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

|  |  |
| --- | --- |
| **Activity** | **Data Type** |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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 | Discreate |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Discreate |
| Gender (Male or Female) | Discreate |

**Q2) Identify the Data types, which were among the following**

**Nominal, Ordinal, Interval, Ratio.**

|  |  |
| --- | --- |
| **Data** | **Data Type** |
| Gender | Nominal |
| High School Class Ranking | Nominal |
| Celsius Temperature | Ratio |
| Weight | Ratio |
| Hair Color | Ratio |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ratio |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Ratio |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

Ans: If three coins are tossed,

Total number of possible combinations = 23 = 8

The combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

Number of combinations that have two heads and one tail = 3, i.e., HHT, HTH, TTH

The probability of two heads and one tail when three coins are tossed simultaneously are

P (Two heads and One tail) = Number of desired outcomes

= ⅜ or 0.375

**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: If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0

B sum is less than or equal to 4

X={(1,3),(2,2),(3,1)}

P(x)=3/36=1/12

Sum is divided by 2 and 3

X1={(1,3),(2,2),(3,1),(1,5),(2,4),(4,2)(3,3),(5,1)(6,6)}

X2={(1,5),(2,4),(3,3)(4,2),(5,1),(6,6)}

Y=X1intersectionx2={(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

=6

P(Y)=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?**

Ans: Total no. of balls = 2R + 3G + 2B = 7

1. No. of ways of drawing 2 balls out of 7:

7C2 = (7 x 6)/ (2 x 1) = 21

No. of balls other than blue = 5

1. No. of ways of drawing 2 balls out of 5:

5C2 = (5 x 2)/ (2 x 1) = 10

Probability = 5C2 / 7C2 = 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 number of candies for a randomly selected child=

(1x0.015+4x0.20+3x0.65+5x0.005+6x0.001+2x0.120)=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:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard Deviation |
| Points: | 3.5965 | 3.695 | Numeric | 0.28588 | 0.53467 |
|  |  |  |  |  |  |
| Score: | 3.2177 | 3.325 | Numeric | 0.95737 | 0.97845 |
|  |  |  |  |  |  |
| Weight | 17.8487 | 17.710 | Numeric | 3.193166 | 1.78694 |

**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: There are total 9 patients therefore the probability of selecting each patient =1/9

Expected value= sum(probability \*value)

=1/9(108+110+123+134+145+167+187+199)

=(1/9)x1308=145.33

Expected value of the weight of that patient =145.33

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

**a)Cars speed and distance [Use Q9\_a.csv]**

**b)SP and Weight(WT) [Use Q9\_b.csv]**

**Ans: a) Skewness**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| **Q9\_a** | **-0.11395477** | **0.78248352** |
|  |  |  |
|  | **SP** | **WT** |
| **Q9\_b** | **1.5814** | **-0.60330993** |

1. **Kurtosis**

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| **Q9\_a** | **-0.57714742** | **0.24801866** |
|  |  |  |
|  | **SP** | **WT** |
| **Q9\_b** | **2.72352149** | **0.819465881** |

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





**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:

The information given is: Since the data is too small so using **t-distribution**

* Sample **mean** of x = 200
* Sample **standard deviation** of s = 30
* Sample **size** of n = 2000

The confidence interval = x(+/-) t\*(s/(n)\*\*1/2)

* **x:** sample mean
* **t:** t-value that corresponds to the confidence level
* **s:** sample standard deviation
* **n:** sample size

|  |  |  |
| --- | --- | --- |
| CI | x-t\*(s/n)\*\*1/2) | x+t\*(s/n)\*\*1/2) |
| 94% | 198.738325292158 | 201.261674707842 |
| 98% | 198.43943840429978 | 201.5605615957022 |
| 96% | 198.62230334813333 | 201.37769665186667 |

**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

Variance=27.125

Standard deviation=5.2081

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

Ans: Symmetrical

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

Ans: Positive skewed

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

Ans: Negatively skewed

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

Ans: It has more in the tails in the normal distribution.

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

Ans: It has less in the tails than the normal distribution.

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



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

Ans: Let’s assume above box plot is about age’s of children in a society. 50% of the children are above 10years old and remaining are less, and children who’s age is above 15 are approx 40%.

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

Ans: Left skewed because(Median>Mean)

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

Ans: Approximately=-8

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

Ans: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.



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)

c. P (20<MPG<50)

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**
2. **Dataset: Cars.csv**

Ans:

Mean= 34.422076

Standard deviation=9.131445

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

**Dataset: wc-at.csv**

Ans:

|  |  |  |
| --- | --- | --- |
|  | Waist | AT |
| Mean | 91.901835 | 101.894037 |
| Median | 90.80 | 96.54 |
| Mode | 94.5,106.0,108.5 | 121.0,123.0,NaN |

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

Ans:

* Z-score of 90% CI = 1.644853626914722
* Z-score of 94% CI = 1.8807936081512509
* Z-score of 60% CI = 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-score of 95%CI = 2.063895616280205
* t-score of 96%CI = 2.1715446760080677
* t-score of 99%CI = 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**

**rcode 🡪 pt(tscore,df)**

**df 🡪 degrees of freedom**

Ans:

Ts=0.4714045

P\_value=1.356651