Q1) Identify the Data type for the Following:

Answer (Q1): Answers embedded in the table below in blue-colored text.

|  |  |
| --- | --- |
| 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.

Answer (Q2): Answers embedded in the table below in blue-colored text.

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

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

Answer (Q3):

The probability (P) of getting heads (H) or tails (T) on a single coin toss is 50% i.e. 1/2.

First toss - P(HHT) = (1/2) \* (1/2) \* (1/2) = 1/8

Second toss - P(HTH) = (1/2) \* (1/2) \* (1/2) = 1/8

Third toss - P(THH) = (1/2) \* (1/2) \* (1/2) = 1/8

Thus, 1/8 + 1/8 + 1/8 = 3/24 i.e. 1/8

Thus, the probability of getting two heads and one tail when three coins are tossed is 1/8.

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

Answer (Q4):

Abbreviations - Probability (P), Favorable Outcomes (FO), Total Outcomes (TO)

1. The minimum value on each dice is 1. So, the probability of equal to 1 is 0 (Zero).
2. Less than or equal to 4 can have six favorable outcomes!

When two dice are rolled:

outcomes less than 4 🡪 (1, 1), (1, 2), (2, 1)

outcomes equal to 4 🡪(1, 3), (3, 1), (2, 2)

Each dice has six sides! So, the total number of possible outcomes will 6 \* 6 = 36.

P = FO / PO | P = 6 / 36 i.e. 1/6

Probability of getting less than or equal to 4 when two dice are rolled is 1/6.

1. To be divisible by both 2 and 3, the sum of both the dice has to be 6!

The sum of both the dice that is divisible by 2 are (1, 5), (2, 4), (3, 3), (4, 2)

The sum of both the dice that is divisible by 3 are (5, 1), (6, 6)

P = FO / PO

P = 6 / 36

i.e. 1/6

Probability of getting the Sum that is divisible by 2 and 3 is 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?

Answer (Q5):

* Non-blue balls = 2 (red) + 3 (green) = 5
* Total balls in the bag = 2 (red) + 3 (green) + 2 (blue) = 7

So, the probability of drawing a non-blue ball from the bag in the first draw is 5/7 because there are 5 non-blue balls out of a total of 7 balls.

And, the probability of drawing a non-blue ball in the second draw, assuming that the first draw was a non-blue ball, is 4/6 because there are 4 non-blue balls left after the first draw and 6 balls left in total in the bag.

Thus, the total probability that none of the balls drawn is blue = (5/7) \* (4/5) = 20/35 i.e. 4/7.

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

Answer (Q6):

Expected candies = (candies of each child \* probability of each child)

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

i.e. Expected candies = 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 = 3.09

Thus, randomly selected child can get 3 candies.

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**

Answer (Q7): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 - Q7.ipynb”

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?

Answer (Q8): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q8.ipynb”

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer (Q9\_a): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q9\_a.ipynb”

**SP and Weight (WT)**

**Use Q9\_b.csv**

Answer (Q9\_b): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q9\_b.ipynb”

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



Answer (10):

Inference drawn:

Histogram of ChickWeight$weight - The histogram of chick weight reveals a positively skewed distribution, indicating that more chick weights are above the mean than below it. The distribution also displays some level of peakedness, suggesting that many chick weights are close to the average.

Boxplot - The points above the T-shaped whiskers, which are 1.5 times the interquartile range, indicate that there are outliers in the data.

**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?

Answer (11): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q11.ipynb”

**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?

Answer (12): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q12.ipynb”

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

Answer (13):

There is no skewness in the data when the mean and median are equal, the data does not exhibit a tail in one direction, be it positive or negative, which is nature of skewed distributions.

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

Answer (14):

When the mean is greater than the median, the nature of the skew is positive, and there is a longer tail on the right side of the distribution due to the presence of larger values.

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

Answer (15):

When the median is greater than the mean, the nature of the skew is negative, and there is a longer tail on the left side of the distribution due to the presence of smaller values.

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

Answer (16):

A positive kurtosis value indicates that the data is more peaked and has heavier tails, indicating the presence of extreme values and less spread in the distribution.

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

Answer (17):

A negative kurtosis value indicates that the data has a flatter peak and lighter tails, indicating the presence of fewer extreme values and more spread in the distribution

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



Answers (18)

What can we say about the distribution of the data?

This data has longer left-side tail compared to the right. This suggests a higher concentration of data points below the median.

What is nature of skewness of the data?

This data has a negative skew. This is evident from the fact that the lower whisker is much longer than the upper whisker.

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

The IQR of the data is approximately 8 by subtracting the lower quartile from the upper quartile. The lower quartile is at 10 and the upper quartile is at 18, so the IQR is 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.

Answer (19):

Boxplot 1 exhibits lower variance compared to Boxplot 2, evident through its shorter IQR in contrast to Boxplot 2.

Neither boxplot displays any outliers, as no data points are plotted beyond the upper or lower bounds of the T-shaped whiskers.

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)

Answer (20): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q20.ipynb”

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

Answer (21): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q21.ipynb”

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

Answer (22): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q22.ipynb”

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

Answer (23): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q23.ipynb”

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

Answer (24): Please refer to the attached notebook “Basic\_Statistics\_Level\_1 – Q24.ipynb”