|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Continuous |
| 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 | Continuous |
| 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 | Nominal |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Nominal |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

(ans) Three coins are tossed:

Total no.of observations= 8

No.of observations with 2 heads and 1 tail= 3

P(2 heads,1 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 2and 3

(ans) Two dice are rolled:

Total no.of observations=36

1. P(sum=1)= 0/36
2. P(sum<=4)=6/36
3. P(sum divisible by 2 & 3)= 6/36

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) no.of ways in which 2 random balls are picked from the bag= 7C2=21

No.of ways that the ball drawn is red or green= 5C2= 10

P(no blue balls)= 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 number of candies for a randomly selected child= ∑(x\*P(x))= 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:

Mean: 3.597

Median: 3.695

Mode: 3.92

Standard deviation: 0.5346787

Variance: 0.2858814

Range: Min. :2.760

Max. :4.930

Score:

Mean :3.217

Median :3.325

Mode: 3.44

Standard deviation: 0.9784574

Variance: 0.957379

Range: Min. :1.513

Max. :5.424

Weigh:

Mean :17.85

Median :17.71

Mode: 17.02

Standard deviation: 1.786943

Variance: 3.193166

Range: Min. :14.50

Max. :22.90

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) P(x)=one patient is chosen at random=1/9=0.111

Expected value= ∑P(x)\*x = 145.188

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

**Cars speed and distance**

**Use Q9\_a.csv**

**(ans)** Speed: Skewness: -0.11751 (left skewed distribution)

Kurtosis: -0.50899

Distance: Skewness: 0.806895 (right skewed distribution)

Kurtosis: 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**(ans)** SP: Skewness: 1.61145(Right skewed distribution)

Kurtosis: 2.977329

WT: Skewness: -0.61475(left skewed distribution)

Kurtosis: 0.950291

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



The Histogram mentioned above is skewed to the right.



The boxplot mentioned above is skewed to the left. The outliers marked shows that’s the data plotted is not correctly analysed.

**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**

1. Find mean,median,variance,standard deviation.

Mean: 41

Median: 40.5

Standard deviation: 5.052664

Variance: 25.52941

1. What can we say about the student marks?

The distribution of the marks of the students, when plotted on a graph shows positive skewness.

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

(ans)When the mean, median of a data are equal, the skewness follows standard normal distribution.

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

(ans)The data is right-skewed when mean>median.

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

(ans)The data is left-skewed when median>mean.

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

(ans)Positive kurtosis value of a data indicates that the data has heavier tails than a normally distributed data.

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

(ans) Negative kurtosis value of a data indicates that the data has lighter tails than a normally distributed data.

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



What can we say about the distribution of the data?

(ans) the median lies near towards the upper whisker, and maximum of the data lies in the range of 10-18, and the data contains no outliers thus indicating no bad unevaluated data.

What is nature of skewness of the data?

(ans) the data is skewed to the left i.e. negatively skewed.

What will be the IQR of the data (approximately)?   
(ans) Q3~18

Q1=10  
IQR of the data= Q3-Q1= 18-10 = 8

Q19) Comment on the below Boxplot visualizations?



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

(ans) 1. The following boxplots show that there is no difference in distribution between the two groups.

2. Boxplot-1 has less dispersed data, while Boxplot-2 has more dispersed data.

3. both of the boxplots contain no outliers.

4. both of the boxplots follow a normal distribution.

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)

(ans) 0.3475941

* 1. P(MPG<40)

(ans)0.7293497

* 1. P (20<MPG<50)

(ans)0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

(ans) Shapiro-Wilk normality test

data: MPG

W = 0.97797, p-value = 0.1764. Thus, MPG follows 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) Shapiro-Wilk normality test:

Waist: p-value = 0.00117

AT: p-value = 0.000654

As both of the p-values are not greater than 0.05, both of the data don’t follow a normal distribution.

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

(ans) at 90% confidence interval, the Z-Score is: 1.645

94% confidence interval, the Z-Score is: 1.88

60% confidence interval, the Z-Score 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) sample size=25

DOF= n-1=25-1= 24

t-value for confidence level:

95%: 1.710882

96%: 1.828051

99%: 2.492159

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) Sample size= 18

Degree of freedom= n-1= 17

Assuming α= 0.05, respective t-score= 1.7396

Thus, P(average life of selected bulbs<260)= 0.050001