|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Ordinal |
| 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 | Discrete |
| Number of tickets in Indian railways | Ordinal |
| Number of times married | Discrete |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Intervals |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

**Answer:**

**Probability of an event (E) = Number of favorable outcomes / Total number of outcomes**

Let, H = Heads, T = Tails

Possible outcomes:

(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 = 8

Number of outcomes that gives two heads and one tail = 3

i.e, (H,H,T), (H,T,H), (T,H,H)

Thus, number of favorable outcomes = 3

Probability of getting two heads and one tail = Number of favorable outcomes / Total number of outcomes

= 3/8

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 2 and 3

**Answer:**

1. The dice are “ fair “, that is, not biassed in any manner.
2. The dice are both six-sided dice, that is both have 6 faces, with each face on each dice, showing one of the numbers, 1 to 6, with no number repeated on the same dice.

Analysis

With two dice, there are ( 6 ) \* ( 6 ) = ( 36 ) possible combinations of numbers.

The minimum sum possible for the two dice thrown is (1, 1) = a sum of (2 )

The maximum sum possible for the two dice thrown is (6, 6) = a sum of (12).

**Sum = (1).**

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

Therefore P( 1 ) = ( 0 )/( 36 ) = 0

**Sum = (4)**

A sum of (4) can be achieved with number combinations ( 2, 2), (1, 3) and (3, 1), that is only with 3 combinations of numbers**.**

P(sum = 4) = ( 4 / 36 ) = ( 1 / 9 )

**Sum = (less than 13 )**

To meet this requirement:

0 < sum =/< 12

The combinations which meet this requirement are:

(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).

which gives a total of 36 combinations.

Note however we needn't have written out all the possible combinations as I did above. All I needed to do, to calculate the.probability that the sum was less than ( 13 ) was to recognize that a sim of (1) to (12) inclusive included every possible combinations of numbers from two dice.

**P(sum < 13) = ( 36 / 36 ) = 1 = 100%**

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?

Answer:

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

n(S)=(7\*6)/(2\*1)

n(S)=21

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

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

n(E)=5c2

n(E)=(5\*4)/(2\*1)

n(E)=10

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

Answer:

1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

= 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

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?

**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?

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

Data is normalized and there is no skewness

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

Negative skewness implies mass of the distribution concentrated on right side

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

Positive Skewness implies mass of the Distribution concentrated on left side

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

Positive kurtosis value indicates that thinner peak and wider tails

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

Negative kurtosis value indicates that wider peak and thinner tails.

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

What can we say about the distribution of the data?

Not normally distributed

What is nature of skewness of the data?

Negative skewness

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

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 of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

pnorm(50,34.422,9.13144)

-

(

1

-

pnorm(20,34.422,9.13144)

)=

0.01311818

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

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

Ans:- 52.86%