|  |  |
| --- | --- |
| aActivity | 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 | Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Ratio |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ratio |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| 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?

ANS:

The Total number of possible outcomes is 23 = 8

S=HHH, TTT, HTT, HHT, THH, TTH, THT, HTH

Number of favorable Outcome=3

P(two H and one T)= 3/8= 0.375

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

ANS:

a) There is no outcomes through which sum is equal to one. Probability is 0.

b) 3 outcomes, 3/36=1/12

c) 6/36= 1/6

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?

ANS:

P (2R, 3G, 2B)

P (5/7, 4/6) = 20/42

= 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:

Expected number of candies for randomly selected child = 1x0.015+ 4x0.20+ 3x0.65+ 5x0.005+ 6x0.01 +2x0.120 =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**

ANS:

* **Points**:

Mean =3.596563, Median= 3.695, Mode= “numeric”,

Variance= 0.2858814, Standard deviation= 0.5346787.

* **Score:**

Mean= 3.21725, Median= 3.325, Mode= “numeric”,

Variance= 0.957379, Standard deviation= 0.9784574

* **Weight:**

Mean= 17.84875, Median= 17.71, Mode= “numeric”,

Variance= 3.193166, Standard deviation= 1.786943

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:

Expected value= Σ[x.p(x)] = (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

ANS:

Speed= -0.117510

Dist=0.806895

Skewness for speed= -0.117510(slightly left skewed)

for distance= 0.8068935(right skewed)

Kurtosis

Index=-1.200000

Speed=-0.508994

Dist=0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness

SP=1.611450

WT=-0.614753

Kurtosis

Unnamed: 0 -1.200000

SP=2.977329

WT=0.950291

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



Ans:

We can notice a long tail towards right so it is heavily right skewed. Therefore, there are outliers lying on the higher side.



Ans:

There are outliers on the upper side and there are less data points between Q1 and bottom point of boxplot.

**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:

X+/-(Z1- α. σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1- σ/2) = 1-0.03 =0.97

1. Confidence interval for 94%= 1.882
2. Confidence interval for 98%= 2.33
3. Confidence interval for 96% = 2.05

**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?

ANS:

1. Mean= 41
2. Median= 40
3. Variance= 24.111
4. Standard deviation= 4.910

There are no outliers.

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

ANS:

Symmetrical

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

ANS:

Right Skewed

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

ANS:

Left Skewed

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

ANS:

The data is normally distributed. Kurtosis value is 0.

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

ANS:

The curve of the given data will have lighter tails and flatter peaks.

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



What can we say about the distribution of the data?

ANS:

The above boxplot is not normally distributed. The value of median is high.

What is the nature of skewness of the data?

ANS:

As median is greater than mean, it is left skewed.

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

ANS:

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 both plots, whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical. Also, there are no outliers.

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:

a) 33 observations in MPG are greater than 38 (33/81)

b) 67 observations in MPG are lesser than 40 (67/81)

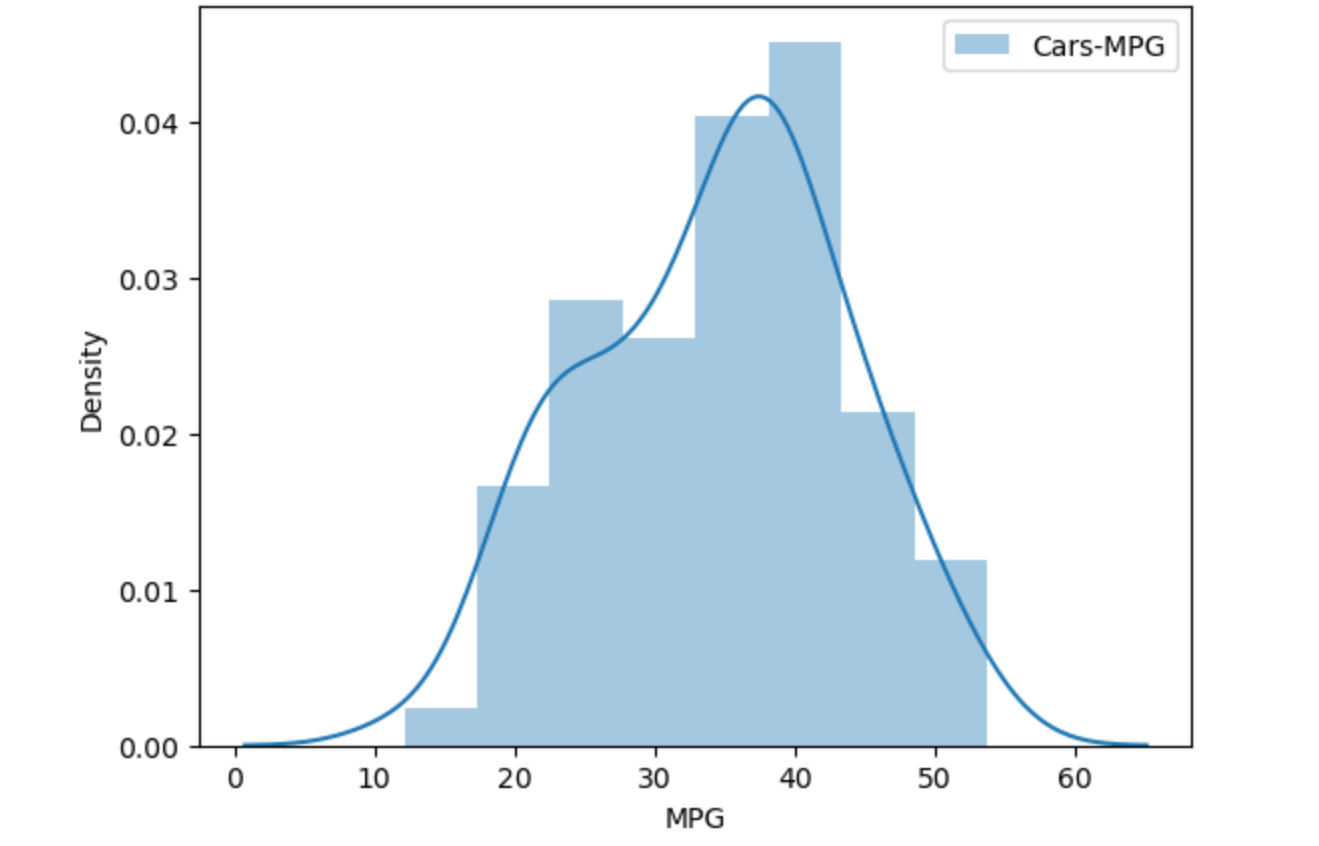
c) 69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

