|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Nominal Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Nominal Data |

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

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

Ans:

Total Events: 8

No. of desired events: 3

P=3/8=0.375

37.5% is possible for getting two heads and one tail

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:

Total n(s) =36

1. Equal to 1

n(x)=0

p(a) =n(x)/n(s)=0/36= 0

1. Less than or equal to 4

n(y)=(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)=6

p(b)=n(y)/n(s)=6/36=0.167

c) Sum is divisible by 2and 3

n(z)=(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

n(z)=6

p(c)=6/36=0.167

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:  
 Red balls= 2

Green balls=3

Blue balls= 2

Total balls = 2+3+2 = 7

Probability:

P= (5\*4/2\*1) / (7\*6/2\*1) =10/21

P= 0.476% Probability of getting no blue balls when randomly picked.

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

= 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

|  |
| --- |
| Expected number of candies=3.14 |

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:

|  |  |  |  |
| --- | --- | --- | --- |
| Column1 | Points | Score | weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.9 | 3.4 | 18.9 |
| Variance | 0.2858814 | 0.957379 | 3.1931661 |
| STD | 0.534679 | 0.978457 | 1.786943 |
| Range | 2.17 | 3.911 | 8.4 |
| Minimun | 2.76 | 1.513 | 14.5 |
| Maximun | 4.93 | 5.424 | 22.9 |

**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:

Weights=108+110+ 123+ 134+ 135+ 145+ 167+ 187+199/9

|  |
| --- |
| Weights=145.33 |

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Speed | -0.1175 | -0.5089 |
| distance | 0.8068 | 0.4050 |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| Sp | 1.6114 | 2.9775 |
| weight | -0.6147 | 0.9502 |

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



Ans:

\***Histogram:**

-The histogram diagram is a right tail

-It is positive Skewness

-Most number of chick weight range is 50-10

\* **Boxplot:**

**-**: The Box plot diagram have many outliers

**-** The Box plot data is positive Skewness

**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:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **94%** | **98%** | **96%** |
| **Upper** | **201.04** | **201.38** | **201.17** |
| **Lower** | **198.96** | **198.62** | **198.83** |

**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 x: 41

Median: 40.5

Standard Deviation: 5.0526

variance:25.5294

1. What can we say about the student marks?

Ans:

Most of student marks got in below 41

Majority of student score range for 35-45

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

Ans: : Zero skewness

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

Ans: positive skewness

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

Ans: negative skewness

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

Ans: distribution data is a center part

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

Ans: distribution data is flatter or wider in a center part of negative Kurtosis

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



What can we say about the distribution of the data?

Ans:

\* The given data lines on right ways

\* It is not symmetrical in nature

What is nature of skewness of the data?

Ans:

\*It is negaitive skewness

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

Ans:

lQR=18-10=8 data

Q19) Comment on the below Boxplot visualizations?



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

Ans:

1. Two Box Plots are Normal Distribution

2.Two Box Plots are same middle value (262)

3. Second BoxPlot is large range for compared to frist Box plot

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)0.6524

b)7293

c)1.2430

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

\* MPG is not Normal Distribution

Reason:

-Skewness is negative

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

Dataset: wc-at.csv

Ans:

\*wc is normal Distribution

Reason:

-Skewness is positive

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

Ans:

|  |  |  |
| --- | --- | --- |
| Confidence interval | Alpha | Z score |
| 94% | 0.03 | 1.880 |
| 90% | 0.05 | 1.644 |
| 60% | 0.02 | 0.841 |

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

Ans:

Sample size n=25

|  |  |
| --- | --- |
| Confidence interval | t\_scores |
| 95% | ±2.060 |
| 96% | ±2.485 |
| 99% | ±2.787 |

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:

t\_score=-0.4714

df= 17

P(t\_score)=0.3216