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

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

**Solution:**

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

No. of combinations that have two heads and one tail = 3, i.e. HTH,HHT,THH

Hence,

P(Two heads and One tail ) = 3/8 i.e.0.375

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

Solution:

The set of possible outcomes when we roll a die = {1,2,3,4,5,6}

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)}

So, when we roll two dice there are 6\*6 i.e. 36 possibilities

1. **Equal to 1**

**Solution:**

Probability that the sum =1 is Zero,

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

Hence, Probability = 0/36, i.e. 0

1. **Less than or equal to 4**

Solution:

Possible outcomes = (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

Hence, Probability = 6/36, i.e. 1/6

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

Solution:

Sum Should be divisible by 2 & 3

Possible outcomes = (1,5), (2,4),(3,3), (4,2), (5,1), (6,6)

Hence, The probability that the sum is divisible by 2 & 3 is : 6/36 i.e. 0.166

**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?**

Solution:

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

n(S) = Number of ways of drawing 2 balls out of 7

=(7\*6)/(2\*1)

=21

Let E = Event of drawing 2 balls, none of which is blue.

n(E) = (2+3)

=(5\*4)/(2\*1)

=10

Hence, p(E) = n(E)/n(S)

=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**

Solution:

Expected number of candies for a randomly selected child :

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

= 3.09

Hence, Expected number of candies for a randomly selected child = 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**

Solution: <https://github.com/NamrataIngle/DS-Assignments.git>

**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?**

Solution:

There are 9 patients,

Hence, Probability of selecting each patient = 1/9

Expected value = (1/9)(108+110+123+134+135+145+167+187+199)

= (1/9)(1308)

= **145.33**

**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**

**Solution :** [**https://github.com/NamrataIngle/DS-Assignments.git**](https://github.com/NamrataIngle/DS-Assignments.git)

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



Observations:

* Histogram of ChickWeight $ Weight has right Skewed data.
* And has tail on right side.
* Mean>Median.



Observation:

* 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?**

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>

**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?**

Solution: <https://github.com/NamrataIngle/DS-Assignments.git>

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

Ans : When mean & median of the data are equal, the sqewness of the

distribution is zero.

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

Ans: When mean>median, the distribution is positively skewed. The tail is towards

right.

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

Ans: When median>mean, the distribution is negatively skewed. The tail is

towards left.

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

Ans: Positive kurtosis value indicates that the curve is more peaked and it is

Leptokurtic.

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

Ans: Negative kurtosis value indicates that the curve is flatter and broader.

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



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

- Data distribution in the above boxplot is not normally distributed.

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

- The data is left skewed.

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

- IQR(Inter Quartile Range) = Upper Quartile - Lower Quartile

= 18-10

= 8

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



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

Ans:

* There are no outliers
* Both the box plot have same median in a range between 250 to 275.
* Data is normally distributed, Zero skewness is there.

**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)**

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>

**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**

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>

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

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>

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

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>

**Q 24)A Government companyclaims 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**

Solution : <https://github.com/NamrataIngle/DS-Assignments.git>