|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Numeric Discrete Data Type |
| Results of rolling a dice | Numeric Discrete Data Type |
| Weight of a person | Numeric Continuous Data Type |
| Weight of Gold | Numeric Continuous Data Type |
| Distance between two places | Numeric Continuous Data Type |
| Length of a leaf | Numeric Continuous Data Type |
| Dog's weight | Numeric Continuous Data Type |
| Blue Color | Categorical Data Type |
| Number of kids | Numeric Discrete Data Type |
| Number of tickets in Indian railways | Numeric Discrete Data Type |
| Number of times married | Numeric Discrete Data Type |
| Gender (Male or Female) | Categorical Data Type |

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 | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal  (disagree, less agree, highly agree) |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal  (A,B,AB,O) |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval  (there is no proper definition of 0) |
| SAT Scores | Interval  (difference between two sat scores is meaningful) |
| Years of Education | Ratio |

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

Answer :

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

Answer :

a)

b)

c)

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?

Answer :

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 (x) | Probability P(X=x) |
| 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

Answer : let E(X) = expected number of candies for a randomly selected child

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**

Answer : python file – “Assignment\_01”

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?

Answer :

We know that, E(X) = mean

E(X) = (sum of observations)/(total no. of observations)

E(X) = 1308/9

E(X) = 145.34

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer : python file – “Assignment\_01”

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer : python file – “Assignment\_01”

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



Answer :

1. Height of the bar represents frequency of weight class.
2. It is a right skewed distribution.
3. Distribution is leptokurtic in nature.
4. Weight class with longest bar represent mode .
5. Mode<median<mean



Answer :

1. Distribution is right skewed.
2. Bold line in the rectangle represents median.
3. Points shown above in the small circular shape are all outliers.
4. Height of rectangle gives idea about variance in data.
5. The lower and upper vertical lines represents the minimum and maximum values respectively.

**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?

Answer : python file – “Assignment\_01”

**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?

Answer : python file – “Assignment\_01”

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

Answer : Symmetric distribution

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

Answer : Right-skewed distribution

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

Answer : Left-skewed distribution

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

Answer : Leptokurtic curve

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

Answer : Platykurtic curve

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



What can we say about the distribution of the data?

Answer : Asymmetric distribution

What is nature of skewness of the data?

Answer : Left-skewed

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

Answer : 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.

Answer : let boxplot 1 represent distribution 1 and boxplot 2 represent distribution 2.

1. Distribution 1 lies in the range 240 – 290 (approx.) with lower whisker = 240, lower hinge = 255, upper hinge = 280 , upper whisker = 290
2. Distribution 2 lies in the range 200 – 350 (approx.) with lower whisker = 200, lower hinge = 225, upper hinge = 312 , upper whisker = 350
3. Both the distribution represented by boxplot 1 and boxplot 2 are symmetric in nature.
4. Both the distribution have same median.
5. Distribution 1 has small IQR while distribution 2 has large IQR.
6. There is no outlier in any distribution.

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)

Answer : python file – “Assignment\_01”

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer : python file – “Assignment\_01”

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

Dataset: wc-at.csv

Answer : python file – “Assignment\_01”

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

Answer : python file – “Assignment\_01”

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

Answer : python file – “Assignment\_01”

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

Answer : python file – “Assignment\_01”