|  |  |
| --- | --- |
| 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) | Continuous |

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**The Probability of two heads and one tail obtained is 3/8 i.e 0.375%**

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

1. Equal to 1 -- **0 (If we toss a coin we will get some number and it is not possible that sum of two dice rolled will be 0)**
2. Less than or equal to 4 -- **1/6 i.e 0.166%**
3. Sum is divisible by 2 and 3 -- **(1+5),(2+4),(3+3),(4+2),(5+1),(6+6) = 6/36 = 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?

**Total ways to pick 2 balls = 7!/2!(7-2)! = 7\*6\*5!/2\*5! = 7\*6/2 = 21.**

**Ways to pick 2 balls from 5(Excluding Blue) = 5!/2!(5-2)! = 5\*4\*3!/2\*3! = 5\*4/2 = 10.**

**The probability that none of the balls drawn is blue is 10/21 i.e 0.476%**

**Formula applied is n!/r!(n-r)!**

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

**Expected No of Candies = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 3.09**

**Expected number of Candies for a randomly selected child is 3.09 ~ 3.1 candies**

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**

**Refer Jupyter Note Book for Solution**

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?

**Expected Value = Sum of all weights/No of Patients**

**The Expected Value(mean) of the weight of a randomly chosen patient will be 145.33 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**

**Refer Jupyter Note Book for Solution**

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



**In the above image, On X-axis we have ChickWeight$weight column data and on Y-axis we have Frequency. As we can see that the histogram is created with repect to values of Weight & Frequency.**



**From the above boxplot we can see that the dataset have some/multiple outliers which need to be fixed. Also it has skewness at top.**

**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?

**Mean = 200**

**Std\_Dev = 30**

**n = 2000**

**df = n-1 = 2000-1 = 1999**

**For 94% Interval the weight will be approx (198.89, 201.10)pounds. Here, we found the tvalue as +-1.646**

**For 98% Interval the weight will be approx (198.597, 201.403). Here, we found the tvalue as +-2.330**

**For 96% Interval the weight will be approx (198.808, 201.192). Here, we found the tvalue as +-1.192**

**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?

**Refer Jupyter Note Book for Solution**

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

**It will not have any Skewness. The nature of Skewness will be Symmetrical.**

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

**If mean>median then the nature of Skewness will be Positive also known as Positively Skewed.**

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

**If mean>median then the nature of Skewness will be Positive also known as Positively Skewed.**

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

**It indicates that the dataset have sharp central peak than the normal distribution.**

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

**The dataset have a Flatter Distribution(like a Plateau). Also the dataset have few extreme values(Outliers).**

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



What can we say about the distribution of the data?

**According to the Image the boxplot is left-skewed i.e Negative Skewed. The tail on the left side of distribution is longer.**

What is nature of skewness of the data?

**It is Left-Skewed(Negative Skewed).**

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

**According to the Image lets consider Q3 = 18 & Q1 = 10. So,**

**IQR = Q3 – Q1 = 18-10 = 8.**

Q19) Comment on the below Boxplot visualizations?



**The Median of both the boxplot visualization looks the same. Also, we don’t have tiny circles i.e the Outliers. The Number 2 dataset is widely spread as compared to Number 1.**

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

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)

**Refer Jupyter Note Book for Solution**

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

**Refer Jupyter Note Book for Solution**

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

**The Z score for 90% interval is 1.645 & for 60% interval is 0.8416.**

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

**To Calculate tscore we need degree of freedom:**

**df = n-1; df = 25-1; df = 24.**

**t score for 95% interval is 2.064.**

**t score for 96% interval is 2.174.**

**t score for 99% interval is 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

**Refer Jupyter Note Book for Solution**