|  |  |
| --- | --- |
| 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) | Categorical |

**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 | Nominal |
| Celsius Temperature | Interval |
| 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: when three coins are tossed sampled space is given as

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

Therefore, n(S)=8

To find probability of two heads and one tail

Probability p(A)=number of favourable outcome n(A)/number of total outcomes n(S)

A= HHT, HTH, THH

P(A)=3/8

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

1. **Equal to 1**

Ans: when we roll two dice there are 6 × 6 = 36 possibilities.

Upon rolling two dice, the minimum sum on their faces would be 1 + 1 = 2.So, the sum is never 1.

Therefore the probability is 0

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

Ans: when we roll two dice there are 6 × 6 = 36 possibilities.

When we roll two dice, the possibility of getting number 4 is (1, 3), (2, 2), and (3, 1).

So,

The number of favorable outcomes = 3

Total number of possibilities = 36

Probability p(A) = The number of favorable outcomes / Total number of possibilities = 3 / 36 = 1/12.

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

Ans: when we roll two dice there are 6 × 6 = 36 possibilities.

The only numbers divisible by 2 and 3 are 6,12

Where 12 can be made by 1 way(double 6) and 6 by 5 ways(1,5)(5,1)(4,2)(2,4)(3,3)

Probability p(A) = The number of favorable outcomes / Total number of possibilities=(1+5)/36=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 number of balls = (2 + 3 + 2) = 7.

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7=7C2 = 21

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

n(E) = Number of ways of drawing 2 balls out of (2 + 3)

balls =5C2 = 10

  Therefore, 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**

Ans: Expected number of candies for a randomly selected child=

1\*0.015+2\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120= **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: refer file assignment1Q7.ipynb in folder Assignment 1.

**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: a). 108+110+123+134+135+145+167+187+199/9=145.33

**Expected value of the weight of the patient is: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**

**Ans: refer assignment1Q9.ipynb**

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



Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

According to the boxplot, it 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?**

Ans: Confidence Interval = X’ ±t(s/√n)

Where X, is Mean = 200

s is standard deviation = 30

N is sample size = 2000

CI94% :

α = 1 - (confidence level / 100) = 1- (94%/100) = 0.06

Critical probability (p\*) = 1 - α/2 = 1-0.03 = 0.97

Degrees of freedom =n-1=2000-1=1999

t-score =1.882

CI94% = 200±1.2620

CI98% :α = 1 - (confidence level / 100) = 1-(98%/100)=0.02

Critical Probability (p\*) = 1 - α/2 = 0.99

Degrees of freedom =n-1=2000-1=1999

t-score = 2.328

CI98% = 200± 1.561

CI96%:α = 1 - (confidence level / 100) = 1 – (96%/100) =0.04

Critical Probability (p\*) = 1-α/2 = 0.98

Degrees of freedom =n-1=2000-1=1999

t-score = 2.055

CI98%= 200±1.378

**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: Mean= sum of all scores / number of scores

Mean(µ)= 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56

​ 18

Mean=738/18=41

Median: The median is the middle value when the data is sorted in ascending order.After sorting the data: 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56.

The median is the average of the two middle values (since there is an even number of data points), which are 40 and 41.

Therefore, the median is 40+41/2=40.5

Variance:

Variance=

=++………..+/18

=25.23

Standard deviation:

Standard deviation=

= =5.022

What can we say about the student marks?

Ans) Maximum no.of students got 35-40 range score.No one get 50-55 range score

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

Ans: Normalized Skewness

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

Ans: Right skew

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

Ans: Left skew

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

Ans: Sharp peak in the plot, less gap between tails to x-axis

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

Ans: Broader peak under the curve and more gap between tails and x-axis

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



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

Ans: The data is distributed in De-assigned format

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

Ans: Left side skew

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

Ans:Q3-Q1=8

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



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

Ans: The Box plot 1 designed with range = 3, Second one range is = 1.5

**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**

**a.P(MPG>38)**

**b.P(MPG<40)**

**c. P (20<MPG<50)**

Ans: REFER Q20.ipynb

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

**a.Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

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

**Dataset: wc-at.csv**

Ans: Refer Q21.ipynb

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

Ans: Z score of 90% = 1.65

Z score of 94% = 1.56

Z score of 60% = 0.26

**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%= 2.064

T score of 94%= 2.171

T score 0f 99%= 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**

ANS: Given:

Population mean (claimed by the CEO): μ = 270 days

Sample size: n = 18 bulbs

Sample mean: x̄ = 260 days

Standard deviation of the population: σ = 90 days

First, let's calculate the standard error of the mean (SEM), which is the standard deviation of the sampling distribution of the sample mean:

SEM = σ / √n

Where:

σ is the population standard deviation

n is the sample size

SEM = 90 / √18 ≈ 21.213

Now, let's find the z-score for x̄ = 260 days:

z = (x̄ - μ) / SEM

z = (260 - 270) / 21.213 ≈ -0.471

Now, we find the probability corresponding to this z-score using a standard normal distribution table or calculator. The probability of finding a z-score less than or equal to -0.471 is approximately 0.3208.

Therefore, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days, given the CEO's claim, is approximately 0.3208 or 32.08%.