|  |  |
| --- | --- |
| 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 | Nominal |
| 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

**ANS:** Total Number ofPossible Combination for 3 coins are Tossed is 2^3=8

Possible Combinations are:

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

Number of combinations that have two heads and one tail : HHT,HTH,THH

Probability = (Number of favorable outcomes) / (Total number of possible outcomes)

Probability = 3/8

Probability = 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:** Each dice has 6 possible combinations

Total Number ofPossible Combination for rolling 2 dice 6\*6=36

Possible combinations are:

(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:**

when rolling two dices, possible outcome for 1 dice rolled is in range 1 to 6 so sum of two dices never be 1 coz minimum outcome of single dice is 1.

**b)Less than or equal to 4:**

possible combinations for less than or equal to 4 is: (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

Probability = (Number of favorable outcomes) / (Total number of possible outcomes)

Probability = 6/36

Probability = 0.166

**c)Sum is divisible by 2 and 3:**

possible combinations for sum is divisible by 2 and 3:

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

Probability = (Number of favorable outcomes) / (Total number of possible outcomes)

Probability = 6/36

Probability = 0.166

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 ball: 2(red)+3(green)+2(blue)=7

Number of ways of drawing 2 balls out of 7

7C2= 7\*6/2\*1

= 42/2

= 21

Total no of ball other than blue: 2(red)+3(green)=5

Number of ways of drawing 2 balls out of 5

5C2= 5\*4/2\*1

= 20/2

= 10

Probability that none of the balls drawn is blue :10/21=0.476

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:**

Multiply each candy count by its corresponding probability for each child.

Expected number of candies =(Probability of Child A)\*(Candies count for Child A) +

(Probability of Child B) \* (Candies count for Child B) +

(Probability of Child C) \* (Candies count for Child C) +

(Probability of Child D) \* (Candies count for Child D) +

(Probability of Child E) \* (Candies count for Child E) +

(Probability of Child F) \* (Candies count for Child 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 = 3.09

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.

**ANS:**

Mean: Median:

Points 3.596563 Points 3.695

Score 3.217250 Score 3.325

Weigh 17.848750 Weigh 17.710

Std: Var:

Points 0.534679 Points 0.285881

Score 0.978457 Score 0.957379

Weigh 1.786943 Weigh 3.193166

Range: Mode:

Points 2.17 Points 3.92

Score 3.9110000000000005 Score 3.44

Weigh 8.399999999999999 Weigh 17.02

Inferences: The columns points and Weigh is multimodal For the Points and score Median> mean , hence the distribution is Right skewed distribution For Weigh Mean> Median, hence the distribution is Left skewed distribution

**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?

**ANS:**

Expected Value = Sum of Weights / Number of Weights

Expected Value = (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199) / 9

Expected Value = 1308 / 9

Expected Value ≈ 145.33

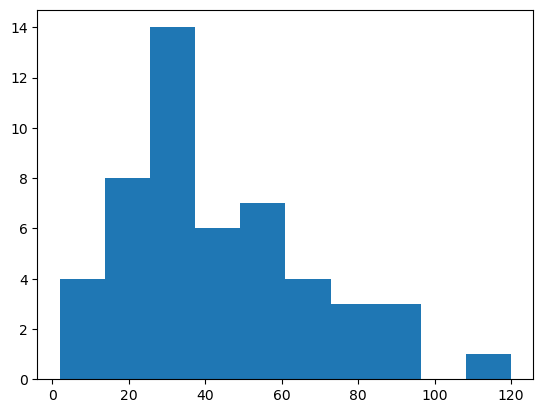
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

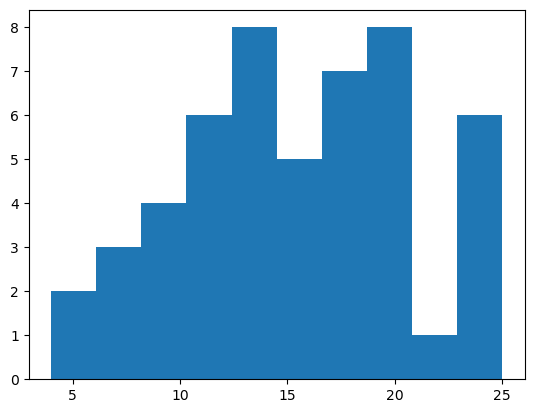
**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | Skeweness | Kurtosis |
| Car speed | -0.117510 | -0.508994 |
| distance | 0.806895 | 0.405053 |

