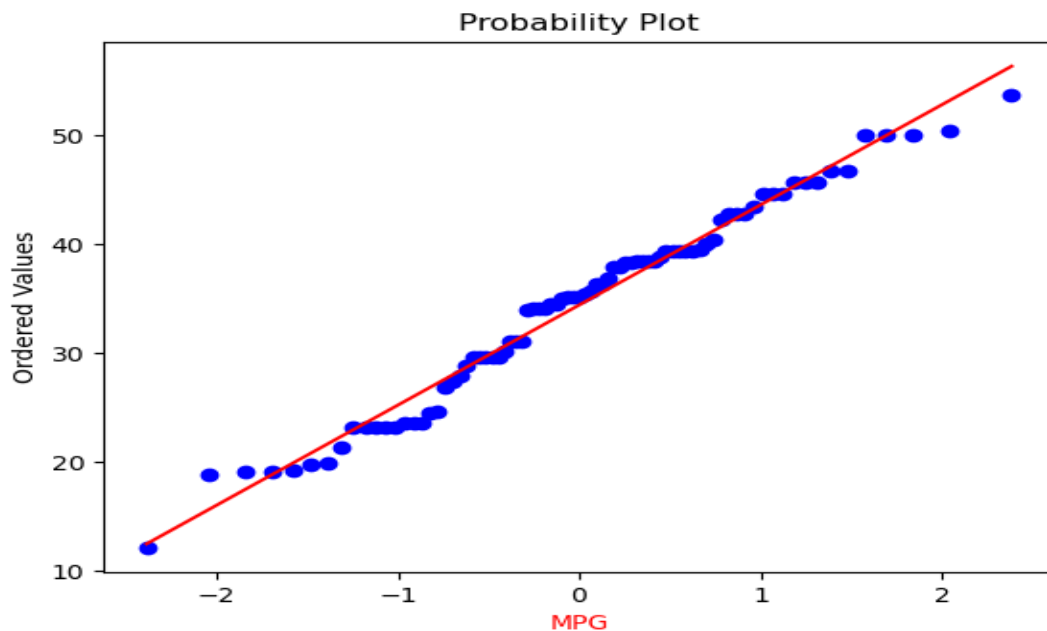
```
MPG <- Cars$MPG
```

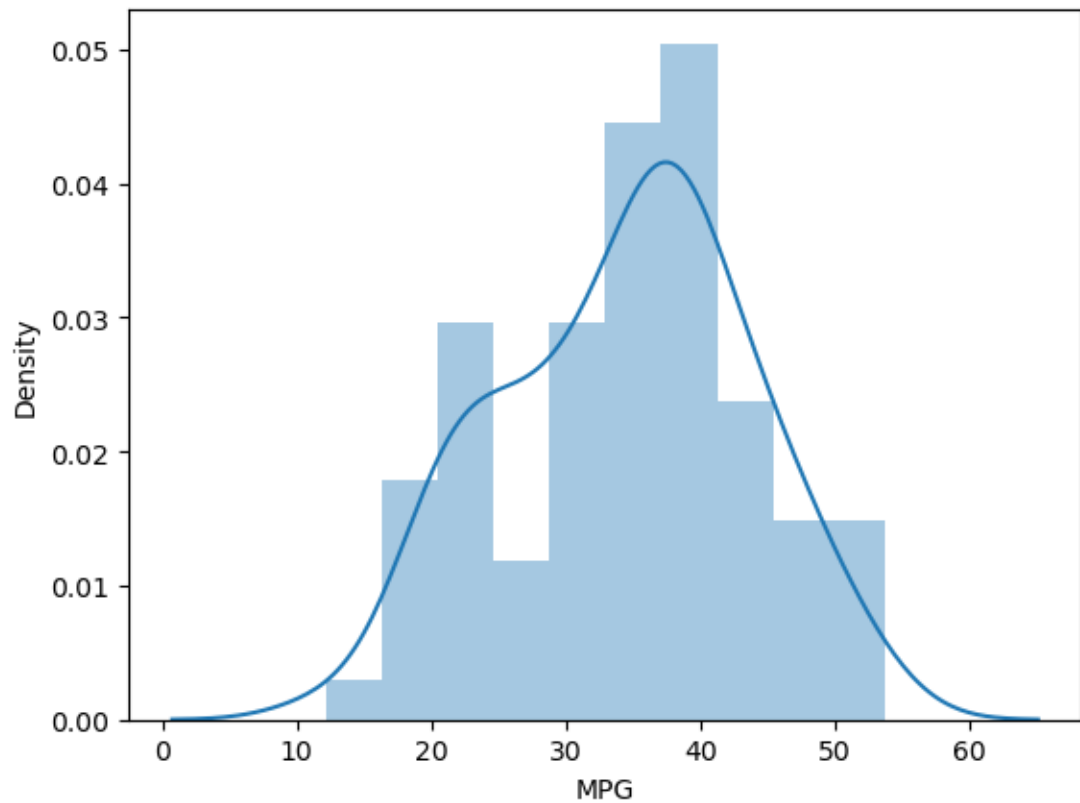
- a. $P(\text{MPG} > 38)$
- b. $P(\text{MPG} < 40)$
- c. $P(20 < \text{MPG} < 50)$

