|  |  |
| --- | --- |
| 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 | 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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

ANS : Total number of possible combination are = 2^3 = 8 .

The combinations are HHH, HHT, HTH, THH, THT, HTT, TTT, TTH.

Combinations that have 2 head and 1 tail is HHT, HTH, TTH

The probability of two heads and one tail is = **3/8 = 0.37**

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 2and 3

ANS: Two dice are rolled = 36

1. 0
2. Sum of 2: 1

Sum of 3: (1,2), (2,1)

Sum of 4: (1,3), (2,2), (3,1)

- Number of favorable outcomes = 1+2+3 = 6

-Total number of outcomes = 36

P(sum <= 4) = Number of favorable outcomes/ Total number of outcomes

= 6/36

1. Sum divisible by 2 = 2, 4, 6, 8, 10, and 12

Sum divisible by 3 = 3, 6, and 9

Now, we need to find the sums that are divisible by both 2 and 3, which mean they need to be multiples of 6. The sums that meet this criterion are 6 and 12.

To get a sum of 6, you can have the following combinations of the two dice:

(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

To get a sum of 12, you can have the following combination:

(6, 6)

Total favorable outcomes = 5 (for sum = 6) + 1 (for sum = 12) = 6 favorable outcomes.

P(sum is divisible by 2 and 3) = Number of favorable outcomes / Total number of outcomes

P(sum is divisible by 2 and 3) = 6 / 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 : Total balls are 2 red + 3green + 2blue = 7

(5/7) \* (4/6) = 20/42 = 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 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090

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 : Range = 32**

**Points : Mean = 3.596563, Median = 3.695000, Mode = ‘numeric’, SD = 0.534679 , Variance = 0.28588**

**Score: Mean = 3.217250, Median = 3.32500, Mode = ‘numeric’, SD =0.978457, Variance = 0.95737**

**Weight = 17.848750, Median = 17.7100, Mode = ‘numeric’, SD = 1.786943, Variance = 3.193166**

**Mean values are closer for both Score and Points.**

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: EX = 108, 110, 123, 134, 135, 145, 167, 187, 199

P(X) = 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9

Expected Value = (1/9)( 108+110+123+134+135+145+167+187+199)

= (1/9) (1308) = 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS : Skewness value for speed is -0.11395 negative , so left skewed,**

**Kurtosis value for speed is -0.577147**

**Skewness value for distance is 0.7824 positive, so right skewed,**

**Kurtosis value for distance is 0.24801**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS : Skewness value for SP is 1.5814 positive , so right skewed,**

**Kurtosis value for SP is 2.723**

**Skewness value for Weight is -0.603 negative, so left skewed,**

**Kurtosis value for Weight is 0.819**

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



ANS: The most of the data points are concerated in the range 50-100 with frequency 200.

Skewness-we can notice a long tail towards right so it is heavily right skewed.



ANS: Median is less than mean right skewed and we have outlier on the upper side 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: X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%(1-σ/2)= 1-0.03) =0.97 for

Confidence interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

**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 = 41, Median = 40.5, variance = 25.529, SD = 5.05266

2) we can say that mean of marks of student is 41 which is slightly greater than median. Most of the students got marks in between 41-42

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

ANS: Symmetrical

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 adata ?

ANS: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

ANS: A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

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%.

What is nature of skewness of the data?

ANS: Left skewed, median is greater than mean.

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

ANS: Approximately= -8

Q19) Comment on the below Boxplot visualizations?



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

ANS: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symetrical.

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)

ANS : a) 0.347

1. 0.7294
2. 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the   
   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 scoresof 90% confidence interval,94% confidence interval, 60% confidence interval

ANS: 90% = 1.644, 94% = 1.88, 60% = 0.841

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

ANS: 95% =1.96 , 96% =2.5 , 99%=2.47

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

Hint:

rcode🡪pt(tscore,df)

df 🡪 degrees of freedom

ANS: µ = 270, x̅ = 260, s = 90, n = 18

SE = s/ sqrt(n)

SE = 90 / sqrt(18) ≈ 21.21

z = (x̄ - μ) / SE

z = (260-270)/21.21 = -0.471

P(z <= -0.471) = stats.norm.cdf(-0.471) = 0.318