|  |  |
| --- | --- |
| 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 | 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 | Nominal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |

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

3ans: The total probability of three coins tossed = 2^3=8

The probability getting two heads one tail =3(HHT,HTH,THH)

Then probability =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

4ans: The total probability of two dice rolled=6^2=36

1. Probability=0

Because the least sum two dice rolled is 2 (sum != 1)

1. P(p<=4)= P(p=1)+P(p=2)+P(P=3)+P(p=4)

P(p=1)=0/36

P(p=2)=1/36

P(p=3)=2/36

P(p=4)=3/36

P(p<=4)=0/36+1/36+2/36+3/36=6/36=1/6

1. Sum is divisible by 2&3 i.e, 6 and 12

The probability of getting 6 (2+4,3+3,4+2)=3/36

The probability of getting 12(6+6)=1/36

Then probability = 1/36+3/36=4/36=1/9

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?

5ans: The total balls =2+3+2=7

From 7 balls we are drawn 2 then probability is=7C2

Probability drawing 2 balls none of the drawn is blue=5C2

Total probability =5C2/7C2=5\*4/1\*2/7\*6/1\*2=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

Expected frequency=cc\*(P(p=1)+P(p=2)+P(p=3)+P(p=4)+P(p=5)+P(p=6))

=1\*0.015+2\*0.120+3\*0.65+4\*0.20+5\*0.005+6\*0.01

=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**

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?

8ans: Expected value = summation i=1 to n(pi\*(xi))

= 1/9(108+110+123+134+135+145+167+187+199)

= 1038/9

=145.33

**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:The histogram has positively skewed and it has long tail on right side, mean>median and has out liers on right side



* **out liers**
* **upper limit**
* **Third quartile**
* **Mean**
* **Frist quartile**
* **Lower limit**
* The space between first quartile and third quartile is called interquartile range
* Boxplot plots are used for identify out liers

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

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

13ans: The Nature of Skewness when mean and median of the data

are equal is ‘skewness=0’

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

14ans: The nature of skewness when mean>median is ‘positively skewed’

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

15ans: The nature of skewness when median > mean is ‘Negatively skewed’

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

16ans: Positive kurtosis indicates for a data ‘flatter tails’ and ‘sharper peak’

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

17ans: Negative kurtosis value indicates for a data ‘flatter peak’ and ‘thinner tails’

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
18ans: 🡪 The most of data distributed towards left

🡪 The nature of skewness of the data is negative

🡪 The IQR = Upper Quartile (UQ)-Lower Quartile(LQ)

Where, UQ=18

LQ=10

IQR=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.

19ans:🡪 The both box plots has same median and it different Inter quartile range,Q1 and Q3 values are different

* No out liers

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

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

Dataset: wc-at.csv

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