|  |  |
| --- | --- |
| 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 | 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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Nominal |

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

Ans : 0.37

Q4) Two Dice are rolled, find the probability that sum is

* Equal to 1 = 0
* Less than or equal to 4 = 1/6
* Sum is divisible by 2 and 3 = 1/6

Explation: Total possible outcomes is 6^2 = 36

Probability that the sum is equal to 1 is 0 as this is not possible because every time the sum will be greater than 1.

Probability that the sum is less than or equal to 4 is 1/6 --> (1,1) ,(1,2), (1,3), (2,1), (2,2), (3,1) = no of interseted events/total no of events = 6/36 = 1/6.

Probability that the sum is divisible by 2 and 3 is 1/6 --> (1,5), (2,4), (3,3), (4,2), (5,1), (6,6) = 6/36 = 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?

Ans : 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

Ans : Expected no 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**

**Ans:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **S.D** | **Range** |
| **Points** | 3.596563 | 3.695000 | 3.92 | 0.285881 | 0.534679 | 2.17 |
| **Score** | 3.217250 | 3.325000 | 3.44 | 0.957379 | 0.978457 | 3.911 |
| **Weigh** | 17.848750 | 17.71000 | 17.02 | 3.193166 | 1.786943 | 8.4 |

**There are outliers in weight column and hence the mean of the weight is influenced.**

Q8) Calculate Expected Value for the problem below

* 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 | -0.117510 | -0.508994 |
| Distance | 0.80689 | 0.405053 |

Data is negatively skewed for speed and it is left skewed by 0.117

Data is positively skewed for distance and it is right skewed by 0.8

**SP and Weight(WT)**

**Use Q9\_b.csv**

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

Data is positively skewd for SP and it is right skewed by 1.611

Data is negatively skewed for Weight and it is left skewed by 0.614

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





Histogram: Data is positively skewed or we can say that data is right skewed

Boxplot: There are outliers present in the upper extreme of the data.In the inter quartile region 50% of data is present. . In the region between upper extreme to upper quartile which is called as upper whisker 25% of data is present. And 25% data is present in lower whisker.

**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**

* Find mean, median, variance, standard deviation.

Mean=41

Median=40.5

Var=25.52

S.D=5.05

* What can we say about the student marks?

Ans: The marks are multimodal in nature.

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

Ans: If the mean and median of the data are equal then we say the data has zero skewness.

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

Ans: When mean is greater than median we say the data is positively skewed.

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

Ans: When median is greater than mean we say the data is negatively skewed.

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

Ans:In positive kurtosis we will get a sharp peak and the loss of the data will be very less which can be negotiable.

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

Ans: In negative kurtosis we will get a flat distribution which will indicate that there is maximum loss.

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



What can we say about the distribution of the data?

Ans: 25% of the data is present in lower whisker (Between the range 1-10, lower whisker is present), 50% of data is present in interquartile range (Between the range 10-18, interquartile is present), 25% data is present in upper whisker(Between the range 18.1-20,upper whisker is present).

What is nature of skewness of the data?

Ans: The data is negatively skewed.

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

Ans: IQR is calculated as Q3-Q1 .

IQR = 18.1 - 10 = 8.1

Q19) Comment on the below Boxplot visualizations?



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

Boxplot1: Range of the data is 230-290, The median is 265 same as boxplot2. The upper whisker lies in the range 276-290, The lower whisker lies in the range 248-251. The interquartile is between 252-278. IQR of boxplot1 is less than boxplot2 which is approximately 39.

Boxplot2: Range of the data is 175-350. The median of the boxplot 2 is same as boxplot1 that is 265. Upper whisker is in the range 315- 350. Lower whisker lies between 175-225. The IQR is between 225-315.

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

* P(MPG>38)
* P(MPG<40)
* P (20<MPG<50)

--> Here X is 38, mean is 34.42, std. dev is 9.13

Probability is 0.347

--> Here X is 40, mean is 34.42, std. dev is 9.13

Probability is 0.72

--> Here we will divide the range such that MPG<50 and MPG>20, for both the range mean is 34.42, std. dev is 9.13

Probability is 0.89

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: The data is not normally distributed as the mean, median and mode are not equal. Also the curve is not symmetric.

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

Dataset: wc-at.csv

Ans: The data is not normally distributed because the curve is not symmetric.

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

Ans: Z score of 90% CI is 1.64

Z score of 94% CI is 1.88

Z score of 60% CI is 0.84

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% CI is 2.06

T score of 96% CI is 2.17

T score of 99% CI is 2.79

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: Population mean = 270

Sample size(N) = 18

Sample mean = 260

Std. dev of sample = 90

As we can see, std.dev of population is not given so we will perform t test here:

T = sample mean - population mean/(std.dev/sqrt(n))

T= 270-260/(90/sqrt(18))

T = 0.471

As sigma that is std.dev of population is not given so we will calculate CI with the help of T distribution with 17 df as n = 18.

Also alpha value is not given so we will consider it as 1- alpha = 95% so alpha value will be 0.95 .We will calculate t vale(crictical value) for 0.97

t0.97,17 = 2.015.

Now we have to calculate the probability that 18 randomly selected bulbs would have an average life of no more than 260 days.

We will calculate this with the help of t test value which is -0.471 and degree of freedom.

So, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days is **0.3218,** From this it can be concluded that bulbs can last for more than 260 days.