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

H H H

H H T

H T H

T H H

T T H

T H T

H T T

T T T

= 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 = 0
2. Less than or equal to 4 = (2,2),(1,1),(1,2),(2,1) = 4/36
3. Sum is divisible by 2and 3 = (1,5),(3,3),(4,2),(5,1),(2,4),(6,6) = 5/36

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?

Total = 7C2 = 21

5 balls green and red = 5C2 = 10

= 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

E(Candies) = (0.015\*1 + 0.20\*4 + 0.65\*3 + 0.005\*5 + 0.01\*6 + 0.120\*2)

=0.015+0.80+1.95+0.025+0.06+0.240

=3.905

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

For Points,

Mean = 3.596563

Median = 3.695

Mode = 3.92

Variance = 0.285881

Standard Deviation = 0.534679

Range = 2.17

For Score,

Mean = 3.21725

Median = 3.325

Mode = 3.44

Variance = 0.957379

Standard Deviation = 0.978457

Range = 3.911

For Weight,

Mean = 17.84875

Median = 17.71

Mode = 17.02

Variance = 3.193166

Standard Deviation = 1.786943

Range = 8.4

**Use Q7.csv file**

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?

Expected value = 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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



The above histogram tells that the data is right skewed. That is it is positively skewed.

The boxplot tells that the data has outliers.

**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?

**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.

Mean = 41

Median = 40.5

Variance = 25.52941

Standard Deviation = 5.052664

1. What can we say about the student marks?

The data can be said that it is normally distributed as both the mean and median are almost same.

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

The nature of skewness is zero skewness. The distribution is symmetric.

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

The nature of skewness is positively skewed/right skewed.

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

The nature of skewness is negatively skewed/left skewed.

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

The distribution of the data is peaked and has thick tails.

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

The distribution of the data is flat has thin tails.

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



What can we say about the distribution of the data?

The distribution of the data is not normal as the whiskers are not of same length.

What is nature of skewness of the data?

The nature of the skewness is left skewed.

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

The inter quartile range of the data is between 10 to 18.

Q19) Comment on the below Boxplot visualizations?



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

The distribution of the boxplots are near to the normal as the whiskers take almost equal percentages of the data.

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Nearly normal.

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

Dataset: wc-at.csv

Nearly normal.

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

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

Q 24**)**A Government companyclaims 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