|  |  |
| --- | --- |
| 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 | ordinal |
| 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) | ordinal |
| Sales Figures | interval |
| Blood Group | nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | interval |
| Number of Children | ordinal |
| Religious Preference | ordinal |
| Barometer Pressure | interval |
| SAT Scores | ordinal |
| Years of Education | nominal |

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

Answer :

Number of outcomes After Tossing the Coin **N=HHH,HTH,HHT,THH,TTT,THT,TTH,HTT = 8**

Formula for finding probability is as fallows =**P(A)=f/n:**

Probablity of getting two head and one tail is as fallows**:HTH,HHT,THH=3**

**P(Two Head and one Tail)=3/8**

\*Probablity of getting two head and one tail is =**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 2and 3

Answer:

If two disc are rolled lets find the number of occurances

(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) **total number of occurance==3**

**P(a)=f/n**

**1)Sum is equal to 1**

There is no match**(f)** =0;

Total number of occurance**(n)**=36;

**P=0/36**

**Answer is probablity eaqual to zero P=0**

P=0 :because there is no match found

**2) Sum of Less than or equal to 4**

Total number of match(F)=**(3,1)(2,2)(3,1)=3**

Total number of occurance**(N)**=36

**P=3/36==0.0833**

Probability of getting 4 is **0.0833%**

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

Total number of match(F)=(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)**F=24**

Total number of occurance**(N)**=36

**P=24/36==0.67**

Probability of sum divisible by 2 and 3 is **0.67%**

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?

Answer :

## **Total number of balls = (2 + 3 + 2) = 7** Let S be the sample space. Then, n(S) = Number of ways of drawing 2 balls out of 7 **=7C2** **=(7\*6)/(2\*1)=21** **=21** Let E = Event of drawing 2 balls, none of which is blue. n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls. **=5C2** **=(5×4)/(2x1)=10** **=10**

## **P(E)=n(E)/n(s) =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

Answer :

Expected number of candies for a randomly selected child

=**0.015\*1+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09**

Expected number of candies for randomly selected child=3.09

=  **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.

**Answer :**

**#Mean values for Points , score ,weigh** :

**Mean values**

**Points 3.596563**   
**Score 3.217250**  
**Weigh 17.848750**

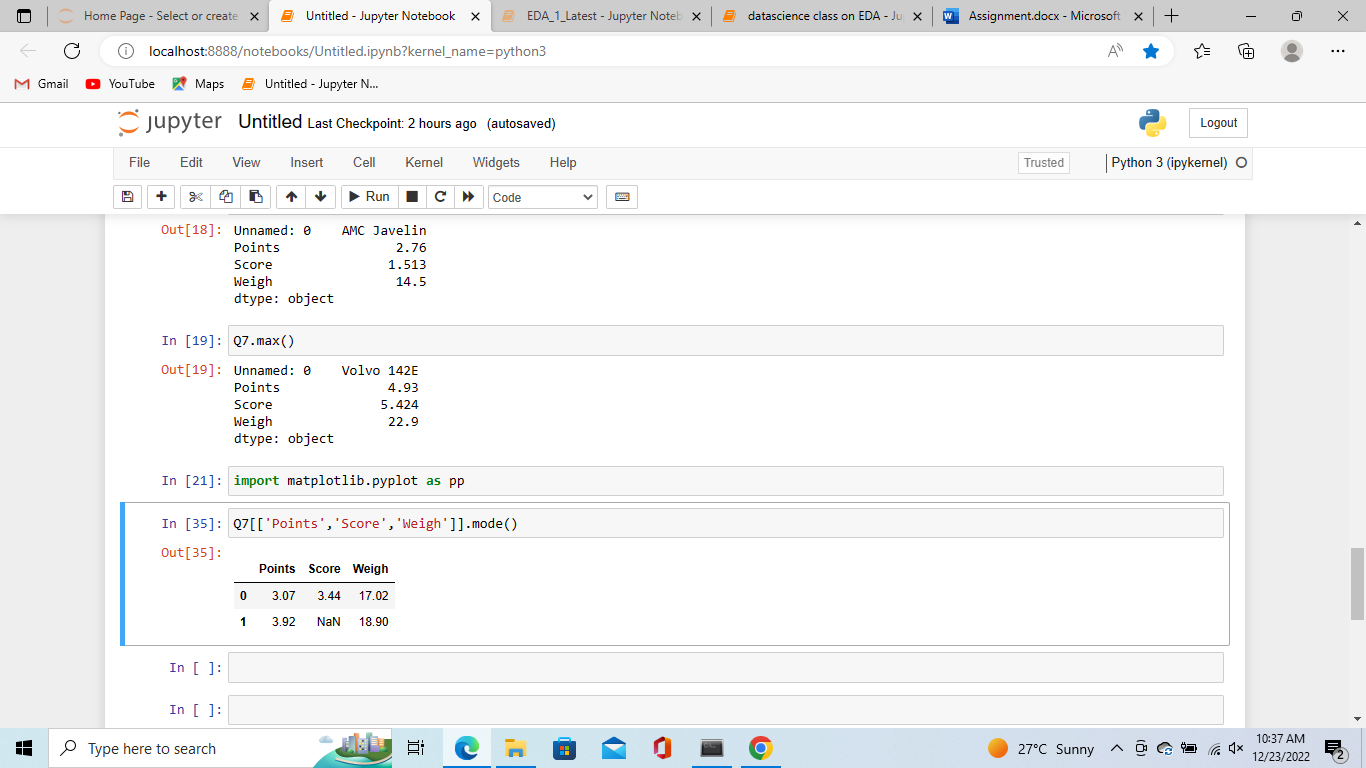
**#Median Values for Points ,Score ,Weigh :**