Inference: from below histogram the Distance data is right skewed means its positive



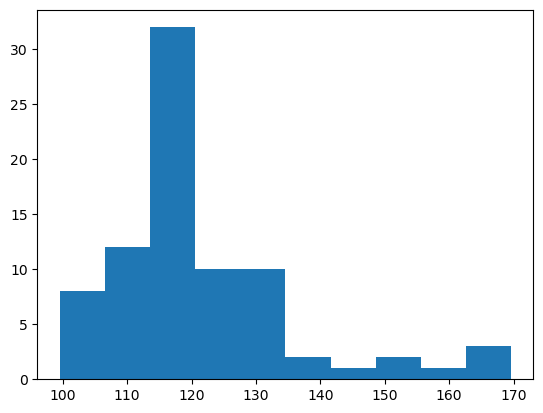
 From below histogram the speed data is left skewed means its negative.



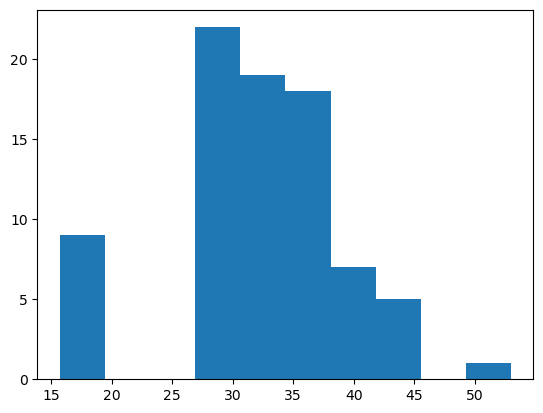
**SP and Weight(WT)**

**Use Q9\_b.csv**

SP and Weight(WT) From the below graph we can infer that the SP data is right skewed and (+ve kurtosis)..



the WT data is left skewed (+ve kurtosis)



|  |  |  |
| --- | --- | --- |
|  | Skeweness | Kurtosis |
| Car speed | 1.611450 | 2.977329 |
| distance | -0.614753 | 0.950291 |

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



1. Histogram of chick Weight is Right Skewed we also called is as positively skewed.
2. Highest chick weight is between 50 to 100
3. Most of the chicken weight is between 50 to 150



1)There are outliers presents 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?

**Ans:**

Sample mean:200

Sample Standard Deviation:30

Sample Size:2000

Confidence Interval = Sample Mean ± (Z \* (Sample Standard Deviation / √Sample Size))

For a 94% confidence interval:

Z = 1.880

Confidence Interval = 200 ± (1.880 \* (30 / √2000))

Confidence Interval ≈ 200 ± 1.26

Confidence Interval ≈ (198.74, 201.26)

For a 96% confidence interval:

Z = 2.05

Confidence Interval = 200 ± (2.05 \* (30 / √2000))

Confidence Interval ≈ 200 ± 1.37

Confidence Interval ≈ (198.63, 201.37)

For a 98% confidence interval:

Z = 2.330

Confidence Interval = 200 ± (2.330 \* (30 / √2000))

Confidence Interval ≈ 200 ± 1.56

Confidence Interval ≈ (198.44, 201.56)

**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.

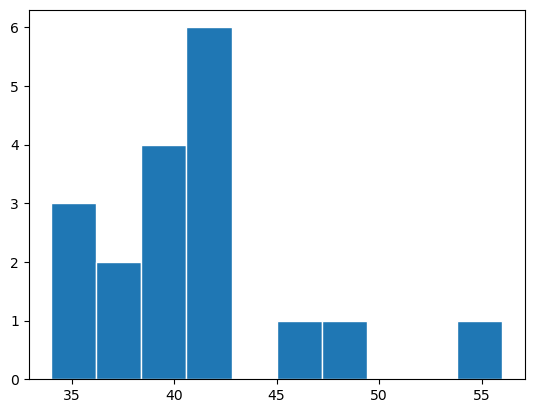
**ANS:** Mean: 41.0

Median: 40.5

Variance: 25.52

Std: 5.052

1. What can we say about the student marks?

**Ans: **

1. Most of the student get marks between 38 to 42
2. It is right skewed students marks histogram which is also called positively skewed.

