|  |  |
| --- | --- |
| 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 |  |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| 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 | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans : When three Coins are Tossed No of total possible outcomes are 8 which are HHH,HHT,HTH,HTT,THH,THT,TTH,TTT

Number of Outcomes When Two heads and one tail is obtained are 3 which are HHT,HTH,THH

Therefore Probability =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 : When two dice are rolled No of outcomes are 36

1. No of possible outcomes Having sum Equal to 1 are zero (0)

Probability = 0

1. No of possible outcomes Having sum Less than or Equal to 4 are 6

Probability = 6/36 =1/6=1.666

1. No of possible outcomes Having sum divisible by 2&3 are 6

Probability = 6/36 =1/6=1.666

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 :

When tow balls are drawn from bang of 2R,3G,2B balls

No of possible outcomes are 7C2 = 21

No of outcomes that none of the balls drawn is blue are 5C2 = 10

Therefore Probability = 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: 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.

|  |  | **Points** | **Score** | **Weigh** |
| --- | --- | --- | --- | --- |
| **0** |  | 3.07 | 3.44 | 17.02 |
| **1** |  | 3.92 | NaN | 18.90 |

|  | **Points** | **Score** | **Weigh** |
| --- | --- | --- | --- |
| **mean** | 3.596563 | 3.217250 | 17.848750 |
| **median** | 3.695000 | 3.325000 | 17.710000 |
| **var** | 0.285881 | 0.957379 | 3.193166 |
| **std** | 0.534679 | 0.978457 | 1.786943 |

Ans: Refer to Q7.ipynb

Mode

Range

Points 2.170

Score 3.911

Weigh 8.400

**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 of Weight of that patient = 145.33

//Mean of all Data

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans: Refer Q9\_a.ipynb

Skewness Kurtosis

speed -0.117510 speed -0.508994

dist 0.806895 dist 0.405053

Speed have negative skewness &kurtosis where dist have positive skewness & kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans: Refer Q9\_b.ipynb

Skewness Kurtosis

SP 1.611450 SP 2.977329

WT -0.614753 WT 0.950291

In above data Column SP is positively Skewed And also Have positive Kurtosis

And WT have negative skewness and positive kurtosis

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



Ans:

Above histogram Shows that Data as positively Skewed & majority of chicks have weight between 50 to 100



Ans: Above data in boxplot have outliers and it 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: For

94% = (201.043, 198.956)

98% = (201.378, 198.621)

96% = (201.378, 198.621)

**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?

Ans:

|  | **0** |
| --- | --- |
| **mean** | 41.000000 |
| **median** | 40.500000 |
| **var** | 25.529412 |
| **std** | 5.052664 |

Student scored Average marks of 41

Above Data has outliers

Above data is positively skewed

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

Ans : Skewness is Zero (0)

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

Ans : Skewness is positive

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

Ans : Skewness is negative

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

Ans: Positive Kurtosis value indicates that data data don’t follow normal distribution And have frequent outliers

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

Ans : Positive Kurtosis value indicates that data don’t follow normal distribution

It has Uniform Distribution And don’t have frequent outliers

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



What can we say about the distribution of the data?

Ans: Distribution of data have skewness

What is nature of skewness of the data?

Ans: Negative Skewness

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

Ans: 8 (18-10)

Q19) Comment on the below Boxplot visualizations?



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

Ans: Above both Box plots have same median but in box plot 1 data is distributed near to the median than box plot 2. And in box plot 2 data is widely distributed than box plot 1

And both boxplots have no outliers

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)

Ans:

1. P(MPG>38) =0.347
2. P(MPG<40) =0.729
3. P (20<MPG<50) =0.898

From the formula of probability

stats.norm.cdf(x,loc=mean,scale=std)

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 data sets don’t follow normal distribution

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

Ans: Z score for

90%=1.2815515655446004

94%=1.5547735945968535

60%=0.2533471031357997

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

Ans: At df = 25 T score for

95%=1.7081407612518986

96%=1.8248284689556018

25%=2.4851071754106413

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

Probability = 0.328

T score = 0.471