|  |  |
| --- | --- |
| 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 | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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. Sample space = total outcomes =2 ^3= 8

{hhh,hht,hth,thh,tth,tht,htt,ttt}

Sample having two heads and one tail = 3

Therefore probability of getting 2 head and one tail is 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 = here n(s) = 36 i.e sample space

When two dice are rolled, sample space is given as:

(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)

Therefore, Total number of possible outcomes = 36

Favorable Cases = Sum is equal to 1 = 0

No Probability means 0 probability.

1. Probability of sum = 1 is 0,since dices start from 1.
2. Probability of sum is less or equal to 4 is 6/36= 1/6.

**:** When two dice are rolled, sample space is given as:

(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)

Therefore, Total number of possible outcomes = 36

1. Favorable Cases = Sum is Less than or equal to 4

= (1,3) (2,2) (3,1) (1,2) (2,1) (1,1)

Probability = 6/36 = 1/6

1. Probability of Sum is divisible by 2 and 3 = 6/36=1/6.

When two dice are rolled, sample space is given as:

(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)

Therefore, Total number of possible outcomes = 36

Favorable outcomes = sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1,5),(2,4),(3,3),(4,2),(5,1),(6,6) Therefore,

Number of favorable outcomes = 6

The probability of an event is given as:

Probability = number of favorable outcomes

Number of possible outcomes

Probability = 6/36

Thus the probability that sum is divisible by 2 and 3 is 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:- Total number of balls = (2 + 3 + 2) = 7

Let ‘S’ be the sample space

To find number of different combinations nCr = n! / r!(n−r)!

Then, n(S) = Number of ways of drawing 2 balls out of 7 = 7C2

n(S) = 7! / 2!(7-1)! = (7×6) / (2×1) = 21

Let E = Event of drawing 2 balls in which none of them is blue.

Therefore, n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls.

N(E) = 5C2 = (5×4) / (2×1) = 10

P(E) = n(E) / n(S)= 10 / 21 = 0.47

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 a randomly selected child should be

= 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.

**Use Q7.csv file**

Ans : Points:-

Mean = 3.596563

Median = 3.695

Mode = 3.07

Variance = 0.285881

Standard Deviation = 0.534679

Range = 2.17

2.Score :

