|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | *Discrete data* |
| 6Results of rolling a dice | *Discrete data* |
| Weight of a person | *Continuous data* |
| Weight of Gold | *Continuous data* |
| Distance between two places | *Continuous data* |
| Length of a leaf | *Continuous data* |
| Dog's weight | *Continuous data* |
| Blue Color | *Discrete data* |
| Number of kids | *Discrete data* |
| Number of tickets in Indian railways | *Discrete data* |
| Number of times married | *Discrete data* |
| Gender (Male or Female) | *Discrete data* |

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 | *Interval* |
| Fahrenheit Temperature | *Interval* |
| Height | *Ratio* |
| Type of living accommodation | *Ordinal* |
| Level of Agreement | *Ordinal* |
| IQ(Intelligence Scale) | *Ratio* |
| Sales Figures | *Interval* |
| Blood Group | *Nominal* |
| Time Of Day | *Ratio* |
| Time on a Clock with Hands | *Ratio* |
| 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:***

*Possibility are 3 (HHT,THH,HTH)*

*Total no.of outcomes 8*

*Probability = 3/8 or 0.375*

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

***Ans:***

* *‘0’ because the least sum favorable outcomes are (1,1) in a dice*
* *favorable outcome = 6, total outcomes = 6\*6 = 36  
  total probability = 6/36 = 1/6 = 0.166*
* *favorable outcome = 1, total outcome = 6\*6 = 36  
  total probability = 6/36 = 1/6 = 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 ways (without blue balls) 5c2 = 5!/(2!\*(5-2)!) = 10  
 total no.of ways (with blue balls) 7c2 = 7!/(2!\*(7-2)!) = 21*

*Probability = 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 no.of Candies = 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0,01+2\*0.120*

*Expected no.of Candies = 3.075*

*So, the Expected number will be 3*

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:- df.mean().round(2)*

* *Points 3.60*
* *Score 3.22*
* *Weigh 17.85*

*Median: -*

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

*Mode:-*

* *Point's : -0 3.07 , 1 3.92*
* *Score : - 0 3.44*
* *Weight:- 0 17.02, 1 18.90*

*Variance:-*

* *Points 0.285881*
* *Score 0.957379*
* *Weigh 3.193166*

*Standard Deviation: -*

* *Points 0.534679*
* *Score 0.978457*
* *Weigh 1.786943*

*Range:*

* *Points 2.17*
* *Score 3.911*
* *Weigh 8.399*

***Source Code is in Attachment File name Assignment.ipynb***

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

* 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/no.of patirnts = 1308/9*

*Expected value = 145.33*

*Source code:*

*Weight = [108, 110, 123, 134, 135, 145, 167, 187, 199]*

*Expected\_Value = sum(Weight)/len(Weight)*

*Expected\_Value*

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

**Cars speed and distance**

**Use Q9\_a.csv**

*Ans:*

* *Skewness*
* *Index 0.000000*
* *speed -0.117510*
* *dist 0.806895*
* *Kurtosis*
* *Index -1.200000*
* *speed -0.508994*
* *dist 0.405053*

**SP and Weight(WT)**

**Use Q9\_b.csv**

*Ans:*

* *Skewness*
* *SP 1.611450*
* *WT -0.614753*
* *Kurtosis*
* *SP 2.977329*
* *WT 0.950291*

Attached python file with name (Assignment.ipynb) in mail

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



*Ans:*

* *The peak is in between 50-100.*
* *The least weight are in between 350-400.*
* *It represents that the data is right skewed*
* *We have outliers on the right side*



*Ans:*

* *The box plot is left skewed (Positively)with having outliers.*
* *The median and mean are not equal and median lies towards the Q1.*
* *The boxplot has the outliers on the 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:*

*Confidence Interval=Sample Mean±(Critical Value× Sample Standard Deviation/sqrt(Sample Size ))*

*​94%:*

*Z-score = 1.88 (approx)*

*CF = 200 ± 1.88(30/sqrt(2000))*

*CF ~ 199.19,200.81*

*98%:*

*Z-score = 2.33 (approx)*

*CF = 200 ± 2.33(30/sqrt(2000))*

*CF ~ 198.98,201.02*

*96%:*

*Z-score = 2.05 (approx)*

*CF = 200 ± 2.05 (30/sqrt(2000))*

*CF ~ 199.03,200.97*

Attached python file with name (Assignment.ipynb) in mail

**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**

* Find mean, median, variance, standard deviation.

