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

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

1. One coin Possibility = (Head ,Trail)

Three coins Possibility = (H, H),(H, T),(T, H),(T, T)

Probability of two heads P(X=1)=1/8

Probability of one tail P(X=2)=2/8

P(X=1)+P(X=2) =+2/8

=3/8=0.3

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

1. Equal to 1
2. 0(its impossible to equal to 1)
3. Less than or equal to 4
4. 1+1=2

1+2=3

1+3=4

3+1=4

2+1=3

2+2=4

1. Sum is divisible by 2 and 3

n(s)=36

Sum is divisible by 2 = 2/36

= 1/18

Sum is divisible by 3 = 3/36

= 1/12

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?

1. Totall no of balls =2+3+2

=7

N(s) = No of ways to draw the 2 balls =7C2

N(e)=None of the blue is =5C2

Probability of none of the balls draw Blue is = 5C2/7C2=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

1. P(X=1)+P(X=4)+P(X=3)+ P(X=5)+ P(X=6)+P(X=2)=(1x0.015)+(4x0.20)+(3x0.65)+(5x0.005)+(6x0.01)+(2x0.120)

=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**

**MEAN**

1. Mean =

**Calculating Mean using Python**

import pandas as pd

data=pd.read\_csv("Q7.csv")

data[['Points','Score','Weigh']].mean()

Points 3.596563

Score 3.217250

Weigh 17.848750

**MEDIAN**

A)Step 1: Arrange the scores in numerical order

Step 2: Count how many scores you have

Step 3: Divide the total scores by 2

Step 4: If you have an odd number of total scores, round up to get the position of the median number.

Step5: If you have an even number of total scores, go to the number in that position and average it with the number in the next higher position to get the median.

**Calculating Median using Python**

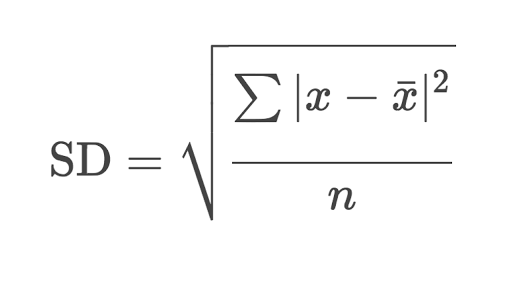
data[['Points','Score','Weigh']].median()

Points 3.695

Score 3.325

Weigh 17.710

**Standard Deviation:**



SD(): standard Deviation

n is the size of the data

x is the each value from the data

x bar means mean

**Calculating Standard Deviation using Python**

data[['Points','Score','Weigh']].std()

Points 0.534679

Score 0.978457

Weigh 1.786943

**Variance:**

Variance =

**Calculating Variance using Python**

data[['Points','Score','Weigh']].var()

Points 0.285881

Score 0.957379

Weigh 3.193166

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?

1. The no of weight of patients = 9

One of the patients is choosen at random is =1/9

Expect value of weight of patient = [108+110+123+134+135+145+167+187+199] = 145.188

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

**Cars speed and distance**

**Use Q9\_a.csv**

1. Skewness mathematically is Given E[(X-Mu/Sigma)]^3

The skewness of the speed = -0.117(no of observations present in left side of distribution)

The skewness of the distance=0.806(no of observations present right side of distribution)

The kurtosis of the speed & Dist is < 3(Flat peak)

**SP and Weight(WT)**

**Use Q9\_b.csv**

The skewness of the SP = 1.611(no of observations present in Right side of distribution)

The skewness of the WT = -0.614(no of observations present left side of distribution)

The kurtosis of the speed & Dist is < 3(Flat peak)

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



Outliear data

Upper extreme

Q3 = 75%

Q2= 50%

Q1 =25%

Lower Extreem



The Histogram of ChickWeight$Weight is describe the no of observations are present in right side of distribution

