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

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

Ans: Possible outcomes:

S = {(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)}

Total number of outcomes = n(S) = 8

Let A be the event that two heads and one tail are obtained

A = {(H,H,T), (H,T,H), (T,H,H)}

n(A) = 3

Hence, the probability of getting two heads and one tail = n(A) / n(s) = 3/8

Thus, the probability of getting two heads and one tail on tossing three coins at once is equal to 3/8.

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

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

Ans: Two Dice are rolled :

Sample Space (S) = {(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6 )

(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6)

(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)

(4, 1), (4, 2), (4, 3), (4, 4), (4, 5) (4, 6)

(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6)

(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)}.

n(S) = 36

1) Let A be the event that sum is equal to 1

But, The minimum possible sum is (1, 1) = ( 2 ).

n(A) = 0

Therefore P( A ) = n(A)/ n(S) = ( 0 )/( 36 ) = 0

Hence, the probability that sum is equal to 1 is 0.

2) Let B be the event that the sum is less than or equal to 4

B = {(1, 1), (1, 2), (1, 3),(2, 1), (2, 2), (3, 1)}

n(B) = 6

Therefore, P(B) = n(B)/n(S) = 6/36 = 1/6

Hence, the probability less than or equal to 4 is 1/6.

3) Let C be the event that the Sum is divisible by 2 and 3

C = {(1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)}

n(C) = 5

Therefore, P(C) = n(C)/n(S) = 5/36

Hence, the probability that sum is divisible by 2 and 3 is 5/36.

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 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

n(S) = 7C2 =(7×6) / (2×1) = 21

Let A be the event of 2 balls, none of which is blue.

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

n(A) = 5C2 = (5×4) / (2×1) =10

Therefore, P(A) = n(A) / 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 + 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

= 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

Q8) Calculate Expected Value for the problem below

* 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: The mean is:

108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 = 145.333.

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



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

And least range of weight is 400 somewere around 0-10.

So the expected value the above distribution is 75.

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



Ans: In this, tha data is concentrated downwards and Q2-Q1<Q3-Q2 which shows positive skewness and has some outliers above the upper quartile.

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?

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

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

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 normally 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?

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 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

* P(MPG>38)
* P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* 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

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

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