*Ans:*

* *mean:- 41*
* *Median:- 40.5*
* *Variance:- 41*
* *STD:- 5.052*

Attached python file with name (Assignment.ipynb) in mail

* What can we say about the student marks?

*Ans:*

*The student's scores are centered around 41-42, with the most common score being 41. There is some variability in the scores, and the distribution is slightly positively skewed. The student's performance is generally good, with a range of scores from 34 to 56.*

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

*Ans: If mean == median then the skewness is* ***Normal***

*and the data is* ***Normal Distribution***

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

*Ans: If mean > median then the skewness is* ***Positive Skewness***

*and the data is Distributed towards* ***Right***

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

*Ans: If mean > median then the skewness is* ***Negative Skewness***

*and the data is Distributed towards* ***Left***

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

*Ans:*

* *A positive kurtosis value indicates that the distribution has heavier tails and more data points in the tails compared to a normal distribution*
* *Positive kurtosis is also referred to as leptokurtic. It means that the data has more outliers or extreme values than a normal distribution.*

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

*Ans:*

* *A negative kurtosis value indicates that the distribution has lighter tails and fewer data points in the tails compared to a normal distribution*
* *Negative kurtosis is also referred to as platykurtic. It means that the data has fewer outliers or extreme values than a normal distribution.*

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



What can we say about the distribution of the data?

*Ans:*

*The boxplot lies between 10-16 and more than 50% of the data lies in above 10 and data above the median has approx 40%*

What is nature of skewness of the data?

*Ans: The Data is right skewed, median is greater than mean.*

What will be the IQR of the data (approximately)?   
  
*Ans: When lower whisker is 10 and upper whisker is 18 and median in 15 (assumption)*

*q1 = 10*

*q3 = 18*

*IQR = q3-q1 = 18-10 = 8  
  
Approximately IQR = 8*

Q19) Comment on the below Boxplot visualizations?



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

*Ans:*

* *Both the box plot median lies on the same place close to 263*
* *The data is Symmetrically distributed around this central value as the mean and median is equal*
* *Both 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)
2. P(MPG<40)
3. P (20<MPG<50)

*Ans:*

*a. (df.MPG > 38).sum()*

*The no.of favourable outcome are 33*

*The no.of Total outcome are 81*

*So. P(MPG>38) is 0.407*

*b. (df.MPG < 40).sum()*

*The no.of favourable outcome are 61*

*The no.of Total outcome are 81*

*So. P(MPG<40) is 0.753*

*c. ((20<df.MPG) & (df.MPG<50)).sum()*

*The no.of favourable outcome are 69*

*The no.of Total outcome are 81*

*So. P(MPG>38) is 0.851*

Attached python file with name (Assignment.ipynb) in mail

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

*Ans:*

*The data is Normally distributed. As Mean is slightly near to median.*

*Mean = 34.422*

*Median = 35.152*

Attached python file with name (Assignment.ipynb) in mail

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

Dataset: wc-at.csv

*Ans:*

* *In Adipose Tissue the data is not Normally Distributed. As Mean > Median .*

*Mean = 101.89*

*Median = 96.54*

* *In Waist Circumference the data is Normally Distributed. As Mean is slightly near to median.*

*Mean = 91.90*

*Median = 90.8*

Attached python file with name (Assignment.ipynb) in mail

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

*Ans:*

* *90%*

*Z-score = 1.645*

* *94%*

*Z-score = 1.881*

* *60%*

*Z-score = 0.842*

Attached python file with name (Assignment.ipynb) in mail

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

*Ans :*

*95% CL:*

*α = 1-cl =0.05*

*df = 25-1=24*

*t-score : 2.064*

*96% CL:*

*α = 1-cl =0.04*

*t-score : 2.172*

*99% CL:*

*α = 1-cl =0.01*

*t-score : 2.797*

Attached python file with name (Assignment.ipynb) in mail

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:*

*sample\_mean = 260*

*sample\_size = 18*

*population\_mean = 270*

*sample\_std\_dev = 90*

*Probability is 0.321 or 32.1%*

*Probability that 18 bulbs would have an average life of no more than 260 days is 32.1% only*

Attached python file with name (Assignment.ipynb) in mail