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

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

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

Soln:

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 where sum is**

1. **Equal to 1**
2. **Less than or equal to 4**
3. **Divisible by 2 and 3**

Soln:

Number of possible outcomes for the above event is

N (Event (Two dice rolled)) = = 36

1. P (sum == 1) = 0.
2. P (sum <= 4) = N (Event (sum <= 4)) / N (Event (Two dice rolled))

= 6 / 36 = 1/6 = 0.166 = 16.66%

1. P (sum / 2 & 3) = N (Event (sum / 2 & 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?**

Soln:

Total no.of balls in the bag = 2 red + 3 green + 2 blue = 7 balls

Total no.of ways to draw 2 balls from the bag:

This can be calculated using combinations.

=

= = 21

No.of ways that none of the balls drawn is blue = 7 – 2 = 5 balls

= = 10

Therefore, the probability of none are blue balls:

P(none are blue) =

**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**

Soln:

Expected no.of candies = (1\*0.015) + (4\*0.20) + (3\*0.65) + (5\*0.005) + (6\*0.001) + (2\*0.120) = 3.085

Therefore, the Expected number of candies for a randomly selected child is approximately 3.085

**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**

Soln:

For Points:

Mean for Points = 3.59,

Median for Points = 3.69,

Mode for Points = 3.07,

Variance for Points = 0.28,

Standard Deviation for Points = 0.53,

Range [Min-Max] for Points [3.59 – 4.93]

For Score:

Mean for Score = 3.21

Median for Score = 3.32

Mode for Score = 3.44

Variance for Score = 0.95,

Standard Deviation for Score = 0.97,

Range [Min-Max] for Score [3.21 – 5.42]

For Weigh:

Mean for Weigh = 17.84

Median for Weigh = 17.71

Mode for Weigh = 17.02

Variance for Weigh = 3.19

Standard Deviation for Weigh = 1.78

Range[Min-Max] for Weigh [17.84 – 22.9]

Draw Inferences



**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?**

Soln:

Expected value = Sum (X \* Probability of X)

Expected value = = 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data (Car’s speed and distance)**

**Use Q9\_a.csv**

Soln:

df1 = pd.read\_csv(‘q9\_a.csv’)

Calculate skewness and kurtosis for speed and distance series:

speed\_skewness = skew(df1.speed)

speed\_kurtosis = kurtosis(df1.speed)

print(“Speed skewness: ”, speed\_skewness)

print(“Speed kurtosis: ”, speed\_kurtosis)

For Cars Speed Skewness value= -0.114 and Kurtosis value= -0.51

distance\_skewness = skew(df1.dist)

distance\_skewness = skew(df1.dist)

print(“Distance skewness: ”, dist\_skewness)

print(“Distance kurtosis: ”, dist\_kurtosis)

For Cars Distance Skewness value = 0.78 and Kurtosis value = 0.41

**SP and Weight (WT)**

**Use Q9\_b.csv**

Soln:

Df2 = pd.read\_csv(‘q9\_b.csv’)

Calculate skewness and kurtosis for SP and Weight(WT) series:

sp\_skewness = skew(df1.SP)

sp\_kurtosis = kurtosis(df1.SP)

print(“SP skewness: ”, sp\_skewness)

print(“SP kurtosis: ”, sp\_kurtosis)

For SP, Skewness value= 1.581 and Kurtosis value= 2.724

wt\_skewness = skew(df1.WT)

wt\_kurtosis = kurtosis(df1.WT)

print(“Weight skewness: ”, wt\_skewness)

print(“Weight kurtosis: ”, wt\_kurtosis)

For Weight, Skewness value= 1.581 and Kurtosis value= 2.724

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



Soln: The Histogram’s peak has right skew and tail is on right.

Mean > Median. Outliers are on the higher side.



Soln: 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?**

Soln:

Considering it as a normal distribution

Given:

n = 2000

= 200 pounds

s = 30

Since population standard deviation is not known,

CI = ± t(1-α,n-1)\*s/n^(1/2)

For 94%,

t(0.97, 1999) = 1.881861

CI = 200± 1.881861 \* 30/2000^(1/2) = 200 ± 1.262 = 198.74 to 201.26

For 98%,

t(0.99,1999) = 2.328215

CI = 200± 2.328215 \* 30/2000^(1/2) = 200 ± 1.5618 = 198.43 to 201.56

For 96%,

t(0.98,1999) = 2.05509

CI = 200± 2.05509 \* 30/2000^(1/2) = 200 ± 1.378596 = 198.62 to 201.388

**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.**

Soln:

Mean =41

Median =40.5

Variance =24.11

Standard Deviation =4.91

1. **What can we say about the student marks?**

Soln: Outliers aren’t present and the data is slightly skewed towards right because mean is greater than median.

