|  |  |
| --- | --- |
| 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(Categorical) |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete(Categorical) |

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 | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

A) P=0.375

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

A) a) p(sum=1)=0

b) p(sum<=4)=1/6=0.16

c) p(sum divisible by 2 and 3)=24/36=0.66

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?

A) p=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

A) Expected number of candies = 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.

A) For points:

* Mean = 3.5965
* Median = 3.6950
* Mode = Bi modal(3.92, 3.07)
* Variance = 0.28588
* Standard deviation = 0.5346
* Range = 2.17(max = 4.93, min = 2.76)

For points the mean and median are nearer to each other. It is bi modal

For Score:

* Mean = 3.21725
* Median = 3.325
* Mode = 3.44
* Variance = 0.957
* Standard deviation = 0.978
* Range = 3.911(max = 5.424,min = 1.513)

For score the mean and median are nearer. It is uni-modal data. It is having high range and outliers are present in the data set.

For Weight:

* Mean = 17.848
* Median = 17.71
* Mode = Bi modal(17.02,18.9)
* Variance = 3.193
* Standard deviation = 1.786
* Range = 8.39(Max=22.9, Min=14.5)



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?

A) 145.3

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

**Cars speed and distance**

**A) skewness(speed) = -0.11**

**Kurtosis(speed) =-0.57**

**skewness(distance) = 0.78**

**Kurtosis(distance) = 0.24**

**Skewnss(sp) = 1.58**

**Kurtosis(sp) = 2.72**

**Skewness(WT) = -0.603**

**Kurtosis(WT) = -0.81**

****

**SP and Weight(WT)**

****

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





A) **Histogram:** Data in the Histogram indicates that the data is positively skewed and the weight of the data is allocated in the left side. The data set is right skewed from this we can infer that frequency of the data is decreasing from mean to right side.

**Box plot:** Box plot here represents that data is positively skewed and the most of the data is allocated between lower whisker and inter quartile range distance is lower. The distance between the higher whisker and the third quartile is larger that indicates the lower density of the data

**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 ?

**A)** Confidence interval for 94% =200±1.26

Confidence interval for 98% =200±1.56

Confidence interval for 96% =200±1.37

**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 = 25.5

Standard deviation = 5.05

1. What can we say about the student marks?

A) Student marks mean and median are not equal. The few marks are away from the mean and out of interval given by the deviation and the outliers are 49 and 56.

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

A) When the **mean** is **equal** to the **median**, then the distribution is symmetric, and the distribution has zero **skewness**.

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

A) The nature of the skewness when mean > median is positvely skewed or right skewed and the weight of the data is located in left side.

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

A) The nature of the skewness when median > mean is negatively skewed or left skewed and the weight of the data is located in right side.

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

A) The positive kurtosis indicates the high tailedness of the curve

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

A) The negative kurtosis indicates the low tailedness of the curve

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



What can we say about the distribution of the data?

A) The data is not symmetric

What is nature of skewness of the data?

A) The nature of the skewness is left skewed or negatively skewed

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

A) 8

Q19) Comment on the below Boxplot visualizations?



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

A) Both the groups have the same median and the data in the boxplot is closer to the median and the data in boxplot2 is not closer while comparing to the data set of boxplot1. IQR of the boxplot2 is larger while compared to the boxplot1.

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)

0.35

* 1. P(MPG<40)

0.729

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

A) MPG follows normal distribution and data is not perfectly normal but it follows up to some extent.

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

Dataset: wc-at.csv

A) AT and Waist follows normal distribution

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

A) Z(90%) = 1.28

Z(94%) = 1.88

Z(60%) = 0.25

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

A) t(95%) = 1.71

t(99%) = 2.49

t(96%) = 1.82

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

A) t score=sample mean-population mean/s/n0.5 =-0.47

Probablility=0.32