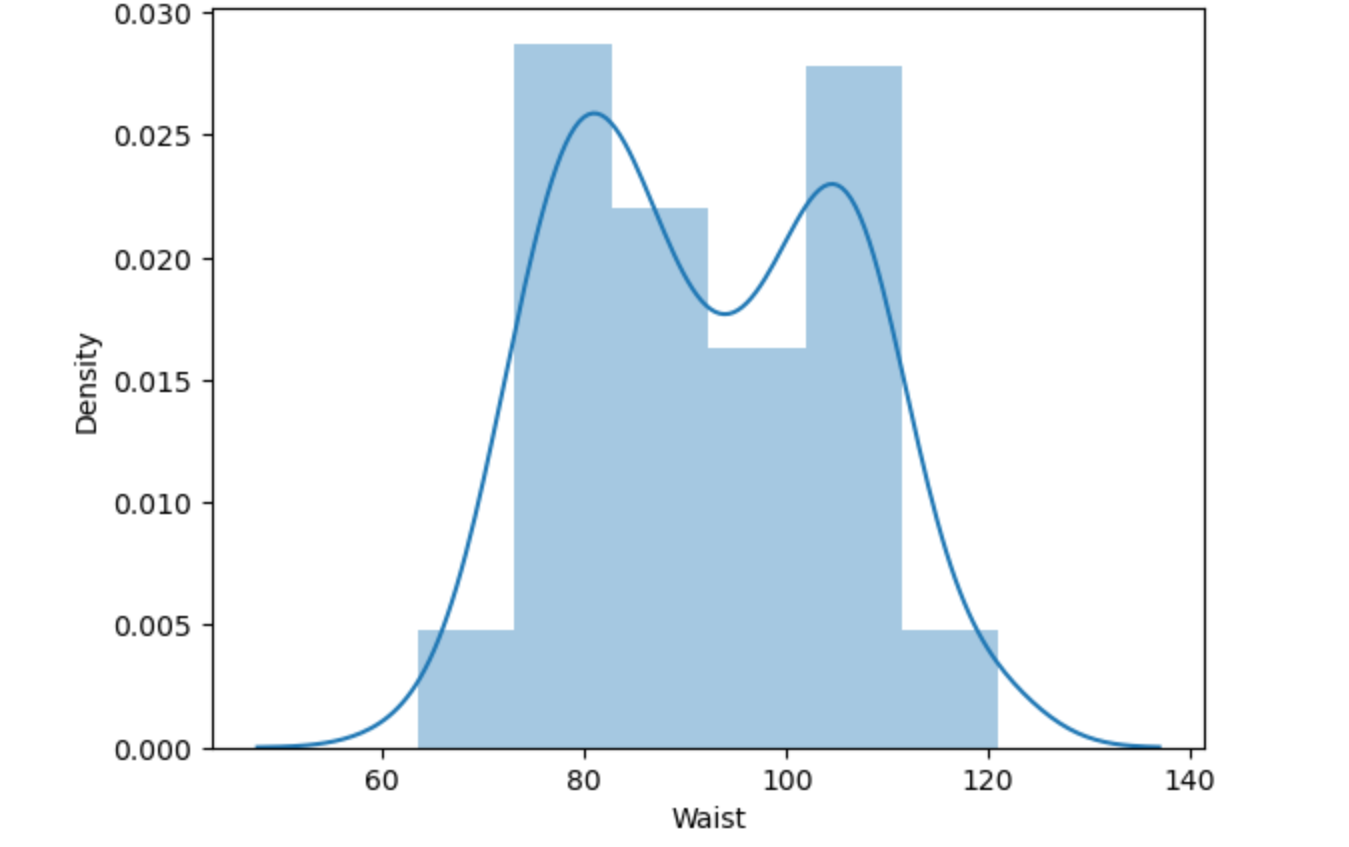
ANS: MPG of cars follows normal distribution

