Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of Gold | Continuous data |
| Distance between two places | Continuous data |
| Length of a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| 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?

Ans:-Three coins are tossed=[HHH, TTT, HTT, THH, THT,HTH,HHT,TTH)

Two Heads and One Tail = [THH, HHT, HTH]

Number of outcomes are 3.

Probability of Two Heads and One Tail = 3/8 = 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

Ans: -

1. Equal to 1 = 0
2. Less than or equal to 4 = [2,2] [1,3] [3,1] = 3/36 = 0.083
3. Sum is divisible by 2 and 3 = 1

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: -

Given,

Red ball = 2

Green ball = 3

Blue ball = 2

Number of balls = 7

Probability =?

* When we took randomly two balls, balance is = 5/7
* And inside the bag there is left with 6 balls with 2 blues =4/6
* So, the probability is (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: - 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=0.015+0.8+1.95+0.025+0.06+0.24

=3.090

=**3.09**

The Expected number of candies for a randomly selected child is **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:-**

****

The value of mean and median are nearly equal so we can say that the curve is nearly **symmetric.**

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: -

Given,

Values=108, 110, 123, 134, 135, 145, 167, 187, 199

Expected value=?

Formulae: -E(x)=(probability \* value)

=**145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: -**In the graph speed data is negatively skewed and distance data is positively skewed.

**SP and Weight (WT)**

**Use Q9\_b.csv**

**Ans: -**In the graph SP data is positively skewed and WT data is negatively skewed.

**CODE: -**

****

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



Ans: -

* The data is positively skewed.
* We can define three categories in this that is Lower weight>50 , Average weight 51-150, Over weight<150.
* There is some outliers may be.



Ans: -

* More outliers are present.
* 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?

Ans: -

Given,

Sample size=2000

Sample mean=200

Standard deviation=30

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| CI = | |  |  | | --- | --- | | X̄ ± Z× | s | |  | | √n | |

For 94% the confidence interval is 198.738-201.262 .

For 96% the confidence interval is 198.622-201.378 .

For 98% the confidence interval is 198.439-201.561 .

**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 = 24.11

Standard deviation = 5.05

1. What can we say about the student marks?

Most of the students got nearly 40 marks.

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

Ans: - if the mean, median of data are equal then the distribution is symmetric and it has zero skewness.

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

Ans: -The distribution is positively skewed.

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

Ans: -The distribution is negatively skewed.

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

Ans: -It indicates that distribution is peaked and possesses thick tails.

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

Ans: -It indicates the distribution has lighter tails 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: - The data is more concentrated towards right side and the median is approximately 15.

What is nature of skewness of the data?

Ans: -It is negatively skewed or Asymmetric.

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

Ans: -Inter quartile range is 8.

Q19) Comment on the below Boxplot visualizations?



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

Ans: -

* The mean and median are same value because the median of the both boxplot is same approximately 262.5.
* Boxplot 1 having less data sample compare to Boxplot 2.
* Both boxplots are symmetric.

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

Ans:-

* 1. P(MPG>38) = 0.34
  2. P(MPG<40) = 0.72

c. P (20<MPG<50) = 0.95

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: - The mean and median are nearly same so it is 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: -

* Adipose Tissue (AT)

The mean and median are nearly same so it is normal distribution.

* Waist Circumference (Waist)

The mean and median are varying so it is not a normal distribution.

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

Ans: -

The Z scores of 90% confidence interval is 1.645.

The Z scores of 94% confidence interval is 1.86.

The Z scores of 60% confidence interval 1.55.

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

Ans: -

The T scores of 95% confidence interval is 2.063.

The T scores of 96% confidence interval is 2.171.

The T scores of 25% confidence interval is 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

Ans: -

Given,

average light bulb lasts = 270 days

sampled bulbs last an average =260 days

standard deviation = 90 days

number of bulbs = 18 bulbs

Ho=mu>=260

Ha=mu<=260

T= 0.4714045207910317

Alpha=0.05

P= 0.6783274643290165

P >alpha

Hence accept Ho.