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

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 | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Nominal |
| Sales Figures | Nominal |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Insterval |
| SAT Scores | Nominal |
| Years of Education | Nominal |

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

Ans: 1/12

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. Sum equal to 1- : Not possible. Hence probability is 0/36
2. Sum less than or equal to 4 : 5/36
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: 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:** Approximately 3

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:** Done in python file named: Answer-Assignment-1.

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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** Done in python file named: Answer-Assignment-1.

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



The above histogram is showing Chick Weight in grams and frequency representing number of chicks on that particular weight. The weight is being divided into eight parts, each with difference of 50 grams, starting from zero till 400 grams. In the chart the lowest weight was lesser than 25 grams. The most number chicks weighed between 50 to 100 grams. About 200 chicks weighed between 50 to 100 grams. Most of the chicks weighed under 200 grams and a few of them lesser than even 25 grams.



The above is a figure of a boxplot. The box plot is a way of graphical representation of data where atleast 60% of the data is present within the range of the box. Hence as the box is formed near lower end of the plot, that is where majority of the data is present. The higher and lower end of the plot is the range of the data, and it presents the data which does not form part of box in the box-plot.

**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:** Done in python file named: Answer-Assignment-1.

**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: 41.0, Median: 40.5, Variance: 24.11, Standard deviation: 4.91

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

Ans: The skewness will be **symmetrical,** if mean and median are equal.

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

Ans: If mean is more than median of a data, it indicates right-skewed distribution. It means there are more values towards lower end of data as compared to higher end of data. The tail will be at left side of distribution.

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

Ans: If median is more than mean of a data, it indicates left-skewed distribution. It means there are more values towards higher end of data as compared to lower end of data. The tail will be at right side of distribution.

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

Ans: Positive kurtosis value indicates that the data has heavier tails than normal distribution. It means there will be relatively more extreme values.

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

Ans: Positive kurtosis value indicates that the data has lesser amount of tails than normal distribution. It means there will be relatively lower extreme values.

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



Q. What can we say about the distribution of the data?

Ans: As we can see in the above boxplot, the maximum value is about 19 and minimum value is about 1. Q1 is 10 and Q3 is 18. So, the middle 50% of the data lies between 10 and 18. Henc IQR of the boxplot is 8. The mean is near to 15, means it lies in the middle of data. It seems there are no outliers. The median is closer to the third quartile (Q3), not the first quartile (Q1).

Q. What is nature of skewness of the data?  
Ans: In summary, this boxplot suggests that the data is somewhat positively skewed

Q. What will be the IQR of the data (approximately)?   
Ans: The IQR is approx. 8

Q19) Comment on the below Boxplot visualizations?



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

Ans: Both the boxplots have similar median, which is 262.5. The data range of box plot 1 is narrower than box plot 2. The range of second boxplot or the blue one is 187.5 to 337.5, while the range of first boxplot is from 237.5 to 287.5. Q1 of first boxplot is about 275 and that of second box plot is 312.5. Q3 of first boxplot is about 255 and that of second box plot is approximately 225.

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:** Done in python file named: Answer-Assignment-1.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

**Ans:** Done in python file named: Answer-Assignment-1.

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

**Ans:** Done in python file named: Answer-Assignment-1.

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

**Ans:** Done in python file named: Answer-Assignment-1.

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:** Done in python file named: Answer-Assignment-1.