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

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

Ans: When 3 coins are tossed, the possible outcomes are HHH, HHT,HTH,THH,TTH,THT,HTT,TTT

In case of 2 head & tail ,we will get the possible outcome as HHT,HTH,TTH

The probability that two heads and one tail =3/8

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

Ans : To find the probability of different sums when two dice are rolled, we can use a probability table. Each die has 6 sides, numbered 1 through 6. There are a total of 6 x 6 = 36 possible outcomes when rolling two dice.

1. Equal to 1

Ans : There is only one way to get a sum of 1, which is rolling a 1 on the first die and a 1 on the second die. So, there is 1 favorable outcome.

P(sum = 1) = (Number of favorable outcomes) / (Total possible outcomes) = 1/36

1. Less than or equal to 4

Ans: Probability that the sum is less than or equal to 4:

The sums less than or equal to 4 can be achieved in the following ways:

- Sum = 2 (1+1)

- Sum = 3 (1+2, 2+1)

- Sum = 4 (1+3, 3+1, 2+2)

There are 10 favorable outcomes in this case.

P(sum ≤ 4) = (Number of favorable outcomes) / (Total possible outcomes) = 10/36 = 5/18

1. Sum is divisible by 2 and 3

Ans: ) Probability that the sum is divisible by 2 and 3:

To have a sum divisible by both 2 and 3, we're looking for sums that are divisible by 6. The only way to get a sum of 6 on two dice is by rolling a 3 on each die.

P(sum is divisible by 2 and 3) = P(sum = 6) = 1/36

So, the probabilities are as follows:

a) P(sum = 1) = 1/36

b) P(sum ≤ 4) = 5/18

c) P(sum is divisible by 2 and 3) = 1/36

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: nCr=n!/n!\*(n-r)!

* Two balls are drawn at random

n=7 , r=2

7C2=7!2!\*(7-2)!

=5040/2\*120

=5040/240

=21 ways

* None of the balls drawn is blue

N=5, r=2

5C2=5!/2!\*(5-2)!

=120/2\*6

=120/12

= 10 ways

* Possible events out of total number of events

=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: . To calculate the expected number of candies for a randomly selected child, you can multiply the candy count for each child by their respective probabilities, and then sum up these values. The formula for calculating the expected value (E) is:

E = Σ (X \* P)

Where:

- E is the expected value (in this case, the expected number of candies).

- X is the value (candy count).

- P is the probability.

Using the provided data:

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

E = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

E = 3.115

So, the expected number of candies for a randomly selected child is 3.115.

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**

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 :** . To calculate the expected value of the weight of a randomly chosen patient, you need to find the mean (average) of the given weights.

The formula for the expected value (E) is:

E = (ΣX) / N

Where:

- E is the expected value.

- ΣX is the sum of all the values (weights in this case).

- N is the number of values (number of patients).

First, sum up the weights:

ΣX = 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199

ΣX = 1098

Now, calculate the expected value:

E = ΣX / N = 1098 / 9 = 122

So, the expected value of the weight of a randomly chosen patient is 122 pounds.

**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**

|  |  |
| --- | --- |
| SPEED | DISTANCE |
| 4  4  7  7  8  9  10  10  10  11  11  12  12  12  12  13  13  13  13  14  14  14  14  15  15  15  16 | 2  10  4  22  16  10  18  26  34  17  28  14  20  24  28  26  34  34  46  26  36  60  80  20  26  54  32 |

|  |  |
| --- | --- |
| SP | WT |
| 104.1854  105.4613  105.4613  113.4613  104.4613  113.1854  105.4613  102.5985  102.5985  115.6452  111.1854  117.5985  122.1051  111.1854  108.1854  111.1854  114.3693  117.5985  114.3693  118.4729  119.1051  110.8408  120.289  113.8291  119.1854 | 28.76206  30.46683  30.1936  30.63211  29.88915  29.56177  30.30848  15.84776  16.35948  30.92015  29.36334  15.75353  32.81359  29.37844  29.34728  29.60453  29.53578  16.19412  29.92939  33.51697  32.32465  34.90821  32.67583  31.83712  28.78173 |
| 114.5985  120.7605  119.1051  99.56491  121.8408  113.4846  112.289  119.9211  121.3926 | 16.04317  38.0682  32.83507  34.48321  35.54936  37.04235  33.23436  31.38004  37.57329 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| SPEED | DISTANCE |
| 4  4  7  7  8  9  10  10  10  11  11  12  12  12  12  13  13  13  13  14  14  14  14  15  15  15  16 | 2  10  4  22  16  10  18  26  34  17  28  14  20  24  28  26  34  34  46  26  36  60  80  20  26  54  32 |

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



**Sol: Medican is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.**



**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 for 94%

200+1.88\*30/√2000

=201.26

OR

200-1.88\*30/√2000

= 198.73

(201.26,198.73)

Confidence interval for 98%

200+2.33\*30√2000

=201.56

OR

200-2.33\*30/√2000

= 198.43

(201.56,198.62)

Confidence interval for 96%

200+2.05\*30/√2000

=201.37

OR

200-2.05\*30/√2000

=198.62

(201.37,198.62)

**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: 1) mean=40

Median=41

SD=4.9

Variance=24.1

2)By considering various measures calculated above,there is a chance of outlayer present in the data.or the data could be normal/skewed.

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

Ans: Symetrical

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

Ans: Right skewed

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

Ans: Left skewed

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

Ans: The data is notmally distributed and kurtosis value is 0.

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

Ans: The distribution of the data has lighter tails and a flatter peaks than the normal distribution

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
Ans: Let’s assume above box plot is about age’s of the students in a school. 50% of the people are above 10 yrs old and remainig are less. And students who’s age is above 15 are approx 40%.

We can say that the data is left skewed.median is greater than mean.

The approximate value of IQR will be 8.

Q19) Comment on the below Boxplot visualizations?



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

Ans: Here there is a representation of 2 box plots.(2) is highly distributed across

the plane.(1) is slightly less distributed .shows variance.

Whiskers in these diagrams show the 100% of the data is spread across values from 350 in 2 whereas it has spread in range 250-290 in (1).

When we compare box plot 1 with box plot 2 we can say that the data in boxplot is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots are equal. And the data spread in both of them are symmetrical.

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)

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

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

Ans:

Z scores

90%

=95+2.5

=97.5

= qnorm(0.975)

=1.96

94%

=qt(0.98,24)

=2.171545

99%

=qt(0.995,24)

=2.79694

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

Ans: TSCORE CALCULATION

T(1,alpha),(n-1))

Here n=25

n-1=24

hence T square value will be

95%

=qt(0.975,24)

=2.063899

96%

=qt(0.98,24)

=2.171545

99%

=qt(0.995,24)

=2.79694

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:

S=90 , x =260 , µ=270 , sample size =18

t = x-µ/s/√n

t = 260-270/90/√18

t= -10/90/√18

t= -0.471

For probability t Value is (n-1)=17 degree of freedom and a t-score of -0.471. This probability is approximately 0.3217 12.Therefore,there would be a 32.17% chance of getting a sample mean of 260 days or less. Since this probability is greater than the significance level of 0.05, we fail to reject the null hypothesis. In other words, there is insufficient evidence to conclude that the average life of a bulb is less than 270 day