|  |  |
| --- | --- |
| 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 | Continuous |
| 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, and 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| 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: 3/8

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

1. Equal to 1 = 0
2. Less than or equal to 4 = 1/6
3. Sum is divisible by 2 and 3 = 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: Total number of outcomes of picking 2 random balls out of 2 Red, 3Green and 2 Blue is 7C2 = 21

The total number of outcomes of getting both non-Blue balls is 5C2 = 10

The Probability of none of the balls drawn blue is = 5C2/7C2 = 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of the 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 – the probability of having 1 candy = 0.015.

Child B – the probability of having 4 candies = 0.20

Ans: The expected number of candies for a randomly selected child will be

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

= 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** | 3.59 | 3.21 | 17.84 |
| **Median** | 3.695 | 3.325 | 17.710 |
| **Mode** | 3.07 | 3.44 | 17.02 |
| **Variance** | 0.28 | 0.95 | 3.193 |
| **Std. Deviation** | 0.53 | 0.97 | 1.78 |
| **Range** | 2.76 – 4.93 | 1.513- 5.424 | 14.5 – 22.9 |

Q8) Calculate the 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?

Expected Value = (1/9)(108) + (1/9)110 + (1/9)123 + (1/9)134 + (1/9)135 + (1/9)145 + (1/9(167) + (1/9)187 + (1/9)199

= (1/9) ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)

= (1/9) ( 1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars’ speed and distance**

**Use Q9\_a.csv**

**Ans** The Skewness of Cars’

Speed = -0.117

Distance=0.806

The Kurtosis of Cars’

Speed = -0.5

Distance= 0.4

The Skewness of Cars’ speed is Negative which means it is left-skewed and Distance is positive which means right skewed

The Kurtosis of Cars’ speed is Negative and Distance is positive

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans** The Skewness of Cars’

SP = 1.611

Weight=-0.614

The Kurtosis of Cars’

SP = 2.977

Weight= 0.95

The Skewness of the Cars’ weight is Negative which means it is left-skewed and SP is positive which means right skewed

The Kurtosis of Cars’ SP and Weight are positive

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



Ans: The histogram’s peak has right skew and the tail is on right. Mean > Median. We have outliers on the higher side.

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%, and 96% confidence intervals.

Ans: For a 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 on tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

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

Ans: Mean =41, Median =40.5, Variance =25.52 and Standard Deviation =5.05

1. What can we say about the student marks?

Ans: we don’t have outliers and the data is slightly skewed towards the right because the mean is greater than the median.

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

Ans: No skewness is present we have a perfectly symmetrical distribution

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

Ans: Skewness and tail are towards Right

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

Ans: Skewness and tail are towards left

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

Ans: Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

Ans: Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

Ans: The above Boxplot has not normally distributed the median towards the higher value

What is the nature of the skewness of the data?

Ans: The data is skewed towards the left. The whisker range of minimum value is greater than maximum

What will be the IQR of the data (approximately)?   
Ans: The Inter Quantile Range = Q3 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 to 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 minimum nor 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)= 0.348
  2. P(MPG<40) = 0.729

c. P (20<MPG<50) =0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

 MPG of cars follows the normal distribution

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

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) and Waist do not follow Normal Distribution

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

Ans: Z score for 60% Conifidence Intervla = -1.6449

Z score for 60% Conifidence Intervla = -1.8808

Z score for 60% Conifidence Intervla = -0.8416

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

T score for 95% Confidence Interval = -2.0639

 T score for 94% Confidence Inteval = -1.974

T score for 99% Confidence Interval = -2.7969

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: import numpy as np Import scipy as stats t\_score = (x - pop mean) / (sample standard deviation / square root of sample size) (260-270)/90/np.sqrt(18)) t\_score = -0.471 stats.t.cdf(t\_score, df = 17) 0.32 = 32%