So the Skewness value is positive

The above boxplot having

Q1( Q1 is the lower Quartile = 25%

Q2( Q2 is the medium Quartile = 50%

Q3(Q3 is the Upper Quartile)=75%

The Upper maximum Quartile having no of outliers there

**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?

1. The Average person in our sample weight is (Mu)=200

The standard deviation of sample is (sigma)=30

94% of Confidence interval = 143.576

256. 42

98% of Confidence interval = 130.209

269.790

96% of Confidence interval =138.387

=261.612

**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?
3. Mean = 41.0

Median= 41

Mode= 41

Variance=25.52

Standard deviation=5.05

1. The no of students marks between 34 to 42

In the boxplot 2 outliers are find out in upper Quartile

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

1. The Skewness value=mean=median=0

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

1. The nature of the skewness is Right skewness

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

1. The nature of the skewness is Left skewness

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

1. The positive Kurtosis value indicates the sharp peak

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

1. The negative Kurtosis value indicates the flat peak

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



1)What can we say about the distribution of the data?

Let us Assume the above boxplot is describe the marks of the students

The box plot fallow the right skewness

2)What is nature of skewness of the data?

The nature of the skewness is left skewness (medium>mean)

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

IQR of the data approximately (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.

1. The above 2 boxplot’s the median Quartile Q2(50%) lies in same Origin at 262.5

The Boxplot 2 is more data comapare to Boxplot 1 data

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)
     1. Mean(Mu)=34.22
     2. Std(Sigma)=9.131
     3. X=38

P(MPG)<38= 0.660

P(MPG) >38=1-P(MPG)<38=0.339

* 1. P(MPG<40)

P(MPG<40) = 0.2633

* 1. P (20<MPG<50) = P(MPG)<50 - P(MPG)<20

=0.9580-0.0596

=0.8983

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. 1)The Mean Value of Cars of MPG is =34.422

2)The Median Value of Cars of MPG is =35.125

3)The Mode Value of Cars of MPG is =29.629936

4) MEAN,MEDIAN,MODE is not equal to Zero.

5)So the Cars of MPF data is not fallow the Normal Distribution

And the skewness fallow the left skewness

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

Dataset: wc-at.csv

1. The Mean Value of Waist & AT is (Waist) = 91.901

(AT) = 101.894

2)The Median Value of Waist &AT is(Waist) = 90.80

(AT) = 96.54

3)The Mode Value of Cars of Waist & AT is

(Waist) = 94.5,106.0,108.5 (This mode is called as Multi model)

(AT) =121.0,123.0(this mode is called as Bi-model)

1. 4) MEAN,MEDIAN,MODE is not equal to Zero.

The (Waist) and (AT) is not fallow the Normal Distribution

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

A) 90% of confidence interval = 1+CL/2=0.95

from scipy import stats

stats.norm.ppf (0.95) =1.64

94% of confidence interval = from scipy import stats

stats.norm.ppf (.97) =2.05

60% of confidence interval = from scipy import stats

stats.norm.ppf (0.8) =0.841

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

A) Sample size(n)=25

Degrees of freedom(df)= n-1= 25-1

=24

95% of confidence interval =1+cr/2

=1+0.95/2=0.975

from scipy import stats

stats.t.ppf(cl,df)

t scores of 95% confidence interval = 1.96

t scores of 96% confidence interval = 2.5

t scores of 99% confidence interval = 2.45

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

1. X =260

Mu =270

Sigma=90

Sample n = 18

Df = n-1

= 17

**Hypothesis**

Ho=Accept null Hypothesis

H1= Accept alternateHypothesis

Ho=Avg life of buib>=260

H1=Avg life of bulb<=260

**T –Statistics =** (X-Mu)/Sigma/sqtr of n

= (260-270)/(90)/sqrt of 18)

= (-10)/90\*Sqrt of 18

= -0.4714

**T-statistics value = -0.4714**

**Probability P – value = (0.3372)**

**Hypothesis test : P –value>0.05**

**Ho=Accept null hypothesis**

**P – value<0.05**

**H1=Accept alternate hypothesis**