|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Ratio  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 | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| 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 | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Interval |

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

Ans: P(H H T) + P(H T H)+ P(T H H)

=1/8+1/8+1/8= 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. There is no outcomes which corresponds sum is equal to 1. It is 0/36. So, probability is 0.
2. (1,3)(2,2)(3,1)=3Outcomes, that is 3/36=1/12.
3. 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:

P (2R, 3G, 2B)

P (5/7, 4/6) = 20/42 = 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: Child A- probability of having 1 candy= 0.015

Child B- probability of having 4 candy = 0.20

Expected no. of candies for randomly selected child = 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: for points dataset;

Mean= 3.596563, Median= 3.695, Mode= 3.07, Variance= 0.285

Minimum range =2.17

Inferences: The data is concentrated around median.

There are no outliers.

The distribution is right skewed

For Score dataset; Mean= 3.217 , Median = 3.325 , Mode= 3.44 , Variance = 0.957,

Range= 3.910

Inference: The data is concentrated around Median.

There are 3 outliers: 5.25, 5.424, 5.345

The distribution is left skewed.

For weigh dataset; Mean= 17. 848 , Median = 17.710 , Mode = 17.02, Variance = 3.193

Range = 8.39

Inference: The data is concentrated around Median.

There is one outlier 22.90.

The distribution is left skewed.

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: 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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer: skewness for speed = -0.113**

**Skewness value is negative so it is left skewed. Since, magnitude is slightly greater than 0 it is slightly left skewed.**

**Skewness for distance = 0.782, Right skewed. Thus, positive skewness.**

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



Answer: The most of the data points are placed in the range 50-100 with frequency 200.

And least range of weight is 400 somewhere around 0-10.

So, the expected value the above distribution is 75.

From the histogram, right skewness is observed.



Answer: Median is less than mean.

Right skewed.

Outliers are on the upper side of box plot.

Less data points between Q1 and bottom point.

**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: confidence interval for 94% is 1.88

Confidence interval for 98% is 2.33

Confidence interval for 96% is 2.05

**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, Variance=24.111, Standard Deviation= 4.910

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

Ans: symmetrical

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

Ans: Right Skewed

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

Ans: left Skewed

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

Ans: The positive kurtosis value indicates for a data that the data is normally distributed and kurtosis value is 0.

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

Ans: it indicates that the data distribution has lighter tails and a flat peak than the normal distribution.

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



What can we say about the distribution of the data?

Ans: let’s assume above box plot is about age of children below 18 to take covid 19 vaccination. 50% of the children are above 10 years old and remaining are less. And children whose age is above 15 are approximately 40%.

What is nature of skewness of the data?

Ans: left skewed, as median is greater than mean.

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

Ans: approximately=-8

Q19) Comment on the below Boxplot visualizations?



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

Answer: Data in boxplot 1 has lower variability than data in boxplot 2. Both boxplots show a normal distribution since all the sections of the boxplot are equally distributed and are symmetrical. Hence, the skewness is 0.

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)

=33/81

* 1. P(MPG<40)= 67/81
  2. P (20<MPG<50)=69/81

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: MPG of cars does follow normal distribution approximately as mean and median are approximately same.

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

Dataset: wc-at.csv

Answer: in case of AT, mean> median, data is positively skewed.

And in case of Waist, mean > median, median is slightly shifted towards left. Data is fairly symmetrically distributed.

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

From the wc-at.csv, z-score of 90% confidence interval = 1.644

z-score of 94% confidence interval= 1.88

z-score of 60% confidence interval = 0.841

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

t score of, 95% confidence interval = 1.96

96% =2.05

99% = 4.50

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: For probability calculations, the no. of degrees of freedom = n-1, so there is a need the t- distributions with 17 degrees of freedom.

X= sample mean =260

Mu= population mean = 270

S = standard deviation of the sample = 90

n = no. of items in sample= 18

t score= -0.417

The probability that t<-0.471 with degrees of freedom assuming the population mean is true, the t- value is less than the t-value obtained with 17 degrees of freedom and t score of –0.417, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.