|  |  |
| --- | --- |
| 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 & Discrete type data |
| High School Class Ranking | Nominal & Discrete type data |
| Celsius Temperature | Interval & Continuous type data |
| Weight | Ratio & Continuous type data |
| Hair Color | Ratio & Discrete type data |
| Socioeconomic Status | Interval & Continuous type data |
| Fahrenheit Temperature | Ratio & Continuous type data |
| Height | Ratio & Continuous type data |
| Type of living accommodation | Ordinal & Discrete type data |
| Level of Agreement | Interval & Discrete type data |
| IQ(Intelligence Scale) | Interval & Discrete type data |
| Sales Figures | Interval & Discrete type data |
| Blood Group | Ratio & Discrete type data |
| Time Of Day | Interval & Continuous type data |
| Time on a Clock with Hands | Interval & Continuous type data |
| Number of Children | Interval & Discrete type data |
| Religious Preference | Ratio & Discrete type data |
| Barometer Pressure | Interval & Discrete type data |
| SAT Scores | Ratio & Discrete type data |
| Years of Education | Nominal & Discrete type data |

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

When three coins are tossed -:

The out-comes = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

For two heads and one tail-:

The out-comes= {HHT, HTH, TTH}

P (Two heads and One Tail) = P (Number outcomes for two heads and one tai) /

P (total outcomes of the event)

=3/8

**P (Two heads and One Tail) = 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-:

P (D1) = {1,2,3,4,5,6} 🡪 n(D1)=6

P (D2) = {1,2,3,4,5,6} 🡪n(D2)=6

N(D1+D2) = 36

1. For ,  
   **P(sum is equal to 1) = 0** ,because the lowest P(D1+D2)=(1+1)=2
2. For,

(sum is less than or equal to 4) ={(1+1),(1+2),(1+3),(2+1),(2+2),(3+1)}

n (sum is less than or equal to 4)= 6

Probability= n (sum is less than or equal to 4)/ n(total outcomes)

=6/36

**P (sum is less than or equal to 4) =1/6**

c)For,

Probability that sum is divisible by 2 and 3,

sum is divisible by 2 and 3= { (1+5),

(2+4),

(3+3),

(4+2),

(5+1),

(6+6)}

n (sum is divisible by 2 and 3) = 6

P (sum is divisible by 2 and 3) = n (sum is divisible by 2 and 3)/ n(s)

=6/36

**P (sum is divisible by 2 and 3= 1/6**

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 = 2red + 3green + 2blue = 7  
Let A be the sample space.  
Then, n(A) = Number of ways of drawing 2 balls out of 7  
 =7C2​  
 =(7\*6)​/(2\*1)  
 =42/2

N(A) =21

Let B = Event of drawing 2 balls, none of which is blue.  
 Then, n(B)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
 =5C2​  
 =(5\*4)​/(2\*1)  
 =20/2  
 =10

P(B)=n(B)/N(A)

=10/21

**Probability that none of the balls drawn is blue=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-:

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

= 3.09

**Expected number of candies for a randomly selected child = 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**

Answer-:

**Please check attached notebook**

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?

Answer-:

There are total 9 patients.

Probability for selecting each patient is = 1/9 🡪 P(X)=1/9

For random selection we can’t pick or assign any value from given data.

For,

Expected Value = total of (probability of selecting each patient \* each given Value)

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

  = (1\*1308)/9

= 145.33

**Expected Value of the Weight of that patient = 145.33**

**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**

Answer-:

**Please check attached notebook**

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





Answer-:

**Histogram-:**

1) The data is distributed with range 0 to 50.

In range 50-100, more data is lies.

2)More data skewed at initial data. It is right skewed type data.

3)Very less data points between 300-400.

4)Less influence of chick weight on frequency as Weight increases.

**Box-plot-:**

1) Outliers at the end of plot.

2)Mean line of the data is near towards 1 IQR.

3)This is positively skewed data.

**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?

Answer-:

We are given the standard deviation for the sample, which is why the

**t-distribution** is used to solve this question.

The information given is:

* Sample mean of xbar=200
* Sample standard deviation of s=30.
* Sample size of n=2000.

The **interval** is:

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%5Cpm%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D

* In which **t** is the critical value for the two-tailed confidence interval.

For ‘n’ number sample with n-1 degree of freedom.

For ‘200’ number sample with n-1=200-1=199 degree of freedom.

Considering a 94% confidence level, the critical value is t = 1.8916,

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.73

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.27

The **94%** confidence interval is **(198.73, 201.27)**

Considering a 96% confidence level,the critical value is t = 2.0673,

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.61

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.39

The **96%** confidence interval is **(198.61, 201.39)**

Considering a 98% confidence level, the critical value is t = 2.3452,

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.43

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.57

The **98%** confidence interval is **(198.43, 201.57)**

**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?

Answer-:

**Please check attached notebook**

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

Answer-:

Skewness gives the idea about data distribution over symmetrical data.

Asymmetric distribution of data can easily identify by skewness curve.

For equal mean, median data the skewness is zero.

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

Answer-:

 If the mean is greater than the median, then the distribution is positively skewed.

Positively skew means right tail data.

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

Answer-:

 If the median is greater than the mean, then the distribution is negatively skewed.

Negatively skew means left tail data.

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

Answer-:

Positive values of kurtosis indicate that distribution is above the normal distribution curve. More data located at the tail.

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

Answer-:

Positive values of kurtosis indicate that distribution is below the normal distribution curve. Distribution is more flat than normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Answer-:

1)From the distribution we can infer that data is left tail.

2) From the box-plot it is observed that data is negatively skewed.

3) (Quantile3- Quantile2)<( Quantile2- Quantile1)

Q19) Comment on the below Boxplot visualizations?



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

Answer-:

1)Both the boxplot having the same mean, but data distribution looks different.

2)From plot it looks like normally distributed data with different data size.

3)Box-plot 1 is comparatively small than box-plot 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)

Answer-:

**Please check attached notebook**

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

Answer-:

**Please check attached notebook**

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

Answer-:

**Please check attached notebook**

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

Answer-:

**Please check attached notebook**

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

Answer-:

**Please check attached notebook**