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

Soln: No skewness is present we have a perfect symmetrical distribution

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

Soln: Skewness and tail is towards Right

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

Soln: Skewness and tail is towards left

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

Soln: Positive kurtosis meSoln the curve is more peaked and it is Leptokurtic

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

Soln: Negative Kurtosis meSoln the curve will be flatter and broader

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



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

Soln: The most of the data points are distributed towards right side of the curve

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

Soln: Negatively Skewed

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

Soln: IQR = Q3 – Q1 = 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.**

Soln: There are no outliers, and both box plots show a similar median around 250 to 275. The distributions are approximately normal, with little to no skewness towards the minimum or maximum whiskers.

**Q20) 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.**

* 1. **P(MPG>38)**
  2. **P(MPG<40)**
  3. **P (20<MPG<50)**

Soln:

N (Event (Total no.of Cars)) = 81

1. P (MPG > 38) = N (Event (MPG > 38)) / N (Event (Total no.of Cars))

= 53 / 81 = 0.654 = 1 – 0.654 = 0.345 = 34.5%

1. P (MPG < 40) = N (Event (MPG < 40)) / N (Event (Total no.of Cars))

= 61 / 81 = 0.729 = 73%

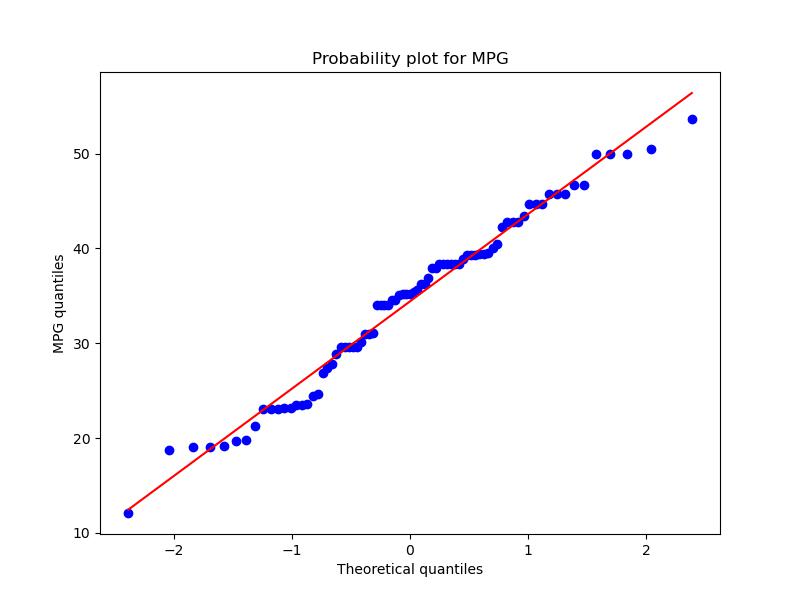
1. P (20 < MPG < 50) = [N (Event (MPG < 50)) - N (Event (MPG < 20))] / N (Event (Total no.of Cars))

= (76 – 7) / 81 = 0.899 = 89%

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

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

Soln: MPG of cars follows normal distribution



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

**Dataset: wc-at.csv**

Soln: Adipose Tissue (AT) and Waist does not follow Normal Distribution

A comparison of a graph

Description automatically generated

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

Soln:

For 90% C.I, z score = 1.645

For 94% C.I, z score = 1.881

For 60% C.I, z score = 0.842

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

Soln:

For 95% C.I, t score = 2.063

For 96% C.I, t score = 2.172

For 99% C.I, t score = 2.797

**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**

Soln:

Since the population standard deviation is not known and sample size is less than 30

Sample mean, = 260

Population mean, = 270

Sample standard deviation, s = 90

Sample size, n = 18

t = / (s / ) = (260 – 270) / (90 / = -0.47