**Median values**

**Points 3.695**  
**Score 3.325**  
**Weigh 17.710**

**#Mode Values for Points,Score,Weigh** :

Mode value are as fallows



# **Variance for Points , score ,weigh** :

Points 0.285881  
Score 0.957379  
Weigh 3.193166

# **Standard deviation for Points , score ,weigh** :

Points 0.534679  
Score 0.978457  
Weigh 1.786943

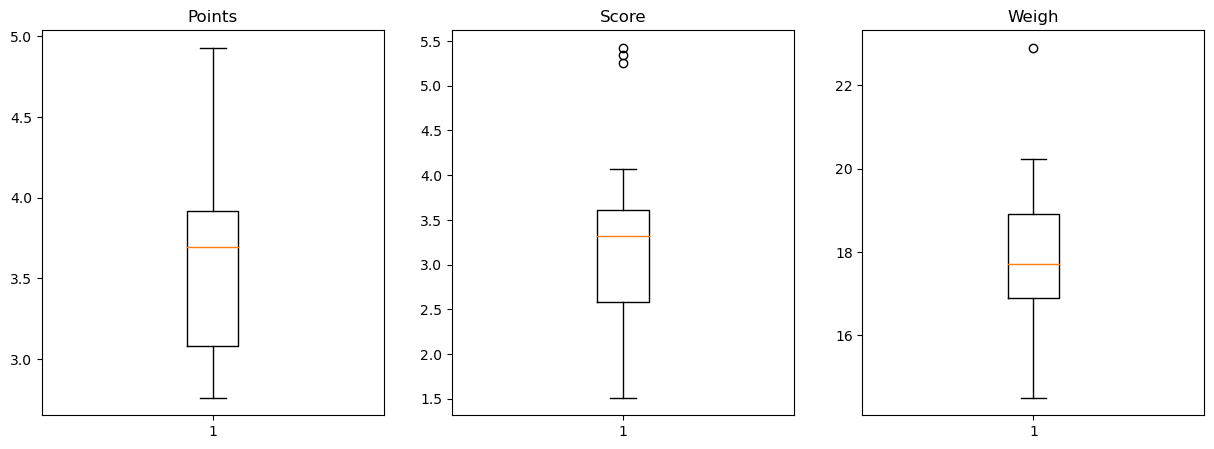
# **Range for Points , score ,weigh** :

Minmum values:

Points 2.76  
Score 1.513  
Weigh 14.5

Maximum values:

Points 4.93  
Score 5.424  
Weigh 22.9



Comments:

For Points dataset:

1) The data is concentrated around Median

2) There are no outliars

3) The distribution is Right skewed

For Score dataset:

1) The data is concentrated around Median

2) There are 3 Outliars: 5.250, 5.424, 5.345

3) The distribution is Left skewed

For Weigh dataset:

1) The data is concentrated around Median

2) There is 1 Outliar: 22.90

3) The distribution is Left skewed

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?

Answer :

Number of patients = 9 #there are nine members of patients

Probability=1/9

Sum of their weight of patients =108+110+123+134+135+145+167+187+199

=1308

Expected weight=sum of probability\*sum of values

=1308/9

Expected weight of patients=**145 pounds**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer :**

