|  |  |
| --- | --- |
| 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 | Ratio |
| Weight | Ratio |
| Hair Color | Ordinal |
| Socioeconomic Status | Ratio |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Interval |
| 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**

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. **0**
2. **1/6**
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: 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: 3.09**

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 points = 3.596563**

**Median points = 3.695**

**Mode points = Bimodal 3.07 3.92**

**Variance points = 0.2858814**

**SD points = 0.5346787**

**Range points = 2.76 4.93**

**Mean score = 3.21725**

**Median score = 3.325**

**Mode score = Unimodal 3.44**

**Variance score = 0.957379**

**SD score = 0.9784574**

**Range score = 1.513 5.424**

**Mean weigh = 17.84875**

**Median weigh = 17.71**

**Mode weigh = Bimodal 17.02 18.90**

**Variance weigh = 3.193166**

**SD weigh = 1.786943**

**Range weigh = 14.5 22.9**

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: 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :**

**Skewness speed = -0.1139548**

**Kurtosis speed = 2.422853**

**Skewness dist = 0.7824835**

**Kurtosis dist = 3.248019**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

**Skewness SP = 1.581454**

**Kurtosis SP = 5.723521**

**Skewness Weight = -0.6033099**

**Kurtosis Weight = 3.819466**

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



**Ans: The histogram is positively skewed and the value of skewness is greater than 0. In this the mean is greater followed by median and then mode.**



**Ans: In this boxplot rthe outliers present only after upper quartile range. The lower quartile range is pretty narrow so more data prsent in this area is in small range. The center line is called as median. The interquartile range contains 50 % of data. The upper quartile range is pretty wide so the 25% of that data is pretty scattered.**

**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:**

**For 94% = (198.73832, 201.26167)**

**For 98% = (198.43943, 201.56056)**

**For 96% = (198.62230, 201.37769)**

**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:**

**Mean: 41**

**Median: 40.5**

**Variance: 25.52941**

**SD: 5.052**

**25% students lies between marks 34 to 38**

**50% students lies between marks 39 to 41**

**Another 25 % students lies between marks 42 to 45**

**49 & 56 are outliers and median is about at 40.5 marks**

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: Distribution is peaked and possesses thick tails**

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

**Ans: Distribution is flat and possesses thin tails**

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



What can we say about the distribution of the data?

**Ans: The data is distributed in De-assigned format**

What is nature of skewness of the data?

**Ans: Left side skewed**

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

**Ans: 8**

Q19) Comment on the below Boxplot visualizations?



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

**Ans: In boxplot 1, the lower quartile area, which contains 25% values, lies between 240 to 250 so its pretty congested. While in boxplot 2, the lower quartile range lies between 200 to 225 so its pretty scattered. In both the boxplots, the median lies at same position i.e at 237.5. But the interquartile range in boxplot 1 lies between 250 to 275 so its pretty congested but in boxplot 2 it lies between 225 to 312.5. and same for upper quartile range. Both the plots does not contain any outliers. So the data in boxplot 1 is lies in small area but in boxplot 2 it lies in vast area.**

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)

**Ans: 0.34759**

* 1. P(MPG<40)

**Ans: 0.72934**

* 1. P (20<MPG<50)

**Ans: 1.24309**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans: It follows normal distribution and it is negatively skewed**

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

Dataset: wc-at.csv

**Ans:**

**Adipose Tissues (AT) follows normal distribution and it is positively skewed**

**Waist Circumference (Waist) does not follow normal distribution.**

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

**Ans:**

**For 90% = 1.64485**

**For 94% = 1.88079**

**For 60% = 0.84162**

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

**Ans:**

**For 95% = 2.06389**

**For 96% = 2.17154**

**For 99% = 2.79693**

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: 0.32167**