|  |  |
| --- | --- |
| 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 | categorical |
| Number of kids | discrete |
| Number of tickets in Indian railways | discrete |
| Number of times married | discrete |
| Gender (Male or Female) | 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 | ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | ratio |
| Type of living accommodation | ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | interval |
| Sales Figures | ratio |
| Blood Group | nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | ratio |
| Number of Children | ratio |
| Religious Preference | nominal |
| Barometer Pressure | nominal |
| SAT Scores | ordinal |
| Years of Education | ratio |

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

(HHH,HHT,HTH,TTT,THT,HTT,TTH,THH)

Total possible events=8,desired events=3

P=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)

a)0

b)1/6

c)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?

A) 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)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.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | sd | range |
| points | |  |  | | --- | --- | |  |  |   3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| score | 3.22 | 3.325 | 3.44 | 0.96 | 0.98 | 3.91 |
| weight | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

No variable has mean=median=mode

Values of weight are much higher than score and points

Score variable has some outliers

**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?

Ans)145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans)Skewness for speed and distance are -0.11,0.80**

**Kurtosis for speed and distance are -0.50,0.40**

**Both speed and distance have positive kurtosis values**

**Speed is negatively skwed whereas dist is positively skwed.Data has distribution more on left for dist while speed has more distribution on right**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**The skewness and kurtosis for Sp are 1.61 and 2.977**

**While for weight are -0.61 and 0.95**

**Sp is positively skewd while weight is negatively skewed**

**Both sp and wt have positive kurtosis**

**Sp has distribution more concentrated on left while wt has more on right**

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



Ans) most of the chicks have weight ranging between 50 to 100 followed by 100 to 150 and 150 to 200

Data is positively skewed

Chick weights can be classified into three types overweight normal and under weight.

Overweight🡪 >200

Normal—50--200

Underweight -- <50



Ans)Data contains outliers.

Data is positively skewed

**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?

94% 98% 96%

Upper 201.26 201.56 201.37

Lower 198.73 198.43 198.62

**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 –40.5

Variance –25.52

Std – 5.052

Data is not normally distributed

Data has outliers

Most of the students marks are between 35-45

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

A)Skewness =0,symmetric bell shaped curve

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

A)Positive skewness data is more distributed on left

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

A)Negatively skewed.Data is more distributed on right

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

A)Kurtosis indicates peakedness of the data.High and narrow peak on the central part of data

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

A)Wider peak on central part of the data

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



What can we say about the distribution of the data?

A)Data is not symmetric.More concentrated on right side.

What is nature of skewness of the data?

A)Negative

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

A)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.

Ans)Data is normally distributed for both plots with centre around 262.First plot has less range while second plot has more range.

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)
  3. P (20<MPG<50)

Ans) a--0.4074074074074074

b ---0.7530864197530864

c ---0.8518518518518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans)Mpg of cars is nearly normally distributed

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

Dataset: wc-at.csv

Ans) Both AT and Waist do not follow normal distribution.

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

A) z scores for

90% ---1.6448

94%-----1.8807

60%-----0.841621

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

A)t scores

95%--------2.063

96%--------2.171

99%---------2.796

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) 0.32167411684460556