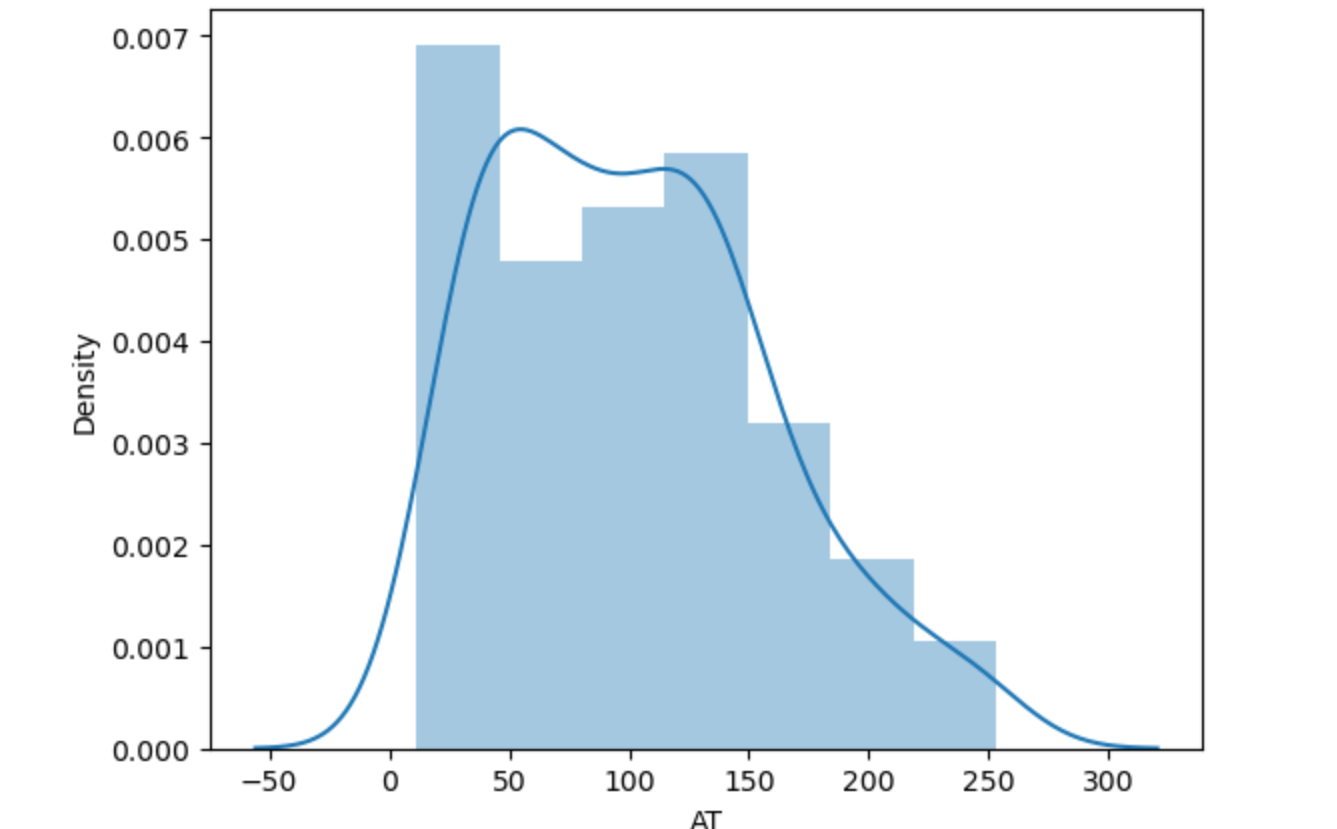


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

ANS: Adipose Tissue (AT) and Waist does not follow Normal Distribution





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

ANS:

1. Z score of 90% confidence interval = 1.65
2. Z score of 94% confidence interval = 1.55
3. Z score of 60% confidence interval = 0.85

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

ANS:

* t score of 95% confidence interval = 1.96
* t score of 96% confidence interval = 2.5
* t score of 99% confidence interval = 2.47

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: 0.32