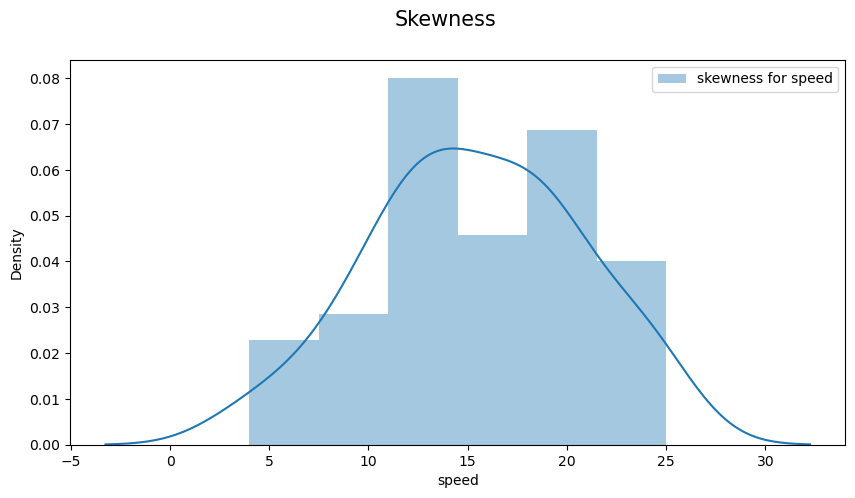
**#Skewness**

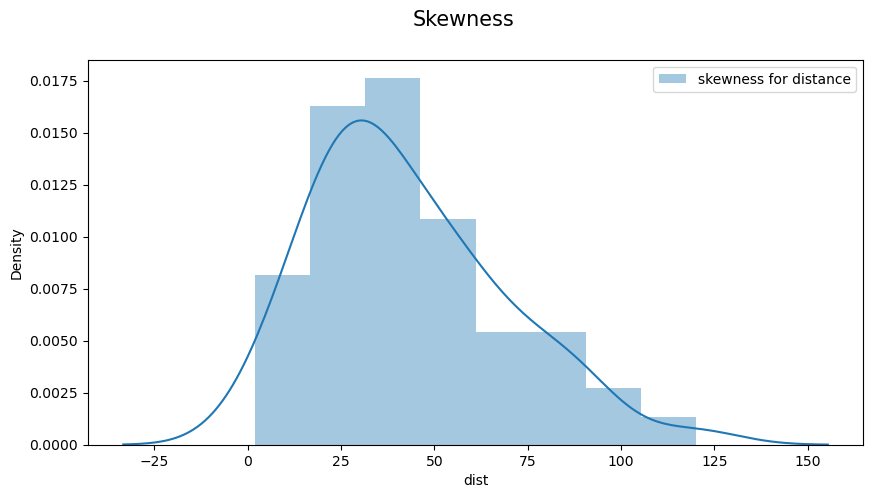
**speed -0.117510**

**dist 0.806895**

**#kurtosis**

s**peed -0.508994**  
**dist 0.405053**





**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer :**

**#Skewness**

**SP 1.611450**  
**WT -0.614753**

**#kurtosis**

**SP 2.977329**  
**WT 0.950291**

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



**Answer :**

The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.



Answer :

The boxplot has outliers in upper quartile range and its Rightly skweed

? **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:

Mean at 94% confidence interval is: [143.5762 256.4238]

Mean at 98% confidence interval is: [130.2096 269.7904]

Mean at 96% confidence interval is: [138.3875 261.6125]

**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= 41.0

Median= 40.5

Variance= 25.529412

Standard deviation= 5.052664

Average marks scored by the student is 41

Minmum marks scored by the student is

|  |  |
| --- | --- |
| **25% of students are getting** | 38.25 marks |
| **50% of students are getting** | 40.5 marks |
| **75% of students are getting** | 41.75 marks |

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

Answer:

If mean and median are equal then Skewness is equal to 0

Because plot will be in symmetrically distributed

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

Answer :

* If mean is greater then median then Skewness will be positive skewness

Because mean lies in positive side

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

Answer :

\*If mean is lesser then median then Skewness will be negative skewness

Because mean lies in negative side

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

Answers :

Positive kurtosis values indicates the peakness and it possesses thick tails compared to normal distribution

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

Answers :

Negative kurtosis indicates less peakness and it also possesses thin tails compared to normal distribution

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



What can we say about the distribution of the data?

Answer :

Distribution of the above plot is Asymmetrical and data are negatively skewed or left skewed , the reason for saying it as asymmetric is because in the inner quartile range (IQR) the median is not at the center of upper and lower quartile

What is nature of skewness of the data?

Answer :

it is negatively skewed or left skewed because the median line is near to upper quartile and the whiser line towards right side is less so it is negatively skewed

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

Answer :

Formula for the IQR is as fallows=upper quartile -lower quartile

Here we found upper quartile value approx : 18

Here we found Lower quartile value approx : 10

IQR=Q3-Q1

=18-10

IQR=8 :the inner quartile range is 8

Q19) Comment on the below Boxplot visualizations?



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

Answer**:** 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 ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Answer :

1. P(MPG>38)=Probability of MPG greater then 38 = 34.75%
2. P(MPG<40)=Probability of MPG Lesser then 40 = 72.94%
3. P(20<MPG<50)=Probability of MPG b/w 20 to 50= 89.88%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer :

MPG of cars are not Normal Distribution because mean and median are not equal for MPG, where mean is lesser then median so it is negatively skewed

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

Dataset: wc-at.csv

Answer :

Waist and AT is not Normal Distribution because mean and median are not equal for both, where mean is greater then median so it is positively skewed

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

Answer :

zscore for 90% CI: 1.64

zscore for 94% CI: 1.88

zscore for 60% CI: 0.84

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

Answer :

Tscore for 95% CI: 2.06

Tscore for 96% CI: 2.17

Tscore for 99% CI: 2.57

Q 24**)**A Government companyclaims 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:

Answer :

Population mean : 270

Sample bulbs : 18

Sample mean : 260

Standard deviation : 90

t=((260-270)/(90/(np.sqrt(18))))

t score = -0.4714

stats.t.cdf(t,17)DOF=sample-1=17,t=t score

probability of bulb less then randomly selected bulbs would have an average life of no more than 260 days

Probability (p) = 32.16%