Mean = 3.21725

Median = 3.325

Mode = 3.44

Variance = 0.957379

Standard Deviation = 0.978457

Range = 3.911

3. Weigh

Mean = 17.84875

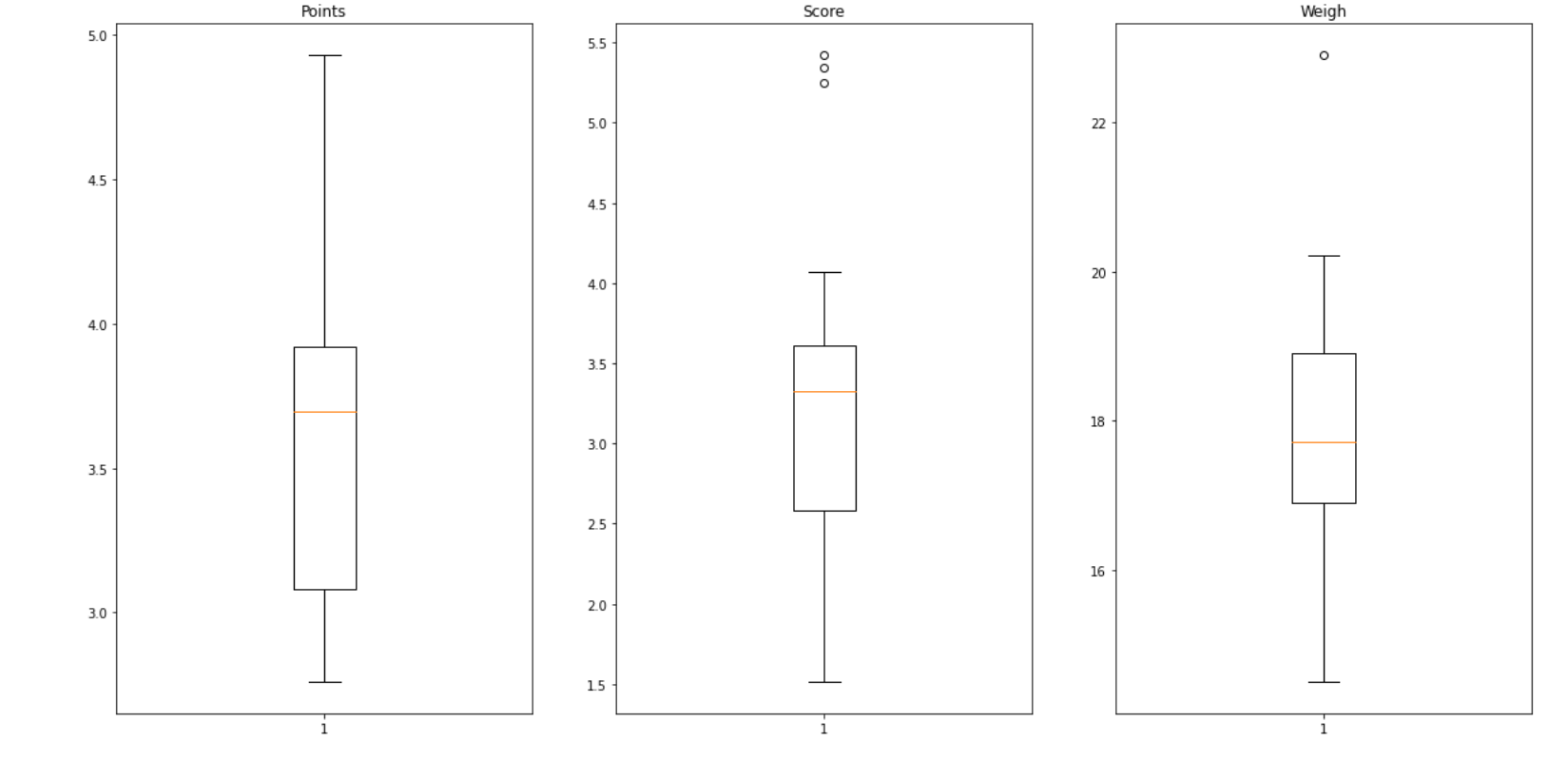
Median = 17.71

Mode = 17.02

Variance = 3.193166

Standard Deviation = 1.786943

Range = 8.4



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- Calculate mean of the given data that will be the weight of the random chosen person.

X= 108+110+ 123+ 134+135+ 145+ 167+ 187+ 199/9

= 1308/9=145.33

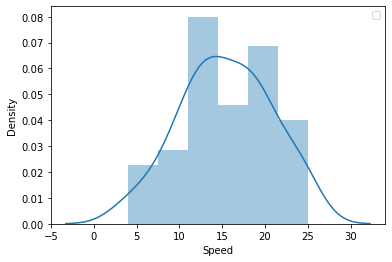
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

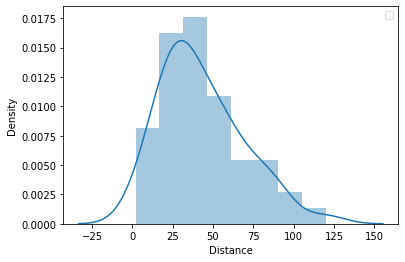
**Cars speed and distance**

**Use Q9\_a.csv**

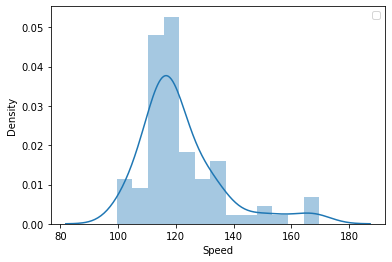
**SP and Weight(WT)**

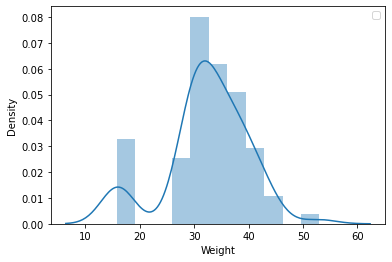
**Use Q9\_b.csv.**

Ans:-Inference:- 



**B:-**





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

Ans- The histogram is right skewed i.e it has Its tail towards the right side.The mass of the data is concent rated towards the left side of the histogram.



Ans- The box plot given here is again positively(Rightly) skewed it show us the measure of asymmetry.the mass of the data is concentrated toward the left side of the centre of the box plot It also shows us some outliers shown by the the smalls dots above the box plot.

