|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Ratio |
| Results of rolling a dice | Nominal |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Ratio |
| Number of tickets in Indian railways | Ratio |
| Number of times married | Ratio |
| Gender (Male or Female) | 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 | 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 | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

The possible outcomes for tossing three coins are HHH,HHT,HTH,THH,TTH

THT,HTT,TTT.The probability of getting output of one tail two heads is 3/8.

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

1. Equal to 1
2. Less than or equal to 4-6/36 = 1/6
3. Sum is divisible by 2and 3-12/36=1/3

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?

The probability is

Number of favourable outcomes/Total number of outcomes

So it is 5/7.

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

E(X)=∑i​P(Xi​)⋅xi​

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

(0.20⋅4)+(0.65⋅3)+(0.005⋅5)+(0.01⋅6)+(0.120⋅2)

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?

**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**

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



**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?

Critical Value×Sample Size

​Sample Standard Deviation​

94% Confidence Interval:

CI94%=200±(1.88×302000)

98% Confidence Interval:

CI98%=200±(2.33×302000)

96%confidence interval:

CI96%=200±(2.05×302000)

**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.

 **Mean (Xˉ):** Mean=∑i=1nXi/n

Mean=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

 **Median:** Since the data is already sorted:

Median=40

 **Variance (s2):** Variance=∑i=1n(Xi−Xˉ)2/n−1

Variance=(34−41)2+(36−41)2+…+(56−41)2/17

 **Standard Deviation (ss):** Standard Deviation=Variance

1. What can we say about the student marks?

Mean(~X):

Mean ≈ 41

Median :

Median = 40

Variance(s2):

Variance ≈ 14.84

Standard deviation(s):

Standard deviation ≈ 3.85

Interpretation:

The mean score is approximately 41.

The median is 40, indicating a slightly right-skewed distribution.

The spread of the scores is moderate, with a standard deviation of around 3.85.

There is a concentration of scores around the mean, with a few higher scores that contribute to the right skewness.

Conclusion:

The student's performance is centered around the mean score of 41.

The majority of scores cluster around the median and mean, with a few higher scores.

The variability in scores is moderate, suggesting a relatively consistent performance with a few notable exceptions

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

Ans:If the mean and median are equal, the distribution is symmetric and has zero skewness.

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

Ans: If the mean is greater than the median, the distribution is positively skewed (right-skewed)

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

Ans: If the median is greater than the mean, the distribution is negatively skewed (left-skewed).

Q16) What does positive kurtosis value indicates for adata ?

Ans: A positive kurtosis value indicates that the data has heavier tails and a sharper peak than a normal distribution. It suggests that the data has more outliers.

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

Ans: : A negative kurtosis value indicates that the data has lighter tails and a flatter peak than a normal distribution. It suggests that the data is less extreme than a normal distribution

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



What can we say about the distribution of the data?

Ans: The distribution can be assessed by looking at the spread, central tendency, and presence of outlier

What is nature of skewness of the data?

Ans: Skewness can be identified based on the position of the median relative to the quartiles

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

Ans: The IQR can be estimated by the length of the box in the boxplot.

Q19) Comment on the below Boxplot visualizations?



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

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

Ans:It rejects the null hypothesis

1. Check whether the MPG of Cars follows Normal Distribution

Ans:It rejects the null hypothesis

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

Ans:It rejects the null hypothesis

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

Ans: confidence\_intervals <- c(0.90, 0.94, 0.60)

z\_scores <- qnorm((1 + confidence\_intervals) / 2)

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

Ans: onfidence\_intervals <- c(0.95, 0.96, 0.99)

degrees\_of\_freedom <- 25 – 1

t\_scores <- qt((1 + confidence\_intervals) / 2, df = degrees\_of\_freedom)

Q 24**)**A Government companyclaims 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

The formula for the z-score is given by:

z=xˉ−μσnz

​σ​xˉ−μ​

where:

xˉ is the sample mean,

μ is the population mean (CEO's claim),

σ is the population standard deviation,

n is the sample size.

In this case:

xˉ=260 days (sample mean),

μ=270 days(population mean according to CEO's claim),

σ=90 days (population standard deviation),

n=18 (sample size).

Plug these values into the formula:

z=260−270/90/18

Calculate the z-score:

z=−109018z=18

z≈−10904.24≈−1021.23≈−0.47

Now, you can look up the probability corresponding to z=−0.47 in the standard normal distribution table. The probability represents the likelihood that the average life of 18 randomly selected bulbs is no more than 260 days.

Using a standard normal distribution table or a calculator, you can find the probability. For z=−0.47 the probability is approximately 0.3192.

Therefore, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is approximately 0.3192or 31.92

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom