|  |  |
| --- | --- |
| 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 | Discrete |
| Dog's weight | Discrete |
| 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 | Ordinary |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinary |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinary |
| Level of Agreement | Ordinary |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinary |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | ratio |
| SAT Scores | interval |
| Years of Education | ratio |

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

Ans:- 3/8 or 0.375

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:- a) 0

b)6/36

c)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: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: 3.085

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 for points=3.596563 Median for points= 3.695**

**Mean for score=3.21725 Median for score= 3.325**

**Mean for weigh=17.84875 Median for weigh= 17.71**

**Mode for points=3.073.92 Variance for points=0.2858814**

**Mode for score= 3.44 Variance for score= 0.957379**

**Mode for weigh=17.0218.90 Variance for weigh=3.193166**

**Standard Deviation for points=0.5346787 Range for points=2.764.93**

**Standard Deviation for score = 0.9784574 Range for score=1.5135.424**

**Standard Deviation for weigh = 1.786943 Range for weigh= 14.522.9**

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: 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:- Skewness for Speed = -0.1105533 & Distance=0.7591268**

**Kurtosis for speed= 2.422853 & Distance= 3.248019**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans: Skewness for SP=1.552258 & WT= -0.6033099**

**Kurtosis for SP=5.723521 & WT= 3.819466**

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



Ans:- : The histograms peak has right skew and tail is on left. Mean > Median. We have outliers on the higher side.



Ans: The boxplot has outliers on the maximum side.

**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: Ans: conf\_94 =stats.t.interval(alpha = 0.94, df=1999, loc=200, scale=30/np.sqrt(2000)) print(np.round(conf\_94,0)) print(conf\_94) For 94% confidence interval Range is [ 198.73 – 201.26] For 98% confidence interval range is [198.43 – 201.56] For 96% confidence interval range is [198.62 – 201.37]

**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=41

Median=40.5

Variance=25.52941

Standard deviation=5.05664

1. What can we say about the student marks?

Ans: Student is performing reasonably well but there is room for

improvement.

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

Ans: :- If the mean and median of the data is equal, the distribution is symmetric, and the skewness is 0.

Therefore, the nature of skewness is zero (or) 0.

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

Ans: Positively skewed (or) the distribution is skewed to the right side (or) right skewed

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

Ans: Negatively Skewed (or) the distribution is skewed to the left side (or) Left Skewed.

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

Ans: A positive kurtosis value indicates that the distribution has a sharper peak and heavier tails than a normal distribution.

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

Ans: A negative kurtosis value indicates that the distribution has a flatter peak and lighter tails than a normal distribution

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



What can we say about the distribution of the data?

Ans: The above boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

Ans: The above boxplot is not normally distributed the median is towards the higher value.

What will be the IQR of the data (approximately)?   
Ans: The inter Quantile Range = Q3 the Upper quartile – Q1 Lower Quartile = 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: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the maximum whisker range.

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:

a) 0.3475

b) 0.729

c) 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:  Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv Ans: a.) MPG of cars 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:  Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv Ans: a.) MPG of cars follows normal distribution

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

Ans:

For 90% Confidence Interval Z-Score is calculated as:

Z = 1.645

For 94% Confidence Interval Z-Score is calculated as:

Z = 1.555

For 60% Confidence Interval Z-Score is calculated as:

Z = 0.253

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

Ans: For a Sample size of 25 and a 95% confidence interval, the t-score is:

t = 2.064

For a Sample size of 25 and a 96% confidence interval, the t-score is:

t = 2.171

For a Sample size of 25 and a 99% confidence interval, the t-score is:

t = 2.797

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: 0.32=32%