|  |  |
| --- | --- |
| **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 | Categorical / Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical / Nominal |

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

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

Ans. The total possible outcome =8

HHH,HHT,HTT,THT,TTH,HTH,THH,TTT

Where H-Head; T-Tail

Number of favorable outcomes =3

Probability(two head and one tail)=**3/8** = **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: a) Total number of outcomes = 36

Number of outcomes sum equal to 1=0

Prob (equal to 1) = **0/36**

b) Total number of outcomes =36

Number of outcomes less than or equal to 4 = 6

(1,1), (1,2), (1,3), (2,1), (2,2), (3,1) = 6

Prob (less than or equal to 4) = 6/36 = **1/6**

c) Total number of outcomes =36

Sum = 6: (1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

Number of outcomes sum is divisible by 2 and 3 =5

Prob (sum is divisible by 2 and 3)=5/36 = **0.139**

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 =7 balls N

.Event (2 balls are drawn randomly from bag) = 7! / 2! \* 5! = (7654321) / (21) \* (54321) N

.Event (2 balls are drawn randomly from bag) = (76)/ (21) = 21

If none of them drawn 2 balls are blue = 7 – 2 = 5 N

.Event (None of the balls drawn is blue) = 5! / 2! \* 3! = (54) / (2\*1) = 10 P

.None of the balls drawn is blue = N (Event (None of the balls drawn is blue) / N (Event (2 balls are drawn randomly from bag) = 10 / 21 = 0.476

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 Value = Σ (Candies count \* Probability)

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

Expected Value = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

Expected Value = **3.09**

So, Expected number of candies for randomly selected child is **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** | **Std. dev** | **Range** |
| **Points** | 3.59 | 3.69 | 3.92 | 0.285 | 0.534 | 2.17 |
| **Score** | 3.21 | 3.32 | 3.44 | 0.957 | 0.978 | 3.911 |
| **Weigh** | 17.84 | 17.71 | 17.02 | 3.193 | 1.786 | 8.399 |

**Comments:**

1. No case has got variable -> Mean=Median=Mode
2. As seen in boxplot ‘Score’ and ‘Weigh’ has got outliers.

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: X =1/9

 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**

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **Speed** | -0.117 | -0.508 |
| **Distance** | 0.806 | 0.405 |

* “Distance” is positively skewed where as “speed” is negatively skewed
  + Thus Distance has distribution of data concentrated on the left whereas speed has distribution on the right. As seen in the above values
* Distance has positive Kurtosis and speed has negative Kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.611 | 2.977 |
| **Weight** | -0.614 | 0.950 |

* “SP” is positively skewed where as “Weight” is negatively skewed
  + Thus SP has distribution of data concentrated on the left whereas Weight has distribution on the right. As seen in the above values
* Both SP and weight has positive Kurtosis.

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



ANS :

1. Majority of the Chicks has weight in range 50 – 100, followed by 100 -150 and 150 – 200
2. The Histogram data is positively Skewed at right side.
3. 3 Categories of weight can be define (Under weight <50; Avg weight 51 – 150; Over weight >150).



Ans: Data has outliers and Data is Positively skewed.

**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:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| Upper | 201.261 | 201.560 | 201.377 |
| Lower | 198.738 | 198.439 | 198.622 |

**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.

ANS:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.50 |
| Variance | 25.53 |
| Std Deviation | 5.05 |

1. What can we say about the student marks?

Ans:

1. Mean is not equal to Median in given data.
2. Data has outlier one outlier value 56.
3. Majority of the students scored between 35 – 45 Marks.

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

ANS: Skewness = 0. Perfectly symmetric in bell shaped curve

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

ANS: Skewness = Positive. Data is distributed more on left

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

ANS: Skewness = Negative. Data is distributed more on right

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

ANS: High and narrow peak on central part of the data/graph.

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

ANS: wider peak on central part of the data/graph.

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



What can we say about the distribution of the data?

ANS : The data is not symmetric because data is more concentrated towards right side.

What is nature of skewness of the data?

ANS: Negative Skewness

What will be the IQR of the data (approximately)?   
ANS: Q1=10; Q3=18

IQR=Q3-Q1

= 18-10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

ANS:

1. Data is Normally Distributed. No Outliers. Center is around 262.5. Comparatively, first graph has less range.
2. Data is Normally Distributed. No Outliers. Center is around 262.5

Comparatively, second graph has more 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 #(Solution is attached in Jupyter file)

* 1. P(MPG>38)

Ans: 0.407

* 1. P(MPG<40)

Ans: 0.753

c. P (20<MPG<50)

Ans: 0.851

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes, MPG is Normal Distribution. (Mean =~ Median)

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

Dataset: wc-at.csv

Ans: Both AT and Waist doesn’t follow Normal Distribution. (Since Mean!=Median).

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

ANS:

|  |  |
| --- | --- |
| 90% | ± 1.644 |
| 94% | ± 1.880 |
| 60% | ± 0.841 |

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

ANS:

|  |  |
| --- | --- |
| 95% | ± 2.063 |
| 96% | ± 2.171 |
| 99% | ± 2.796 |

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: (Solution in Jupyter file)

t-score = -0.4714,  
Degree of freedom = 17   
P(t) = 0.3216725