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

Q1) Identify the Data type for the Following: import scipy.stats as stats

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 | Nominal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

* S={HHH , HHT, HTH, THH, TTH, HTT, THT, TTT}

P(x =2)= probability of occurrence of 2 head and one tail is

{HHT, HTH, THH}=3/8

Q4) Two Dice are rolled, find the probability that sum is

Equal to 1

* Solution: 1/36

Less than or equal to 4

* Possible outcome = {(1,3), (3,1),(2,2),(1,1),(2,1),(1,2)}

=6/36 =1/6

Sum is divisible by 2 and 3

* Possible outcomes= {(3,3), (4,2), (2,4), (1,5), (5,1)}

=5/36

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?

* N=7

Total number of ways to select 2 balls out of 7 balls=7c2

Total number of ways to select 2 balls out of 5 balls which are not blue

=5c2 =10

Probability (none of the balls drawn is blue)

=7c2/5c2

=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

* Expected no. of candies= (1\*0.015)+(4\*0.20)+(3\*0.65)+(6\*0.01)+(2\*0.120)

=0.015+0.80+1.95+0.02+0.06+0.24

=3.085

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?

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

=1308/9

=145.333

**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**



* Histogram:= The most of the datapoints are concentrated 50 to 100 with frequency 200. And least range is 400 is between 0 to 10.the above distribution is 75.
* Boxplot median is less than mean rightskewed and outlier on the upper side of boxplot and there is less datapoints between q1 and bottom.

Q1<Q3

**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?

* N=2000

Mean u=200

Std.=30

Percent= 94%,98%,96%

**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 =4.910

1. What can we say about the student marks?

* The student has a range of score from 34 to 56 .the lowest score of student is 34 and highest is 56.

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

* Skewness is measure how asymmetric a distribution of data.

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

* +ve (right) skewness

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

* -ve(left) skewness

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

* Data is normaly distributed an kurtosis value is 0.

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

* The distribution with light tails and chance of lack of outliers.

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



What can we say about the distribution of the data?

* No outliers, not a normal distribution. Q1>Q3, left skewed.

What is nature of skewness of the data?

* Left skewed

What will be the IQR of the data (approximately)?   
 IQR= Q3-Q1

= 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.

* Boxplot follows normal distribution with same median, mode, mean.

Plot 1 having lesser range compared to 2

Plot 1 has less sample

Q1=Q3 for both 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)

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

* Mean= 270, X=260, std=90, n=18, df=19

T=(x-u)/SE