|  |  |
| --- | --- |
| 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 | Continuous |
| Number of tickets in Indian railways | Continuous |
| Number of times married | Continuous |
| 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 | Interval |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Interval |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Interval |
| Blood Group | Ordinal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

**Ans:- two head 3/8,one tail 1/8**

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:- a)probability that sum is equal to 1 is 0**

**b) probability that sum is Less than or equal to 4 is 6/36=1/6**

**c) probability that sum is divisible by 2 and 3 is 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:- probability that none of the balls drawn is blue is 5/7\*4/7=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:-** **The expected number of candies for a randomly selected child is 3.135.**

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 Q 7.csv file

**Ans:-**

**Points Score weigh**

**Mean 3.5965 3.2172 17.847**

**Mode 3.92 3.44 17.02**

**Median 3.695 3.325 17.710**

**Variance 0.2858 0.9573 3.1931**

**Standard**

**Deviation 0.5346 0.9784 1.7869**

**Range 2.17 3.911 8.39**

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=1/9(108+110+123+134+135+145+167+187+199)=145.33**

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

Cars speed and distance

Use Q9\_a.csv

**Ans:-**

**Speed Distance**

**Skewness -0.1175 0.806**

**Kurtosis -0.5089 0.4050**

SP and Weight(WT)

Use Q9\_b.csv

**Ans:-**

**Speed Weight**

**Skewness 1.6114 -0.3147**

**Kurtosis 2.9773 0.9502**

Q10) Draw inferences about the following boxplot & histogram



**Ans:-** **Ans:-**

**Histogram**

**Skewed Left (positively): If the left tail is longer, it indicates that there are some lower values in the data.**

**From the above boxplot there are some outliers**

**Outliers:**

**Isolated Bars: Bars that stand alone may indicate outliers or data points significantly different from the rest.**

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%=****between 197.741-202.258**

**98%= between 199.2479-200.7520**

**96%= between 198.4953-201.5046**

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 =41.0**

**Median=41.5**

**Variance=24.11**

**Standard deviation=4.9103**

1. What can we say about the student marks?

**Ans:-the students marks are improving by each interval gradually**

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

**Ans:-When the data is symmetrically distributed and there is no skewness**

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

**Ans:-It indicates a right skewed distribution**

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

**Ans:-It indicates a left skewed distribution**

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

**Ans:-****It indicates a distribution with heavier tails and a sharper peak then normal distribution. The distribution is leptokurtic**

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

**Ans:-** **It indicates a distribution with lighter tails and a flatter peak then normal distribution. The distribution is flatyurtic**

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



What can we say about the distribution of the data?

**Ans:-Positive distribution**

What is nature of skewness of the data?

**Ans:-Positive skewness**

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

**Ans:- IQR=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:-**

**The distribution of data for boxplot1 with respect to boxplot2**

**The data in boxplot1 is between 250 to 275 wbs and the data in boxplot2 is between 225 to 312.5 wbs**

**There are no outliers in both boxplot1 and boxplot2**

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)

c. P (20<MPG<50)

**Ans:-**

**A. P(MPG>38) = 0.407**

**B. P(MPG<40) = 0.753**

**C. P(20<MPG>50) = 0.938**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:- NO**

**It not follow the 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:-NO**

**Its not follows the normal distribution**

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

**Ans:- Z score of**

**90% = 1.64**

**94% = 1.88**

**60% = 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% = 2.06**

**96% = 2.171**

**99% = 2.796**

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:-**

**0.3216**