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

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

ANS=

Three coins are tossed,

P={(H,H,H),(H,H,T),(H,T,T),(T,T,T),(H,T,H),(T,T,H),(T,H,H),(T,H,T)}=8

TOTAL EVENTS=8

Probability getting two head one tail=

{(H,H,T),(H,T,H),(T,H,H)}=3

Interested events=3

Probability=no of events/no of total events=3/8=0.37

Probability two heads and one tail is 37%.

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 the dice is rolled***

***{(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)}***

***=36.***

***Sum is equal to 1***

***A={}=0***

B.

Sum is less than or equal to 4=

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

P=I.E/P.E

P(B)=6/36

P(B)=0.16

Probability having sum is equal to 4 is 16%

C.

Sum is divisible by 2 and 3=

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

=6

P=I.E/P.E

P(C)=6/36

P(C)=0.16

Probability having sum is divisible by 2 and 3 is 16%.

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=

Red ball - 2 green ball - 3 blue ball - 2

Total balls=7

S={2R+3G+2B} - 7BALLS

**When two balls drawn at randomly out of seven ,then probability of seven balls with drawn two is**

**nCr =n!/r!(n-r)!**

**-7C2**

**= (7!)/2!/(7-1)!**

**= (7\*6\*5!)/2!\*5!**

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

**=42/2**

= 21

T=21 Total events when randomly two are drawn

**Then, out of them they can’t contail blue balls**

**There are two balls having blue colour**

**It contain 5 balls only out of 7**

Using formulae of ncr..

**nCr =n!/r!(n-r)!**

nCr =5C2

=5!/2!(5-2)!

=5\*4\*3/2\*3

=5\*4/2\*1

=20/2

=10 \_\_\_(interested event)

Probability of none of the balls are drawn blue

P=I.E/T.E

P=10/21

P=0.47

PROBABLITY Of NONE OF THE BALLS DRAWN IS BLUE IS 47%.

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

Sol-

Probability of expected number of candies for a randomly selected child.

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=0.015+0.8+1.95+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.

**Using Q7.csv file:-**

**ANS=**

|  |  |  |  |
| --- | --- | --- | --- |
| **Q7.csv** | **Points** | **Score** | **Weigh** |
| **Mean** | **3.596** | **3.217** | **17.848** |
| **Median** | **3.695** | **3.325** | **17.71** |
| **Mode** | **3.92** | **3.44** | **17.02** |
| **Variance** | **0.285** | **0.957** | **3.193** |
| **Std.Deviation** | **0.534** | **0.978** | **1.786** |
| **Range** | **2.17** | **3.911** | **8.399** |

**Inferences-- Data set have outlier**

**Dataset have positive skewed**

**Points column are shown in right skewed**

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?

**SOLUTION=**

weight of patient

X=(108,110,123,134,135,145,167,187,199)

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

Mean =X.P(X)

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

= 145.33

Expected value of patient is 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS=**

|  |  |  |
| --- | --- | --- |
| **Q9\_a.csv** | **speed** | **distance(dist)** |
| **Skewness** | **-0.1175** | **0.8068** |
| **Kurtosis** | **-0.5089** | **0.4050** |

**Inferences\_**

**Data are positively skewed.**

**Speed are left skewed.**

**Distance are right skewed.**

**SP and Weight(WT)**

**ANS=**

|  |  |  |
| --- | --- | --- |
| **Q9\_b.csv** | **SP** | **WT** |
| **Skewness** | **1.6114** | **-0.6147** |
| **Kurtosis** | **2.9773** | **0.9502** |

**Inferences\_**

**Dataset are positively skewed.**

**SP is right skewed.**

**WT is right skewed.**

**Both have outlier.**

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



**ANS=**

**Inferences\_**

**Histogram is positively skewed.**

**Between 300 and 400 have outlier.**

**Histogram is positively high skewed**



ANS=

Inferences\_

In boxplot outlier on upper extreme.

Boxplot is right skewed.

**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=200

Std(dev)=30

Sample mean=3000000

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**

**Apply tdistribution**

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

**Confidence interval 94%=(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.
2. What can we say about the student marks?

ANS=

ANS =

**Mean** = 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

=738/18

**Mean** = 48

**Median** =34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56

=40+41/2

=81/2

**Median** =40.5

**Variance = (34-41)²+(36-41)²+ (36-41)²+(38-41)²+(38-41)²+(39-41)²+**

**(39-41)²+(40-41)²+(40-41)²+(41-41)²+(41-41)²+(41-41)²+**

**(41-41)²+(42-41)²+(42-41²)+(45-41)²+(49-41)²+(56-41)²/18**

**= (-7)²+(-5)²+(-5)²+(-3)²+(-3)²+(-2)²+(-2)²+(-1)²+(-1)²+(0)²+(0)²+**

**(0)²+(0)²+(1)²+(1)²+(4)²+(8)²+(15)²/18**

**= 49+25+25+9+9+4+4+1+1+0+0+0+0+1+1+16+64+225/18**

**= 434/18**

**Variance = 24.11**

**Standard Deviation(S.D)....**

**S.D = sqrt.σ**

**= sqrt.24.11**

**S.D = 4.9101**

**2- Student mark is ascending order.**

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

ANS=

**The nature of skewness is symmetric.**

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

ANS=

**The nature of skewness is right skewed.**

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

ANS=

**The nature of skewness is left skewed.**

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

ANS=

**Implies that the data has sharp peaks and wide tails.**

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

ANS=

**Implies that data have wider peak and thin tails.**

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



What can we say about the distribution of the data?

ANS=

**Normally distributed.**

What is nature of skewness of the data?

ANS=

**Left skewed.**

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

ANS=

**The Interquartile Range is 18-10. That 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 both the sides of first boxplot have same length.**

**In the second boxplot it is also a symmetrical because the whisker on the both side are equal.**

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)

c. P (20<MPG<50)

ANS=

A. **P (MPG >38 ) = 0.3475.**

**B. P (MPG <40) = 0.7293.**

**C. 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,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=

**Yes,it follows Normal Distribution.**

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

**ANS=**

**- For 90% C.I. Area = 0.95 then ,**

**Z-Score =stats.norm.ppf(.95)**

**Z-Score =1.6448**

**- For 94% Area = 0.97 then**

**Z-Score =stats.norm.ppf(.97)**

**Z-Score =1.8807**

**- For 60% Area = 0.80 then,**

**Z-Score =stats.norm.ppf(.80)**

**Z-Score =0.8416**

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

**ANS=**

**Degrees Of Freedom = 25 -1 = 24**

* **For 95% C.I. Area = 0.975 then ,**

**t-Score =stats.t.ppf(.975,df=24)**

**t-Score =2.0638**

* **For 96% Area = 0.98 then**

**t-Score =stats.t.ppf(.98,df=24)**

**t-Score =2.1715**

* **For 99% Area = 0.995 then**

**t-Score =stats.t.ppf(.995,df=24)**

**t-Score =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=**

**Df = 18 – 1 = 17**

**S=90**

**(Population mean) = 260**

**(sample mean) = 270**

**(Sample size) = 18**

**t = (270-260)/90 /sqrt(18)**

**t = 0.4714**

**p Value = 0.3221.**