|  |  |
| --- | --- |
| 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 Scale |
| Weight | Ratio Scale |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval Scale |
| Height | Ratio Scale |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio scale |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval Scale |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval Scale |
| SAT Scores | Interval Scale |
| Years of Education | Ordinal |

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

Ans. Total number of possibilities= HHH,TTT,HHT,HTT,HTH,THT,TTH,THH.

The probability of getting HHT=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.** Total number of possibility=36

**a.** probability that sum is equal to 1 =0

**b.** Possibilities of getting number sum is less than or equal to 4 is (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

Probability of sum is less than or equal to 4 =6/36=0.1667

**c.** Possibilities of sum is divisible by 2 and 3 is (1,5), (2,4), (3,3), (4,2), (5,1), (6,6)

P(sum is divisible by 2 & 3)= 6/36=0.1667

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.** Total number of balls=(2+3+2)=7

Let S be the sample space

Then, n(S)=Number of ways of drawing 2 balls out of 7

== =21

Let E =Event of drawing 2 balls, none of which is blue.

n(E)=Number of ways of drawing 2 balls out of (2+3) balls.

= = = 10

P(E)=n(E) / n(S)=10/21=0.476

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.** Expected Value = (1/6)**{** 0.015+0.8+1.95+0.025+0.06+0.24}=0.515

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.

**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.** Expected Value= (1/9){108+110+123+134+135+145+167+187+199}

=145.3

**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**



**Ans.** Histogram: The histograms peak has right skew.

Boxplot: The boxplot has outliers on the maximum side.

**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.** Sample size = 200

The interval is

For 94% confidence interval

df=200-1=199

t=1.8916

For 94% confidence interval is (198.73,201.27).

For 98% confidence interval

t = 2.3452

For 98% confidence interval is (198.43,201.57).

For 96% confidence interval

t = 2.0673

For 96% confidence interval is (198.61,201.39).

**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=µ= = 41

Median= = 40.5

Variance = 25.52

Standard deviation==5.05

1. What can we say about the student marks?

**Ans.**

* It doesn’t have outliers.
* The Data is skewed towards right.

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

**Ans.** No skewness is present we have a perfect symmetrical distribution.

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

**Ans.** Skewness towards Right.

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

**Ans.** Skewness towards Left.

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

**Ans.** Positive kurtosis value indicates that distribution is peaked and possess thick tails .

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

**Ans.** Negative Kurtosis means the curve will be flatter and broader.

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



What can we say about the distribution of the data?

**Ans.** Boxplot is not normally distributed the median is towards the higher value.

What is nature of skewness of the data?

**Ans.** The data is skewed towards left.

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

**Ans.** IQR= Upper Quartile – Lower Quartile

=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.**

* There are no outliers
* Both Boxplot has same Median

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

**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.** AT and Waist does not follow Normal Distribution

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 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