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

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

* 3 coins tossed = {HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}

Total no. of outcomes = 8

Probability that 2 heads and 1 tail are HHT,HTH,THH

i.e Probability that 2 heads and 1 tail = 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

* 2 dice rolled = {(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)

(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)

(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)

(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)

(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)

(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)}

Total no. of outcomes= 36

1. Sum equal to 1 = 0 = 0/36

Probability that sum equals to 1 = 0

1. Sum is less than or equal to 4={(1,1)(1,2)(1,3)(2,1)(2,2) (3,1)}=6

Probability that Sum is less than or equal to 4 = 6/36 = 1/6

1. Sum is divisible by 2 and 3={(1,5)(2,4)(3,3)(4,2)(5,1)(6,6)} =6

Probability that sum is divisible by 2 and 3 = 6/36 = 1/6

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 bag contains 2 red, 3 green and 2 blue balls= 2R+3G+2B =7

Choosing 2 balls from all 7 balls = 7C2=21

7balls – 2 Blue balls = 5

The possible outcome of 2 ball from 5 non blue = 5C2=10

Probability that none of the balls drawn is blue=10/21

Q6) Calculate the Expected number of candies the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)for a randomly selected child

Below are

|  |  |  |
| --- | --- | --- |
| 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

* Expected number of candies for a randomly selected child =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

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**

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?

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

Probability of value=1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9

Expected Value= 1/9\*108+1/9\*110+1/9\*123+1/9\*134+1/9\*135+1/9\*145+1/9\*167+1/9\*187+1/9\*199

=12+12.22+13.66+14.88+15+16.11+18.55+20.77+22.11

=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**



* THE HISTOGRAM SHOWS POSITIVE SKEWNESS AND POSITIVE KURTOSIS.



* THE BOXPLOT SHOWS OUTLIERS IN LOWER QUARTIL.

**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?**

**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 = 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=738/18

=41

* Median = 40+41/2

= 40.5

* Mode = 41
* Standard deviation =

=

=

=

1. **What can we say about the student marks?**

* THE STUDENT GETS THE AVERAGE MARKS 41, THE STUDENT HAS SCORED 41 MARKS MOST OF THE TIMES.

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

* THE NATURE IS NORMALLY DISTRIBUTED THAT IS ZERO SKEWNESS.

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

* POSITIVE SKEWNESS

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

* NEGATIVE SKEWNESS

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

* THE KURTOSIS HAS THIN PEAK AND THE TAIL IS WIDE THAT IS THE DISTRIBUTION IS LESS IN THE PEAK.

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

* THE NEGATIVE KURTOSIS VALUE INDICATES FOR DATA THAT THE DISTRIBUTION IS FLAT AND THE TAIL IS THIN.

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



**What can we say about the distribution of the data?**

* THE DISTRIBUTION OF BOXPLOT IS LEFT SKEWED AND HAS NO NORMAL DISTRIBUTION.

**What is nature of skewness of the data?**

* THE BOXPLOT IS NEGATIVE SKEWNESS

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

* 15.5

**Q19) Comment on the below Boxplot visualizations?**



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

* BOTH THE BOXPLOT HAS SAME MEAN. 1ST BOXPLOT SHOWS THE NEGATIVE SKEWNESS, 2ND BOXPLOT HAS NO SKEWNESS THAT IS IT HAS NORMAL DISTRIBUTION. FOR 1ST BOXPLOT THE LOWER EXTREME IS 287.5 APPROXIMATELY, AND THE UPPER EXTREME IS 240 APPROXIMATELY. FOR 2ND BOXPLOT THE LOWER EXTREME IS 337.5 APPROXIMATELY, THE UPPER EXTREME IS 187.5 APPROXIMATELY. THERE ARE NO OUTLIERS IN BOTH THE BOXPLOT.

**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)**

**Q 21) Check whether the data follows normal distribution**

1. **Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

* THE “DATASET: CARS.CSV” BOXPLOT IS NOT A NORMAL DISTRIBUTION. IT IS POSITIVE SKEWNESS

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

**Dataset: wc-at.csv**

* THE ADIPOSE TISSUE (AT) AND WAIST CIRCUMFERENCE (WAIST) FROM WC-AT DATA SET DO FOLLOWS 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**

* T score =

=260-270 /90 /4.24

= -10/21.22

T score = -0.471