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

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

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

Ans :- P ( Two Heads and One Tail ) = 3/8 = 37.50 %

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 :- a) P (Equal to 1) = 0

b) P (Less than or equal to 4) = 6/36 = 1/6 = 16.67 %

c) P (Sum is divisible by 2 and 3) = 24/36 = 2/3 = 66.67 %

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 = P ( None of the balls drawn is blue )10/21 = 47.61 %

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 Number of Candies = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 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**

**Ans :- Que 7 Ans .ipynb**

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 :- Probability of Selecting Each Patients P(X) = 1/9

Expected Value = Summation [(Weights (X) \* P(X) ]

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

= 145.33 (in pounds)

Expected Value Of the Weight of that patients = 145.33 (in pounds)

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

Cars speed and distance

Use Q9\_a.csv

Ans :- Que 9 Ans.ipynb

1. The data is positively skewed or Right skewed for dist.
2. The data is negatively skewed or Left skewed for speed.
3. For Kurtosis the distribution is Normal for both dist & speed.

SP and Weight(WT)

Use Q9\_b.csv

Ans :- Que 9 Ans. Ipynb

1. The data is Positive skewed or Right skewed for SP .
2. The data is Negative skewed or Left skewed for WT .
3. For the kurtosis the distribution is Leptokurtik for SP & for WT the distribution is Normal or Mesokurtic.

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





Ans:- Histogram :- 1. Chick weight data is rightly skewed or we casn say Positively Skewed .

2. Most Weight is lying between 50-200 or we can say 50% of data is between 50-200.

3. Between 50-100 weight have highest frequency whereas 350-400 have lowest frequency.

4. The median will be lower than the mean because the mean is more sensitive to the higher

Values and is drawn towards the tail of the histogram.

Box Plot :- 1. The data is right skewed .

2. Plotted box plot has more outlier on upperside of the data.

3.That means most data is in upper quartile whereas less data in lower quartile

**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 :- Que 11 Ans .ipynb

**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:- 1. Que 12 Ans . ipynb

2. The average marks for the students is 41, As per data maximum student is below average i.e., out of 18 the 9 student are below average and 8 are above average. And there are approximately 5 students are more than the average.

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

Ans :- When mean and median of the data are equal then we can say data symmetric that means the data is equally distributed 50 % on both side of the median.

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

Ans :- When the mean is greater than median we can say data is positively skewed or right skewed that means most of value from data is lying left tail of the distribution.

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

Ans :- When the medain is greater than mean we can say data is negatively skewed or left skewed that means most of value from data is lying right tail of the distribution.

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

Ans :- Positive kurtosis value means Leptokurtic. The positive kurtosis indicates that the distribution is peaked or have thick tails. The Distribution more numbers are located in the tails of the distribution instead of around the mean.

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

Ans :- Negative kurtosis value means Platykurtic. The negative kurtosis indicates that a distribution is flat and has thin tails. The distribution is less peaked when compared with the normal distribution with the fever value is lighter and thinner tails.

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



What can we say about the distribution of the data?

Ans :- Approximately we can say distribution is normal or left skewed distribution. Here Distribution is shifted towards left that means mean is less than median.

What is nature of skewness of the data?

Ans :- As we see can most of data in box plot is on right side then we can say data left skewed.

What will be the IQR of the data (approximately)?   
Ans :- IQR = Upper Quartile – Lower Quartile = 18.1 – 10 = 8.1 (Approximately) .

Q19) Comment on the below Boxplot visualizations?



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

Ans :- From data There are no outlier.Both the box plot have same median. That is approximately between 250 to 275. They are normally distributed at Zero. No skewness at minimum or maximum whisker range.

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 :- Que 20 Ans .ipynb

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

Ans :- Que 21 Ans .ipynb

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

Ans :- Z score = (1+C.L)/2

Z score at 90% = (1+0.90)/2

= 1.90/2

= 0.95

So from the z score table close value is 1.65 and 1.64, Hence mean as

Z score at 90% = (1.65 + 1.64)/2 = 1.645

Z score at 94% = (1+0.94)/2

= 1.94/2

= 0.97

So from the z score table close value is 1.88 and 1.89, Hence mean as

Z score at 94% = (1.88+1.89)/2 = 1.885

Z score at 60% = (1+0.60)/2

= 1.60/2

= 0.80

So from the z score table close value is 0.86 and 0.85, Hence mean as

Z score at 60% = (0.86+0.85)/2=0.855

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

Ans :- Que 23 Ans .ipynb

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

Ans :-

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

As the sample size less than 30 use t test

t = (260-270) / [90/sqrt(18)]

t = -0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that **t < - 0.471 with 17 degrees of freedom** assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of **0.3217** assuming the mean life of the bulbs is 300 days.

Que 24 Ans .ipynb