|  |  |
| --- | --- |
| 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 | Discrete |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Continuous |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans : S={HHH,HHT,HTT,TTT,TTH,THH,THT,HTH}=8

Probability of two head and one tail(HHT,HTH,THH)=3

Probabilty = 3/8=0.37

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

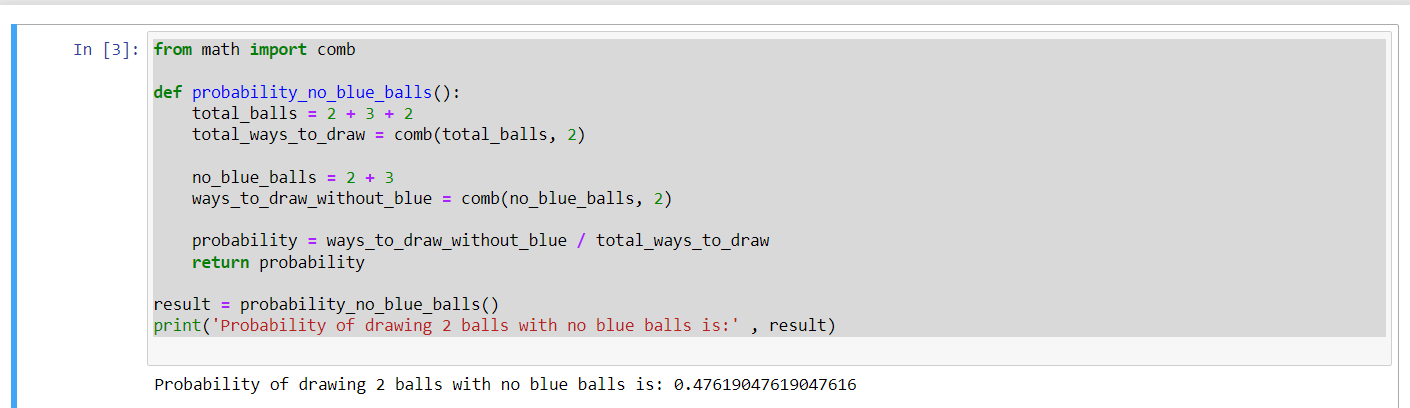
1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans : Sample set will be size of 36

1. probability that sum is Equal to 1= 0
2. probability that sum is Less than or equal to 4 = 6/36=1/6
3. probability that sum is divisible by 2 and 3 = 2/36=1/18

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 : code in python



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 number of candies = sum of product of individual candy count and their probability

i.e. sumation (xipi)=(1 x 0.015)+(4 x 0.20)+(3 x 0.65)+(5 x 0.005)+

(6 x 0.01)+(2 x 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:** In Jupyter notebook .

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 or mean =

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

=135.33

So, the patient whose weight is chosen at random is 135 pounds.

**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 :**

In Jupyter Notebook .

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



Ans : Histogram follow positve skewness and there is high possibility that

there exist outlier greater than weight of 350 approx.



Ans : There exist outlier above upper quartile means value must be greater

than Q3 and median is closer to lower quartile means it shows

positive skewness.

**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: In Jupyter Notebook .

**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, variance=25.52 , standard deviation=5.05

1. What can we say about the student marks?

Ans: Maximum student has nearly avg marks and marks positively skewed.

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

Ans: It means data symmetric and has bell shaped curve which means data is normal distribution .

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

Ans: Right tailed skewed means positively skewed.

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

Ans: Left tailed means data is negatively skewed .

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

Ans: It means data is more peaked at center and has more outlier.

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

Ans : It means data is flatter peaked at center than normal distribution means it has low outlier.

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



What can we say about the distribution of the data?

Ans: Data is not symmetric and it doesn’t follow normal distribution.

What is nature of skewness of the data?

Ans: Data is negatively skewed since median is closer to upper quartile.

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

Ans: IQR=Q3-Q1=18-10=2 .  
  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans: It seems that there is large variability in boxplot 1 than 2 and median of boxplot 1 is closer to lower quartile means it is positively skewed while boxplot 2

show normal distribution.

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 : In Jupyter Notebook .

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: With the help of Q-Q plot we can say that MPG column follow 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: By plotting Q-Q plot it is clear that both wc and at column follow normal distribution.

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

Ans : In Jupyter Notebook .

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

Ans : In Jupyter Notebook .

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 : In Jupyter Notebook .