|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discreate |
| Results of rolling a dice | Discreate |
| 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 | Categorical |
| Number of kids | Discreate |
| Number of tickets in Indian railways | Discreate |
| Number of times married | Discreate |
| Gender (Male or Female) | Categorical |

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 | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ordinal |
|  |  |

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

1. Each coin has two outcomes

Total no of outcomes=2\*2\*2=8

Two heads and one tail outcomes =(HHT,HTH,THH)=3

Probability=3/8=0.375

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
4. Total outcomes =6x6=36

Sum of two dice can never be 1.so probability =0

1. Possibe outcomes={(1,1),(1,2),(1,3),(2,1),(3,1),(2,2)}=6

Required probability=6/36=0.167

1. Total out come=36

Sum is divisible by(2 and3)

Out comes=(1,5)(2,4)(3,3)(5,1)(6,6)

Out comes=5

Probability is 5=5/36

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= total number ball are=(2+3+2)=7

Number of ball=7

Number of way in which 2balls can be drawn out of 7=7C2=21

N(S)=7C2n(S)=(7x6)/(2x1)n(s)=21

Number of red and green balls are 2+3=5

Number of way in which 2balls can be drawn from green and red balls are =5C2

=10

N (E)=5C2n(S)=(5x4)/(2x1)n(s)=10

N(E) =10

I.E P(E) = 10/21=0.48

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 =Expected number of candies=sum of (candies count x probability)=

1x0.015+4x0.20+3x0.5+5x0.005+6x0.01+2x0.120=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**

**A )** Given Df is having 4 columns i.e, Points, Score, Weight

Calculating the Mean, Median, Mode, Variance, Standard Deviation, and Range

For points

* Mean = df.Mean() - 3.5965625
* Median=df.Median()- 3.695
* Standard Deviation- df.std() – 0.53467
* Range = [2.76,4,93]
* Variance =0.285881

For score

* Mean =3.21725
* Median=3.325
* Standard Deviation=0.972845
* Variance=0.9573789
* Range=1.513,5.424

For Weight

* Mean= 17.8487500
* Median=17.71
* Standard Deviation =3.193166
* Range= 14.5,22.9
* Variance=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?

Probability of randomly one is mean in the case

Import statistics

List=[108, 110, 123, 134, 135, 145, 167, 187, 199]

Statistics.mean(list) = 145.33333334

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Skew of speed = -0.117510**

**Skew of distance = 0.806895**

**Kurtosis of speed = -0.508994**

**Kurtosis of distance =0.405053**

**SP and Weight(WT)**

**Use Q9\_b.cs**

**For skew**

* **SP = 1.611450**
* **WT = -0.614753**

**For kurtosis**

* **SP = 2.977329**
* **WT = 0.950291**

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



From this histogram we can see that:

1. We can see that mode is between 50-100
2. We can see that median is approx. 200
3. The histogram is right skewed



* The box plot is left skewed

There apporax 7 layers

**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. for 94% C.I is the average age of an adult male is between:(198.73798748893051,201.26201251106949)
2. for 96% C.I is the average age of an adult male is between:(198.62193448096605,20.37806551903395)
3. for 98% C.I is the average age of an adult male is between:(198.4390205765186,201.5609794234814)

**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.

Output :

Mean=41

Median=40.5

Variance=25.529411764705884

Standard deviation=5.05266382858645

2)What can we say about the student marks?

* Average marks obtained is 41
* The median is 40.5 i.e,50% of student scored below 40.5

And rest all scored above 40.5

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

A ) Normal distribution

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

1. Right Skewness

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

1. Left Skewness

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

1. distribution are heavy tails than the normal distribution

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

1. distribution are Lighter tails than normal distribution

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



What can we say about the distribution of the data?

1. The data range has from 1to19 and there is no outliers in data

What is nature of skewness of the data?

1. It is left skewed data

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

A)Q1=10 ,Q3=18

IQR=Q3-Q1

18-10=8 approximately it is 8

Q19) Comment on the below Boxplot visualizations?



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

1. Data 1 and data2 have same median value(262.5)
2. These two data are symmetrically distributed clearly seen from

Nature of the box plots

1. Data 1 is smaller than data 2 and it is wider

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)

A) P(MPG>38 = 0.3475939251582705

B) P(MPG<40 = 0.7293498762151616

C) P(20<MPG<50) = 0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* 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

* At does not follow normal distribution

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

A) Z- score of 90%C.I is=1.6448536269514722

Z- score of 94%C.I is =1.8807936081512509

Z-score of 60%C.I is = 0.8416212335729143

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

A) T-score od 95% is =2.0638985616280205

T-score od 96% is =2.1715446760080677

T-score od 99% is = 2.796939504772804

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. Mean = 260
  2. Samp\_mean =270
  3. N=18
  4. Samp\_std=90

T-score is:--0.47140

Corresponding probability is :-0.678325