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

**ANS:** Data is normally distributed. there is no skewness to the data

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

**ANS:** When the mean of a data is greater than the median .the nature of skewness is Positive in nature which is Right skewed distribution

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

**ANS:** When the median of a data is greater than the mean . the nature of skewness is Negative in nature which is Left skewed distribution

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

**ANS:** Positive kurtosis for a data Implies that the data is of more peaked distribution than the normal distribution

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

**ANS:** Negative kurtosis for a data Implies that the data is of flatter distribution than the normal distribution

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



What can we say about the distribution of the data?

**ANS:**Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

**ANS:**The data is a skewed towards left.

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

**ANS:** Inter Quartile Range = 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.

**ANS**: 1) Both the boxplots do not contain any outliers

2) both the boxplot indicates same Median

3) Range of both the boxplot median is approximately in between 250 to 275

4) They are normally distributed

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)

**ANS:** (1 - stats.norm.cdf(38, loc= df.MPG.mean(), scale= df.MPG.std()))

0.34759392515827137

* 1. P(MPG<40)

**ANS:** (stats.norm.cdf(40, loc= df.MPG.mean(), scale= df.MPG.std()))

0.7293498762151609

* 1. P (20<MPG<50)

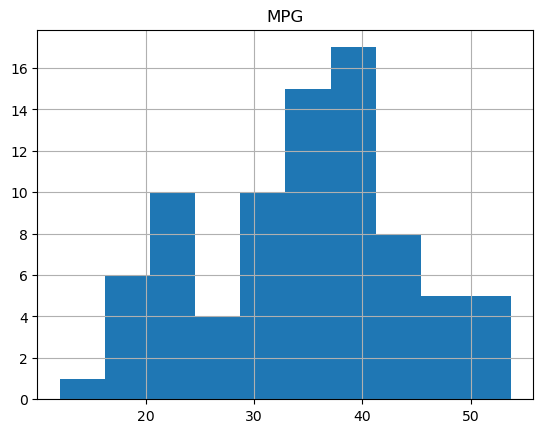
**ANS**: stats.norm.cdf(0.50,Cars.MPG.mean(),Cars.MPG.std())-stats.norm.cdf(0.20,Cars.MPG.mean(),Cars.MPG.std())

1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**ANS: **

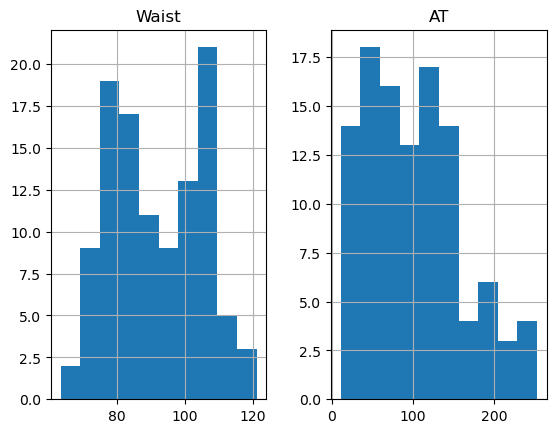
**Cars['MPG'].mean() =** 34.42207572802469

**Cars['MPG'].median() =** 35.15272697

 Median >mean . hence it is left skewed 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**: data['Waist'].median() =90.8

data['Waist'].mean() =91.90

data['AT'].mean()= 101.8

data['AT'].median()= 96.54

Mean > Median , The distribution is Right Skewed distribution.

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

**Ans:**

stats**.**norm**.**ppf(0.95)

For a 90% confidence interval Z-score ≈ 1.645

stats**.**norm**.**ppf(0.97)

For a 94% confidence interval Z-score ≈ 1.880

stats**.**norm**.**ppf(0.8)

For a 60% confidence interval Z-score ≈ 0.841

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

**Ans:**

stats**.**t**.**ppf(0.975,24)

For a 95% confidence interval t-score ≈ 2.064

stats**.**t**.**ppf(0.98,24)

For a 96% confidence interval t-score ≈ 2.172

stats**.**t**.**ppf(0.9,24)

For a 99% confidence interval t-score ≈ 2.797

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:

population mean = 270

sample size = 18

sample mean = 260

sample std = 90

df = sample size - 1

t = (sample mean - population mean) / (sample std / (√sample size ))

=(260-270)/(90/√18)

probability = stats.t.cdf(t, df)

probability= 0.32167253567098364