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

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

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

SAMPLE SPACE:{HHH,THH,HTH,HHT,TTT,HTT,THT,TTH}

A={THH,HTH,HHT}

P(A)=3/8

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

1. Equal to 1=0
2. Less than or equal to 4=1/6
3. Sum is divisible by 2 and 3= 4/6

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?

n(S)=7C2=21

n(E)=5C2=10

P(E)=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

ANS-1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.12

=0.015+0.8+1.95+0.025+0.06+0.24

=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**

POINTS SCORE WEIGH

MEAN 3.596563 3.21725 17.84875

MEDIAN 3.695 3.325 17.71

MODE 3.92 3.44 17.02

VARIANCE 0.285 0.957 3.193

STD 0.534 0.978 1.786

RANGE

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

= (1/9) (1308)

= 145.33

EXPECTED VALUE OF THE WEIGHT OF THAT PATIENT IS 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SKEWNESS**

Index 0.000000

speed -0.117510

dist 0.806895

dtype: float64

**THE GRAPH OF SPEED HAS NEGATIVE SKEWNESS AND POSITIVE KUTOSIS**

**KURTOSIS**

​Index -1.200000

speed -0.508994

dist 0.405053

dtype: float64

THE GRAPH OF DIST HAS POSITIVE SKEWNESS AND POSITIVE KURTOSIS

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SKEWNESS**

Unnamed: 0 0.000000

SP 1.611450

WT -0.614753

dtype: float64

**KURTOSIS**

Unnamed: 0 -1.200000

SP 2.977329

WT 0.950291

dtype: float64

**THE SKEWNESS OF SP IS NEGATIVE AND KURTOSIS IS NEGATIVE**

**THE SKEWNESS OF WT IS NEGATIVE AND KURTOSIS IS POSITIVE**

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



HISTOGRAM

-CHICK WEIGHT DATA IS RIGHT SKEWED OR POSITIVELY SKEWED

-MORE THAN 50% CHICK WEIGHT IS NETWEEN 50 TO 150

-MOST OF THE CHICK WEIGHT IS BETWEEN 50 TO 100



BOXPLOT

-THE DATA IS RIGHT SKEWED

-THERE ARE OUTLEIR ON THE UPPER SIDE

**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:Sample mean=200,Sample Sd=30,n=2000

94% CI=(198.738325292158, 201.261674707842)

98% CI=(198.43943840429978, 201.56056159570022)

96% CI=(198.62230334813333, 201.37769665186667)

**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:41.0

MEDIAN:40.5

SD:4.9103

VARIANCE:24.111

1. What can we say about the student marks?

The scores are in uniform distribution and in ascending order.

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

ANS:When mean and median are equal ,the skewness is zero.

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

ANS:When mean is greater than median the skewness is positively skewed.

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

ANS:When median is greater than mean the skewness is negatively skewed.

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

ANS:Positive value of kurtosis indicates that a distribution is peaked and possess thick tails.

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

ANS:Negative value of kurtosis value indicates that a distribution is flat and has thin tail.

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



What can we say about the distribution of the data?

-The data is distributed in DE-assigned format.

What is nature of skewness of the data?

Left side skewed.

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

Q3-Q1=18-10=8

8 IS THE IQR

Q19) Comment on the below Boxplot visualizations?

There is a representation of 2 box plot 2) is highly distributed across the plane and the plane are1) is slightly less distributed.

It is seen that 100% of the data is spreded across values from 350 in 2 whereas it spread in range 250-290 in 1.



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

After comparing box plot 1 with box plot 2 the data in box plot 1 is widely spread.The data range varies high in box plot 2 it is hard to make prediction in box plot 2.The median in the 2 box plots are equal and the data spread in both of them are symmetrical.

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

P(MPG>38)= 0.3475939251582705

In [67]:

P(MPG<40)= 0.7293527263719559

In [6]:

c. P (20<MPG<50)= 0.8988689146142506

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

-IT IS NORMALLY DISTRIBUTED

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

Dataset: wc-at.csv

-BOTH ARE NORMALLY DISTRIBUTED

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

-CI 60% = 0.8416212

-CI 94% = 1.644854

-CI 96% = 1.880794

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

-CI 95% = 2.063899

-CI 96% = 2.171545

-CI 99% = 2.79694

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

-x=mean of the sample of bulbs=260

u=population mean=270

s=sd of sample=90

n=18

t=-0.471

The probability that t<-0.471 with 17 degree of freedom assuming the population mean is true,the t-valu is less than t-value obtained with 17 degree of freedom and a t-score of-0.71,the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of bulbs is 300 days.