|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous  Data |
| Distance between two places | Continuous  Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete 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 |
| 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 | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

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

For   ={HHT,HTH,THH}

   =3/8

And the probability of getting one Tail is

  =3/8

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- When two Dice are rolled sample space is

   S= {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6) (2,1) (2,2) (2,3)  (2,4) (2,5) (2,6) (3,1) (3,2) (3,3) (3,4) (3,5) (3,6) (4,1) (4,2) (4,3) (4,4) (4,5) (4,6) (5,1) (5,2) (5,3) (5,4) (5,5) (5,6) (6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

Total number of possibility is =36

(a)-Sum is equal to 1

Ans-Zero (0)

(b)- sum is less that or equal to 4

  B= {(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)}

  N(B)=6

  P(B)=6/36=1/6.

(c)- Sum is divided by 2 and 3

    C={(1,1), (1,2), (1,3), (1,5), (2,1), (2,2) (2,4) (2,6) (3,1) (3,3) (3,5) (3,6) (4,2)(4,4) (4,5) (4,6) (5,1) (5,3) (5,4) (5,5) (6,2) (6,3) (6,4) (6,6)}

N(C)=24

P(C)=24/36

        =2/3

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- Given- Red Ball=2   Green Ball=3   Blue Ball=2

                Total number of Balls =7

      S=Two Ball Drawn Randomly Then,

       N(S)=7C2 =7!/2!5!

                        =21

    Let consider M = None of the balls drawn is blue,

     N(M)=5C2=5!/3!2!

                       =10

     Then P(M)=N(M)/N(S)

                      =10/21

The probability of none of the ball blue is 10/2.

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- Probability of expected no. of candies for randomly selected child=

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-**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Point** | **Score** | **Weigh** |
| **Median-** | **3.695** | **3.325** | **17.71** |
| **Mean-** | **3.596563** | **3.21725** | **17.84875** |
| **Mode-** | **3.92** | **3.44** | **17.02** |
| **Variance-** | **0.285881** | **0.957379** | **3.193166** |
| **std.Dev.-** | **0.534679** | **0.978457** | **1.786943** |
| **Range-** | **2.17** | **3.911** | **8.4** |

* Median of Weigh is greater then Score and Point, but median of point and score is approximately same.
* Mean of Weigh is greater then Score and Point, but mean of point and score is approximately same.
* Mode of Weigh is greater then Score and Point, but mode of point and score is approximately same.
* Variance of Weigh is greater then Score and Point, but Variance of point and score is approximately same.
* Std Deviation of Weigh is greater then Score and Point, but Std.deviatin of point and score is approximately same.
* Range is different for points, score and weigh.

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- Consider, where is weight of patient.

X=(108,110,123,134,145,167,187,199,)

P(X)=(1/9+1/9+1/9+1/9+1/9+1/9+1/9+1/9+1/9)

Sum of P(X)=9/9=1

We know that

E(X) = mean = X.P(X)

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

=145.33

Expected value of the weight of patient chosen at random is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.1139548 | 0.7824835 |
| Kurtosis | -0.6730924 | 0.1193971 |

Cars speed is negatively high skewed where as Distance is positively skewed.

The speed is mesokurtic where as Distance is less leptokurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
|  | SP | Weight |
| Skewness | 1.552258 | -0.5921721 |
| Kurtosis | 2.583072 | 0.7257402 |

SP is positively high skewness and Weight is negatively moderate skewed.

SP is mesokurtic where WT is Platykurtic.

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



* From the histogram is observed that histogram is positively high skewed and it is seen that is has some flatness so we can say that it is platykurtic.
* The boxplot is not symmetric because the whiskers on boat the sides of boxplot are not equals .and also some outlier on the upper extreme of boxplot ,the deviation from upper quartile to upper extreme is high .where as deviation is low from quartile to lower extreme.

**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-

n=2000

std dev(σ)=30

sample mean=300000

std error= =0.67

α=1-(confidence interval/100)=0.06

Degree of freedom(DF)=2000-1=1999

Critical Probability=1-(α/2)=0.97

We have std dev of sample so we apply t Distribution

Critical value= 1.881, error=critical value \*std error=1.26

Confidence interval94%

=(200+1.26, 200-1.26) = (201.26, 198.74)

FOR 98% confidence interval,

α= 1-(confidence level/100)=0.02,

Critical probability= 1-(0.02/2)=0.99,

DF=1999,

critical value at df= 2.33,

error= critical value\*std error= 1.56

Confidence interval of 94%= (200+1.56,200-1.56) = (201.56, 198.44)

96% confidence interval

std error= 0.67,

α= 1-(confidence level/100)=1-0.96=0.04,

Critical probability= 1-(α/2)=0.98.

by using t distribution ,

critical value=2.055

error= critical value\*std error=1.37,

96% CONFIDENCE INTERVAL =(200+1.37, 200-1.37)= (201.37, 198.62)

**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, Std.Dev-5.052664

1. What can we say about the student marks?

Ans- If we draw the graph of student marks then it observed that the graph is positively skewness and kurtosis is mesokurtic.

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

Ans- Mean and Median are equal it means symmetrical about mean.

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

Ans- If mean is greater then median then it is positively skewed.

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

Ans- If median is greater then mean then it is negatively skewed.

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

Ans- Positive kurtosis value indicates that no wider peak and it tail is at peak.

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

Ans- Negative kurtosis value indicates that thinner tail and wider pick.

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



What can we say about the distribution of the data?

Ans- The whisker on left and whisker on right on boxplot are not equal so the distribution is not symmetric.

What is nature of skewness of the data?

Ans- Boxplot is negatively skewed .Median is greater then mean.

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

Ans-The approx. and Interquartile range is 8.

Q19) Comment on the below Boxplot visualizations?



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

Ans-

* First boxplot is symmetrical that is whisker on boat side of first boxplot have same length.
* In second boxplot it is also a symmetrical because the whisker on the both side are equal.
* By calculating IQR of the both boxplot then the IQR will be different because they have the different data sets that is we can say that have different values of quartiles.

So conclusion is both boxplot are symmetrical but in nature they are different.

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-

1. P(MPG>38)=0.3475
2. P(MPG<40)= 0.7293

3) P (20<MPG<50)=0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans-

Yes, the data set 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-

Yes, the data set follows normal distribution

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

Ans- Z scores of 90% confidence interval

Qnorm(0.95)

[1] 1.644854

94% confidence interval

Qnorm(0.97)

[1] 1.880794

60% confidence interval

Qnorm(0.8)

[1] 0.841621

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

Ans-

**T score of 95% confidence interval**

1+0.95/2=0.975

Degree of freedom n-1=24

(0.975,24)

R code >qt(0.97,df=24) = 2.063899

**T score of 96% confidence interval**

1+0.96/2=0.98

(0.98,24)

R code > qt (0.98,df=24) = 2.171545

**T score of 99% confidence interval**

1+0.99/2=0.995

(0.995,24)

R code > qt (0.995,df=24) = 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

Ans-

mean (µ)=260 , (n)=18, x=270, std. dev.(s)=90

t stat formula:-

=

=0.47

Degree of freedom =n-1

Then df=18-1

=17

P score = (0.47, 17)

By using log table,

= 0.32 as 32%

R code pt(0.47, df=17) = 0.3221639