Q 21) Check whether the data follows normal distribution

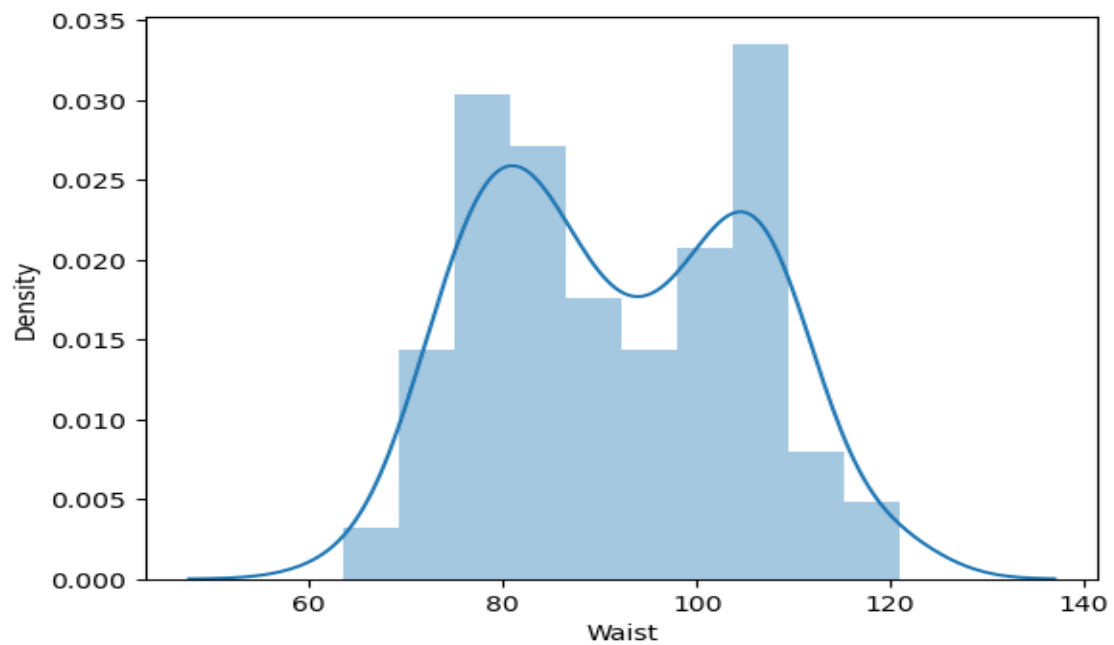
a) Check whether the MPG of Cars follows Normal Distribution

