**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| 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 |

**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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?**

**Ans** P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three

coins tossed)) = 3/8 = 0.375 = 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 2 and 3

**Ans:**

1. Equal to 1 = The set of possible outcomes when we roll a die are {1, 2, 3, 4, 5, 6} So, when we roll two dice there are 6 × 6 = 36 outcomes. There is no any moment when the sum of dice is 1 Therefore, the P (sum is equal to 1) = 0/36=0.

b) Less than or equal to 4 = (Sum is less than or equal to 4) = N (Event (Sum is less than or equal to 4)) / N (Event (Two dice rolled)) = 6 / 36 = 1/6 = 0.166 = 16.66%

c)Sum is divisible by 2 and 3 = N (Event (Sum is divisible by 2 and 3)) / N(Event (Two dice rolled)) = 6 / 36 = 1/6 = 0.16 = 16.66%

**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**: 0.47619047619047

**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:** = 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.120 = 0.015+0.8+1.95+0.025+0.06+0.240 =3.090

**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 of points data = 3.5965625

Median of points data = 3.695000

Mode of points data = 3.07

Variance of points data =0.2858813

Standard deviation of points data = 0.5346787

Range of points data= 2.17

Mean of score data = 3.217250

Median of score data = 3.325

Mode of score data = 3.44

Variance of score data = 0.9573789

Standard deviation of score data = 0.978245

Range of score data =3.911000

Mean of weigh data=17.84875

Median of weigh data =17.71

Mode of weigh data =17.02

Variance of weigh data =3.19316

Standard deviation of weigh data= 1.78694

Range of weigh data= 8.3999

**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 = Sum (X \* Probability of X) = (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199) = 145.33

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

**Car’s speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:**

1. **Car’s speed and distance**

Skewness

Speed= -0.11175

Distance= 0.8068

Kurtosis

Speed= -0.5089

Distance = 0.40505

1. **SP and Weight (WT)**

Skewness

SP= 1.6114

WT= -0.6147

Kurtosis

SP= 2.9773

WT= 0.9502

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



**ANS:** The histogram’s peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

The boxplot has outliers on the maximum 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:** For 94% confidence interval Range is [ 198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

**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:** 1) mean = 41.0, median =40.5, variance= 25.529411764705884, standard deviation =5.05266382858645

2)By looking at the boxplot I can say that most of the student’s marks lies between 35 to 45.

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

**ANS:** Skewness refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed. Skewness can be quantified as a representation of the extent to which a given distribution varies from a normal distribution. A normal distribution has a skew of zero. When mean, median and mode is equal then it is a normal distribution. So, the nature of skewness is zero when mean and median of data is equal.

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

**ANS:** When the value of mean is greater than median and mode then it is called positive skewed. So, the nature of skewness is positive when mean > median.

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

**ANS:** When the value of mean is less than median and mode then it is called negative skewed. So, the nature of skewness is negative when median > mean.

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

**ANS:** Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. An extreme positive kurtosis indicates a distribution where more of the numbers are located in the tails of the distribution instead of around the mean. It means that the dataset has more outliers or extreme values compared to a normal distribution.

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

**ANS**: A negative kurtosis value indicates that a dataset has lighter tails and a flatter peak compared to the normal distribution, which is known as platykurtic distribution. In simpler terms, it means that the dataset has fewer outliers or extreme values compared to a normal distribution.

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



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

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

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

**ANS:**

1) The above Boxplot is not normally distributed the median is towards the higher value

2**)**The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

3) The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 18 – 10 =8

**Q19) Comment on the below Boxplot visualizations?**



**ANS:**  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 of Cars for the below cases.**

**MPG <- Cars$MPG**

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

**c. P (20<MPG<50)**

**ANS:** a. P(MPG>38) = 0.34759392515827137

b. P(MPG<40) = 0.7293498762151609

c. P (20<MPG<50) = 0.898868916968204

**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:**

**a)** Yes, MPG of cars follows normal distribution.

**b)** Adipose Tissue (AT) and Waist does not follow Normal Distribution.

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

**ANS:**

Z scores of 90% confidence interval = 1.6448536269514722

Z scores of 94% confidence interval = 1.8807936081512509

Z scores of 60% confidence interval = 0.8416212335729143

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

**ANS:**

t scores of 95% confidence interval for sample size of 25 = 2.0638985616280205

t scores of 96% confidence interval for sample size of 25 = 2.1715446760080677

t scores of 99% confidence interval for sample size of 25 = 2.796939504772804

**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:**  Null Hypothesis: Avg life of Bulb >= 260 days

Alternative Hypothesis: Avg life of Bulb < 260 days

Population mean(pmean) =260

Sample mean(smean)=270

Sample size(n)=18 bulbs

Sample standard deviation(sd)=90 days

t=(pmean-smean)/(sd/math.sqrt(n)

t=-0.4714052

pvalue= 1-stats.t.cdf(abs(t), df=18-1)

pvalue=0.32167411

Therefore, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days, if the CEO's claim were true, is approximately 0.321 or 32%.