|  |  |
| --- | --- |
| 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 |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Ans: Total possible Event : 8 , No.of desired events: 3

Probability(P)= 3/8 =0.375

37.5% probability to get two heads 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 2 and 3

Ans:

1. Probability of Equal to 1= 0
2. Less than or equal to 4= 6/36 = 0.167=16.7%
3. Sum is divisible by 2 and 3 = 5/36=0.139=13.9%

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 ball=2+3+2=7

1. Probability of drawing the first non-blue ball=5/7
2. Probability of drawing the second non-blue ball=4/6

Two balls are drawn at random = 5/7 \* 4/6 = 0.476

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: 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**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | **3.596563** | **3.217250** | **17.848750** |
| **Median** | **3.695000** | **3.325000** | **17.710000** |
| **Mode** | **3.07** | **3.44** | **17.02** |
| **Std** | **0.534679** | **0.978457** | **1.786943** |
| **Variance** | **0.285881** | **0.957379** | **3.193166** |
| **Range** | **2.17** | **3.911** | **8.4** |
| **Max** | **4.930000** | **5.424000** | **22.900000** |
| **Min** | **2.760000** | **1.513000** | **14.500000** |

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: 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | -0.11 | -0.50 |
| distance | 0.80 | 0.40 |

• From the above Table we can say that Speed is negatively Skewed and Distance is positively Skewed.

• Due to both Speed and Distance has lighter tails and fewer outliers than a normal distribution. Such distributions are called platykurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans**

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtosis |
| speed | 1.611450 | 2.977329 |
| Weight | -0.614753 | 0.950291 |

• From the above Table we can say that Speed is positively Skewed and weight is negatively Skewed.

• Due to both Speed and Weight has lighter tails and fewer outliers than a normal distribution. Such distributions are called platykurtic.

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



ANS:

• The data is positive Skewed

• The above histogram has right tail

• Most number of chick weigh from the range of 50-100 and followed by 100-200



Ans:

Data of Box plot shows us that it has many outliers on upper whisker and the data is positively Skewed

**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** | **201.26** | **201.56** | **201.37** |
| **Lower** | **198.73** | **198.43** | **198.62** |

**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?

Ans:

1) Mean: 41

Median: 40.5

Mode:41

Standard Deviation: 5.05

Variance:25.53

2) Most of the marks of student got is below 41

Majority of score lies between to the range of 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: Positively skewed

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

Ans: Negatively skewed

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

Ans: the distribution of data is peaked in center part

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

Ans: the distribution of data is flatter or wider in center part

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



What can we say about the distribution of the data?

Ans: The given data lies on right side, so it not symmetrical in nature and the distribution is left hand tail

What is nature of skewness of the data?

Ans: Negatively skewed

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

Ans: IQR=UQ-LQ = 18-10 =8((approximately)

Q19) Comment on the below Boxplot visualizations?



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

* Both the Box Plots are Normally Distributed
* Both have Same Median around 261(app)
* Comparatively Second BoxPlot has high variability compared to 1st Box plot

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)
  3. P (20<MPG<50)

Ans:

P(MPG>38) = 40.74%

P(MPG<40) =75.30%

P (20<MPG<50) =0%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: MPG of cars data followed Normal Distribution

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

Dataset: wc-at.csv

Ans: Both AT and Waist doesn’t follow Normal Distribution

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

|  |  |  |
| --- | --- | --- |
|  | Alpha a | Z score |
| 60% | 0.20 | 0.253 |
| 90% | 0.05 | 1.281 |
| 94% | 0.03 | 1.554 |

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

|  |  |
| --- | --- |
| 95% | ±1.71 |
| 96% | ±1.82 |
| 99% | ±2.49 |

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-score: -0.4714

Degree of freedom(df): 17

P(t)=0.3216 = 32.16%