**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:- Confidence interval for 94% is :

(198.43943840429978, 201.56056159570022)

* Confidence interval for 98% is :

(198.43943840429978, 201.56056159570022)

* Confidence interval for 96% is :

(198.62230334813333, 201.37769665186667)

**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 ,Median = 40.5.

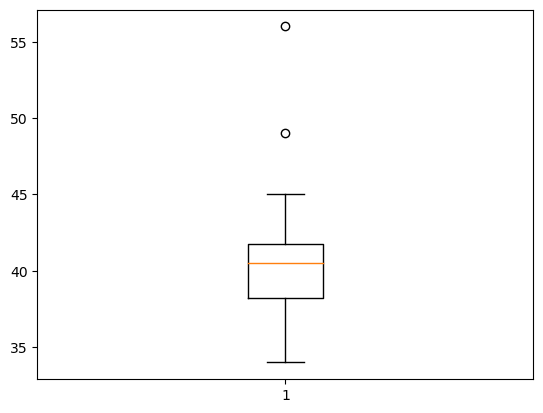
Mode =41, variance= 25.529411.

Standard deviation = 40.5.

2).There are 2 outliers in the marks list

i.e 49 and 56.

Command=plt.boxplot(marks)



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

Ans-If the Mean and Median are equal means the skewness of the data is zero .Means the data is distributed Symmetrically ex-Normal distribution.

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

Ans-If mean > Median the data is Positively skewed.

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

Ans- if the mean is less than the median, the distribution is negatively skewed.

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

Ans- Positive kurtosis means the data is peaked,the tail of the graph are concentrated towards the mean of the data.(Leptokurtic)

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

Ans-Negative kurtosis the data has wider peaks (Platykurtic).Distribution of the data has lighter tail & 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?

Ans-If we consider the above shown boxplot is of marks of students in a class then we can say that , 50% of students in a class having marks more than 10 & remaining are less.

What is nature of skewness of the data?

Ans- it is negative skewed(Median>Mean)

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

Ans- Its interquartile range(IQR) lies between (10-18).

Q19) Comment on the below Boxplot visualizations?



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

Ans- 1.The boxplot 1 and the boxplot 2 have the same median i.e approx 262.5.

2. The IQR 1= approx. 255 to 280 and for IQR 2 = 225 to 300.

3.Both the box plots don’t show outliers.

4.The spread of values for plot 1st = 237.5 to 277.5.

And for 2nd = below 200 to 337.5.

5.both the plots are normally skewed mean the values are distributed symmetrically.

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

Ans:-

Probability for MPG>38 is: 0.3475939251582705

* 1. P(MPG<40)

Probability for MPG<40 is: 0.7293498762151616

* 1. (20<MPG<50)

ANS : Probability for 20<MPG<50 is: 0.013116469610523339

Q 21) Check whether the data follows normal distribution

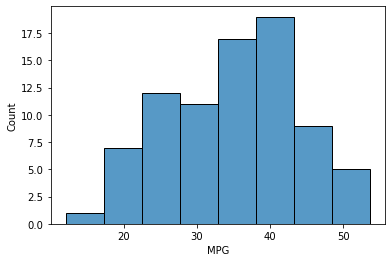
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

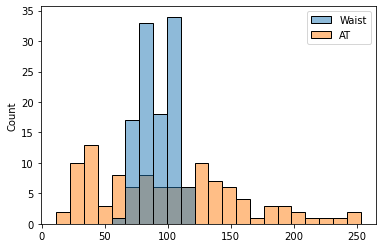
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv.

Ans:-



* The data set is not normally distributed, it is rightly skewed or positively skewed.

Ans B:- 

* Both the Adipose Tissue (AT) and Waist Circumference(Waist) data set do not follow the normal distribution approximately**.**

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

Ans –

|  |  |
| --- | --- |
| Confidence Interval | Z- Scores |
| 60% | 0.25 |
| 90% | 1.64 |
| 94% | 1.55 |

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

Ans –

|  |  |
| --- | --- |
| Confidence interval | T scores |
| 95% | 2.06 |
| 96% | 2.17 |
| 99% | 2.79 |

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- t=(260-270)/(90/18\*\*0.5)

t=-0.4714045207910317

p\_value=1-stats.t.cdf(abs(-0.4714),df=17)

p\_value=0.32167411684460556