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

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

**Ans:-**

a=2H & 1T

n(a)= Number of favourable outcomes.

n(s) = Sample space

p= Probability

n(s) = {HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}

n(s) = 8

n(a) = {HHT,HTH,THH}

n(a) = 3

p = n(a)/n(s)

p = 3/8

**p = 0.375 or 37.5%**

**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**

**Ans:-**

Sample space:-

n(S) = {(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)}

n(s) = 36

1. **A=Sum Equal to 1**.

n(A) = 0

p = n(A)/n(S)

**p= 0/36 = 0**

1. **Sum less than or equal to 4.**

n(A) = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

n(A) = 6

p = n(A)/n(S)

p = 6/36

p = 1/6

**p = 0.1667 or 16.67%**

1. **Sum is divisible by 2 and 3**

n(A) = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

n(A) = 6

p = n(A)/n(S)

p = 6/36 = 1/3

**p = 0.1667 or 16.67 %**

**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?**

**Ans:-**

Total number of balls = 2+3+2= 7

n(S) = 7C2 = (7\*6)/(2\*1) = 21

n(A) = 5C2 = (5\*4)/(2\*1) = 10

p = n(A)/n(S)

p = 10/21

**p = 0.4762 or 47.62%**

**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 randomly selected child.

E(x) = ∑(x) \* P(x)

E(x) = (1+0.015) + (4\*0.20) + (3\*0.65) + (5\*0.005) + (6\*0.01) + (2\*0.120)

E(x) = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

**E(x) = 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:- Refer Jupiter Notebook file.**

**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:- Refer Jupiter Notebook file.**

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

**1)Cars speed and distance**

**Use Q9\_a.csv**

**2)SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:- Refer Jupiter Notebook file.**

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



**Ans:-**

1. From histogram we can say that data is positively skewed.
2. Weight between 50 to 100 is more.
3. Box plot shows there are some outliers present in the data set.

**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:- Refer Jupiter Notebook file.**

**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:- Refer Jupiter Notebook file.**

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

**Ans:-** When mean is equal to median, data is symmetrically distributed and in that case skewness is zero.

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

**Ans:-** When mean is greater than median, the nature of skewness is positive(data is positively skewed).

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

**Ans:-** When median is greater than mean, the nature of skewness is negative(data is negatively skewed).

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

**Ans:-** Positive kurtosis indicate that distribution is peaked and possesses thick tails.Which means there are more chances of outliers.

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

**Ans:-** Negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



1. **What can we say about the distribution of the data?**

**Ans:-**  Here we can see that left whisker is longer than the right whisker, so the distribution is negatively skewed.

**2) What is nature of skewness of the data?**

**Ans:-** The nature of skewness of the data is negative.

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

**Ans:- IQR =Upper Quartile – Lower Quartile**

IQR ≈ 8.2

**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 whiskers length in the Boxplot-1 & Boxplot-2 is almost same, so the distribution in both cases is symmetric.

1. The inter quartile range in Boxplot-1 is less than in Boxplot-2.
2. Mean in the both cases is same.
3. Range in Boxplot-2 is more than Boxplot-1.
4. More observation taken in Boxplot-2 than Boxplot-1.

**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)

c. P (20<MPG<50)

**Ans:- Refer Jupiter Notebook file.**

**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:- Refer Jupiter Notebook file.**

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

**Ans:-**

**Z-Scores of 90%,94% & 60% confidence interval:-**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confidence Interval** | **90%** | **94%** | **60%** |
| **Z-Scores** | **1.64** | **1.88** | **0.84** |

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

**Ans:- Given, Sample Space (n) = 25.**

**Degree of freedom = n-1**

**= 25-1**

**Degree of Freedom(DOF) = 24**

**t Score of 95%,96% & 99% confidence interval at DOF = 24:-**

|  |  |  |  |
| --- | --- | --- | --- |
| **Confidence Interval** | **95%** | **96%** | **99%** |
| **t Score** | **2.064** | **2.208** | **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

**Ans:- Refer Jupiter Notebook file.**