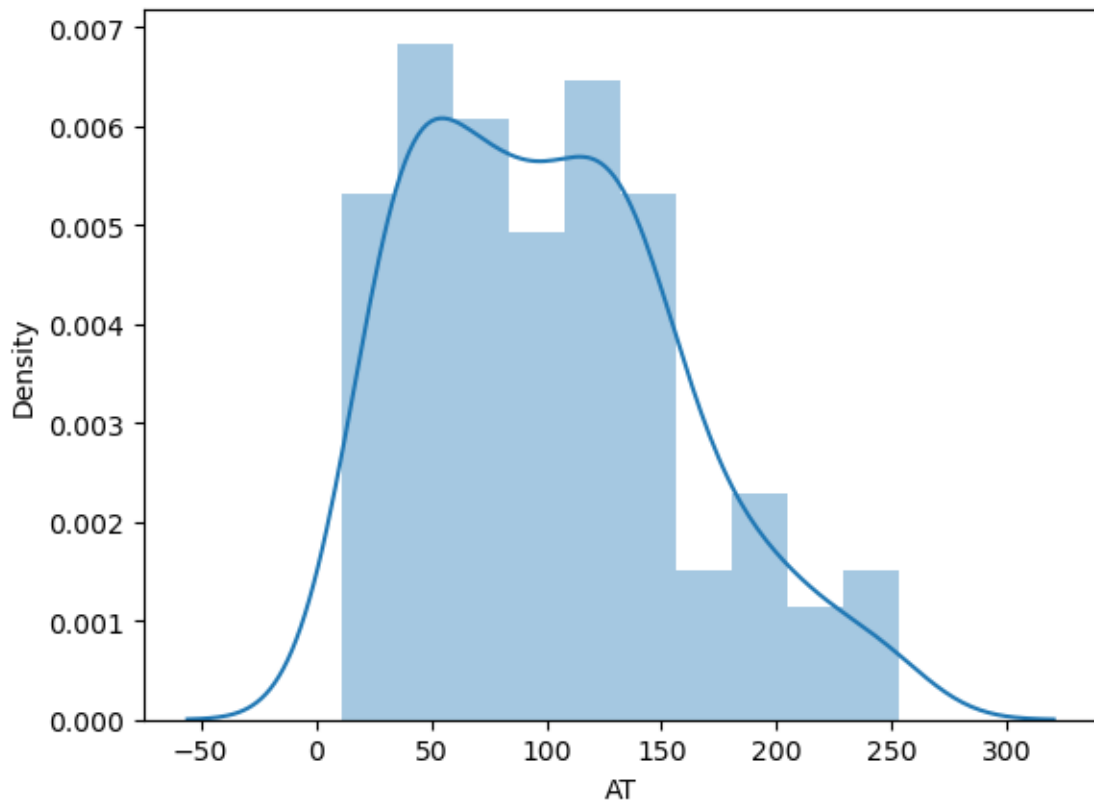
Dataset: Cars.csv





b) 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

Answer: # z value for 90% confidence interval

```
print('Z score for 60% Confidence Intervla =', np.round(stats.norm.ppf(.05), 4))
```

Z score for 60% Confidence Intervla = -1.6449

z value for 94% confidence interval

```
print('Z score for 60% Confidence Intervla =', np.round(stats.norm.ppf(.03), 4))
```

Z score for 60% Confidence Intervla = -1.8808

z value for 60% confidence interval

```
print('Z score for 60% Confidence Intervla =', np.round(stats.norm.ppf(.2), 4))
```

Z score for 60% Confidence Intervla = -0.8416

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

Answer: # t value for 95% confidence interval

```
print('T score for 95% Confidence Interval = ', np.round(stats.t.ppf(0.025, df=24), 4))
```

T score for 95% Confidence Interval = -2.0639

value for 94% confidence interval

```
print('T score for 94% Confidence Interval = ', np.round(stats.t.ppf(0.03, df=24), 4))
```

T score for 94% Confidence Interval = -1.974

t value for 99% confidence interval

```
print('T score for 99% Confidence Interval = ', np.round(stats.t.ppf(0.005, df=24), 4))
```

T score for 99% Confidence Interval = -2.7969

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: To determine the probability that 18 randomly selected bulbs would have an average life of no more than 260 days, we can use the concept of the sampling distribution of the sample mean. If the CEO's claim were true, the distribution of the sample means would follow a normal distribution with the same mean and a standard deviation equal to the population standard deviation divided by the square root of the sample size.

Given information:

- Population mean (μ) = 270 days
- Sample size (n) = 18
- Sample mean (\bar{x}) = 260 days
- Sample standard deviation (σ) = 90 days

Calculate the standard error of the mean:

standard error of the mean is given by the formula: $SE = \sigma / \sqrt{n}$

```
In [22]: import math

population_mean = 270
sample_size = 18
sample_mean = 260
sample_std_dev = 90

standard_error = sample_std_dev / math.sqrt(sample_size)
```

```
In [23]: standard_error
```

```
Out[23]: 21.213203435596427
```