|  |  |
| --- | --- |
| 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?

Ans: Total possible Events=8 No of desired outcomes=3

P(Two heads and one tail) = Number of possible outcomes

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)dice 1 dice 2 sum

1 1 2

1 2 3

1 3 4

2 1 3

2 2 4

3 1 4

6/36 = 1/6

c)(1,5),(3,3),(4,2),(5,1),(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?

Ans: red = 2 green = 3 blue = 2

C(5,2) = 5!/(2!\*(5-2)!)=10

C(7,2) = 7!/(2!\*(7-2)!)=21

Probability = Favorable outcomes/Total outcomes

=10/21 or 0.4762

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.

**Use Q7.csv file**

Ans:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Std | 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 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

From the given data, there is no such case for variable mean = median =mode and thus, as seen in the graph after plotting from the given data there are outliers in “Score” and “Weigh”.

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 | kurtosis |
| speed | -0.118 | -0.51 |
| dist | 0.81 | 0.41 |

“dist” is positively skewed and has positive kurtosis where as “speed” has negatively skewed and has negative kurtosis.

The distribution of dist from the given data has been concentrated on the left, as speed has distribution on the right.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | skewness | kurtosis |
| SP | 1.611 | 2.98 |
| WT | -0.614 | 0.95 |

“SP” has positively skewed where as “WT” is negatively skewed and Both “SP” and “WT” has positive kurtosis.

The distribution of SP from the given data has been concentrated on the left, as WT has distribution on the right

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



Ans:

Majority of the chicks has in the range of 50-100, followed by 100-150 and 150-200

Data is positively skewed

We can define 3 categories: underweight<50, average weight 51-150, overweight>150



Ans:

Data is positively skewed

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?

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 94% | 98% | 96% |
| upper | 201.04 | 201.38 | 201.17 |
| lower | 198.96 | 198.62 | 198.83 |

**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.

ANS:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.50 |
| Variance | 25.53 |
| Standard deviation | 5.05 |

1. What can we say about the student marks?

Ans:

Data is not normally distributed

Data has outliers

Majority of the students have marks in the range of 35-45

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

Ans: Skewness=0. Perfectly symmetric bell-shaped curve

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

Ans: Skewness = positive. Data is distributed more on the left side

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

Ans: Skewness = negative. Data is distributed more on the right

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

Ans: A positive kurtosis value indicates High and narrow peak on central part of the data.

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

Ans: A negative kurtosis value indicates 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?

Ans: Data is not symmetric. Data is concentrated towards right side.

What is nature of skewness of the data?

Ans: Skewness = negative

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

Ans: IQR of the data is 8

(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 the graphs and median around 262.5, but first graph has less range and second graph has more range. No outliers are present in both the graphs.

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)

Ans: 0.348

* 1. P(MPG<40)

Ans: 0.729

* 1. P (20<MPG<50)

Ans: 0.013116469610523374

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 follows Normal Distribution

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 the normal distribution

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

Ans:

|  |  |
| --- | --- |
| 90% | 1.6449 |
| 94% | 1.8808 |
| 60% | 0.8416 |

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

Ans:

|  |  |
| --- | --- |
| 95% | 2.0638985616280205 |
| 96% | 2.1715446760080677 |
| 99% | 2.796939504772804 |

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: t-score = -0.4714

Degrees of freedom = 17

P(t) = 0.3216725