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

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

Ans:- Probability of two heads =3/8 (HHT,HTH,THH)

Probability of one tail =7/8 (TTT,TTH,THT,THH,HTT,HTH,HHT)

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

1. Equal to 1

Ans:-Probability is zero

1. Less than or equal to 4

Ans:-6/36=1/6

1. Sum is divisible by 2 and 3

Ans:-18/36=1/2 and 12/36=1/3

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:- Probability=10/21

Total no.of balls=2+3+2=7

No of ways of drawing two balls (s)=2 balls out of 7

7c2=(7\*6)/(2\*1)=21

N = Event of drawing 2 balls, none of which is blue.

No of ways of drawing two balls (N)=2out of 5

5c2=(5\*4)/(2\*1)=10

P(E)=n(n)/n(s)=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\*0.120=3.090

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:-** :-

Points Score Weigh

Mean : 3.596563 3.21725 17.84875

Median : 3.695 3.325 17.71

Mode : 3.07 3.92 3.44 17.02 18.90

Variance : 0.285881 0.957379 3.193166

S.Deviation : 0.534679 0.978457 1.786943

Range : 2.17 3.91 8.4

**Points**  **Score**  **Weigh**

Min. :2.760 Min. :1.513 Min. :14.50

1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89

Median :3.695 Median :3.325 Median :17.7

Mean :3.597 Mean :3.217 Mean :17.85

3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90

Max. :4.930 Max. :5.424 Max. :22.90

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:- No. of patients =9

Probability of each patients weight is 1/9 =1/9\*108+1/9\*110+1/9\*123+1/9\*134+1/9\*135+1/9\*145+1/9\*167+1/9\*187+1/9199

=145.33

**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**

**Ans:** >skewness(Q9\_a$speed) >skewness(Q9\_a$dist)

-0.1139548 0.7824835

Negative skewed =left skew Positive means =Right skew

Here, Mean<Median Here,Mean>Median

>kurtosis(Q9\_a$speed) >kurtosis(Q9\_a$speed)

2.422853 3.248019

Positive kurtosis Positive kurtosis

So, it is sharp distribution So, it is sharp distribution

And also 0.248019 excess kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

>skewness(Q9\_b$SP) >skewness(Q9\_b$WT)

1.581454 -0.6033099

Positive skewness =Right skewed Negative skewness =Left skewed

Here,Mean>Median Here,Mean<Median

>kurtosis(Q9\_a$SP) >kurtosis(Q9\_a$WT)

5.723521 3.819466

Positive kurtosis Positive kurtosis

So, it is sharp distribution So, it is sharp distribution

Excess kurtosis

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



Ans:- Right skewed data, here Mean>Median

Nature of skewness is Positive.



Ans:- Boxplot gives information about 5-points.

1. Median

2. 1st Quartile

3. 3rd Quartile

4. Lower limit

5. Upper limit

Lower limit =1st Quartile-1.5(IQR)

Upper limit =3rd Quartile-1.5(IQR)

**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: The 94% CI is (198.7383,201.26167)

The 94% CI is (198.4393,201.56056)

The 94% CI is (198.6223,201.37769)

**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.

Ans: Mean = 41

Median = 40.5

Variance = 25.52941

Standard deviation = 5.052664

1. What can we say about the student marks?

Ans: Mean>Median So skewness is also positive

Kurtosis is also >3 so excess kurtosis

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

Ans: Perfectly symmetrically

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

Ans: Right skewed data

Positive skewness

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

Ans: :- Left skewed data

Negative skewness

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

Ans: Sharp Peak

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

Ans: Flat peak

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



What can we say about the distribution of the data?

Ans: Mean<Median (Left skewed data)

What is nature of skewness of the data?

Ans: Negative skewness

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

Ans: IQR = Upper Quartile-Lower Quartile

IQR =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.

Ans: Boxplot 1:-

IQR=275-250=25

Lower limit= 1ST Quartile-1.5(IQR)

Lower limit=250-1.5(25)=212.5

Upper limit=3rd Quartile-1.5(IQR)

Upper limit =275+1.5(25)=312.5  
 Boxplot 2:-

IQR=300-225=75

Lower limit= 1ST Quartile-1.5(IQR)

Lower limit=225-1.5(75)=112.5

Upper limit=3rd Quartile-1.5(IQR)

Upper limit =300+1.5(25)=412.5

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)

Ans: stats.norm.cdf(38,loc=mu,scale=sd)

0.6524058977429089

* 1. P(MPG<40)

Ans: 0.6524058977429089

* 1. P (20<MPG<50)

Ans: stats.norm.cdf(50,Car.MPG.mean(),Car.MPG.std())- stats.norm.cdf(20,Car.MPG.mean(),Car.MPG.std())

0.8988689170528976

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: No, it doesn’t follow Normal distribution because here mean<median. If mean=median, then it 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: No, here also mean is not equal to median. So it doesn’t follow normal distribution

For AT mean,median =(101.8940,96.54)

For Waist mean,median(91.901,90.8)

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

Ans: :- Z scores of 90% confidence interval =1.645

Z scores of 94% confidence interval =1.88

Z scores of 60% confidence interval =0.84

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

Ans: stats.t.ppf(CI,df)

t-score of 95% CI=1.710

t-score of 96% CI=2.1715

t-score of 95% CI=2.7969

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: T Statistical value =260-270/90/√18 = 0.471

P value=pt(0.471,17)= 0.3218