|  |  |
| --- | --- |
| 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 | Ratios |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratios |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratios |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratios |
| Religious Preference | Nominal |
| Barometer Pressure | Ratios |
| SAT Scores | Interval |
| Years of Education | Ratios |

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

Solution

Find the probability of two heads and one tail:  
Three coins are tossed at once.  
The possible toss outcomes will be (HHH, TTT, HTT, THT, TTH, THH, HTH, HHT).  
We have to find the probability of two heads and one tail:  
So, favorable outcomes = THH, HTH, HHT  
Number of favourable outcomes = 3   
Hence required probability is:   
Probability of event to happen P(E) = (Number of favourable outcomes) / (Total Number of outcomes)   
Probability (2H & 1T) = 3/8  
Hence, when 3 coins are tossed. the probability of two heads and one tail is 3/8.

ANS: P(E) = n(E) / n(S) = 3/8 = 0.375

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

(a). Equal to 1

Solution

If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

(b). Less than or equal to 4

Solution

If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum >=4) = 6 {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

Hence, probability is 6/36 or 0.166.

(c). Sum is divisible by 2 and 3

Solution

If two dices were rolled, then total possible cases =36

Total Favourable cases (Having Sum is divisible by 2 and 3) = 6 {(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)}

Hence, probability is 6/36 or 0.166.

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?

Solution

Total number of balls in the bag = 2 (red) + 3 (green) + 2 (blue) = 7

C(7, 2) = 7! / (2! \* (7-2)!) = 7! / (2! \* 5!) = (7 \* 6) / (2 \* 1) = 21

C(5, 2) = 5! / (2! \* (5-2)!) = 5! / (2! \* 3!) = (5 \* 4) / (2 \* 1) = 10

So, the number of favorable outcomes is 10.

probability of not drawing any blue ball is given by:

Probability = (Number of favorable outcomes) / (Total number of outcomes)

Probability = 10 / 21 ≈ 0.4762

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

Solution:

Expected number of candies for a randomly selected child

= (1 x 0.015 ) + (4 x 0.20) + (3 x 0.65) + (5 x 0.005) + (6 x 0.01) +( 2 x 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.

ANS:

**Mean: Median: Mode:**

Points => 3.596563 Points => 3.695 Points => 3.07, 3.92

Score => 3.21725 Score => 3.325 Score => 3.44

Weight => 17.84875 Weight => 17.71 Weight => 17.02, 18.90

**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: = (1/9) X (1308)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

ANS:

**Skewness: Kurtosis:**

Speed=> -0.117510 Speed=> -0.508994

Distance=> 0.806895 Distance=> 0.405053

From the skewness of speed we can say that the data of speed is symmetric and from the distance we can say that the data is positively skewed.

**SP and Weight(WT)**

**Use Q9\_b.csv**

ANS:

**Skewness: Kurtosis:**

Speed=> 1.611450 Speed=> 2.977329

Distance=> -0.614753 Distance=> 0.950291

From the skewness of SP we can say that the data of SP is positively skewed and from the skewness of weight we can say that the data is negatively skewed.

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



ANS: This is right skewed Histogram and its skewness is seen to be positive.



ANS: Medican 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.

**C.L. Lower Limit Upper Limit**

94% 198.738325 201.261675

98% 198.622303 201.377697

96% 198.622303 201.560562

**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.

ANS:

Mean => 41.0 Variance => 24.1111

Median => 40.5 Standard deviation => 4.9103

1. What can we say about the student marks?

ANS:

Student scores 41 marks as an average.

Mean is greater than Median

It is right skewed and contains no outliers.

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

ANS: It is perfectly symmetrical. Hence, skewness value will be zero.

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

ANS: It is right skewed. Hence, skewness value will be Positive.

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

ANS: It is left skewed. Hence, skewness value will be Negative.

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

ANS: It indicates that a distribution is peaked and possess thick tails

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

ANS: It indicates that a distribution is flat and has thin tails

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



a). What can we say about the distribution of the data?

Answer :

We can say that there is a big difference between upper quartile and upper extreme compared to the lower quartile and lower extreme.

b). What is nature of skewness of the data?

Answer :

The nature of skewness is negatively skewed. Or Left Skewed

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

Answer :

IQR = Q3-Q1

IQR = 18-10

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

ANS: Medium is same for both graph, there is difference between upper limit and lower limit (IQR1-IQR2) and there is no outliers in both the boxplot.

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)

Ans. 0.347594

* 1. P(MPG<40)

Ans. 0.729350

* 1. P (20<MPG<50)

Ans. 0.898869

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans. From above plot and values we can say that data is fairly symmetrical, i.e fairly normally distributed.

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

Dataset: wc-at.csv

Ans. From the above plots we can say that, the data is normally distributed for ‘Waist’ and data is positively skewed for ‘AT’.

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

Ans.

Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

Ans.

T-Score of 95% confidence interval = 2.0638985616280205

T-Score of 96% confidence interval = 2.1715446760080677

T-Score of 99% confidence interval = 2.796939504772804

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.

t = -0.4714045207910317

Probability: 0.32167411684460556