|  |  |
| --- | --- |
| 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 | 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | interval |
| SAT Scores | interval |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained? Ans:3/8

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

1. Equal to 1=0/36=0
2. Less than or equal to 4=1/12
3. Sum is divisible by 2 and 3=29/36

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?=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: 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 \* 0.12

= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=       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**

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?

Sol: there are 9 patients

Probability of selecting each patient = 1/9

Ex  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

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

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

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed=Skew=** **-0.11395477012828319**

**Kurtosis=**-0.5771474239437371

**and distance = skew=0.7824835173114966**

**Kurtosis=**0.24801865717051808

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SP= skew=** 1.581453679442373

**WT= skew=** -0.6033099322115126

**SP=** **Kurtosis=** 2.7235214865269173

**WT=** **Kurtosis=** 0.8194658792266849

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



Ans: For Histogram

- Chick weight data is right skewed or positively skewed

- More than 50% Chick Weight is between 50 to 150.

-Most of the chick weight is between 50 to 100.

For Boxplot

-The data is right Skewed

-There are outliers at upper side

**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?

Sol:

Given: n=2000 mean =200 SD=30

Zvalue

for 94%=1.88

for 98%=2.33

for 96%=2.06 0.67

CI=Z(/)

CI for 94%=[200+1.88\*(30/)], [200-1.88\*(30/)]= (201.26,198.74)

CI for 98%=[200+2.33\*(30/)], [200-2.33\*(30/)]=(201.561,198.4381)

CI for 96%=[200+2.06\*(30/)], [200-2.06\*(30/)]=(201.38,198.62)

**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 deviatioA

Mean=41, Median=40.5, SD=4.91, Variance=24.11

1. What can we say about the student marks?

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

Ans=Zero Skewness

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

Ans=Positive Skewness

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

Ans=Negative Skewness

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

Ans: distribution is peaked and has thick tails

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

Ans: distribution is flat and has thin tails

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



What can we say about the distribution of the data?

Ans: The data is not normally distributed.

What is nature of skewness of the data?

Ans:Nature of Skewness is Positive

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

Ans:  **IQR** = Q3 – Q1

Q3=Upper quartile=18 and Q1=Lower Quartile=10  
  
So,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: 1.Both the Boxplot are Normally Distributed

2. Neither Boxplot shows any Outlier

3. Short box i.e Boxplot1 mean its data points consistently hover around the center values. Whereas Taller box i.e. Boxplot 2 having more variable data.

3.IQR for both Boxplot is different

IQR1=Q3-Q1=275-250=25

IQR2=Q3-Q1=312.5-225=87.5

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) =0.346
  2. P(MPG<40)=0.730

c. P (20<MPG<50)=0.900

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv=Ans: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:Dose not follow Normal Distribution

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

Ans: Z Score for 90 % Confidence Interval: 0.8289438736915182

Z Score for 94 % Confidence Interval: 0.8339767539364704

Z Score for 60 % Confidence Interval: 0.7881446014166034

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

Ans: T Score for 95 % Confidence Interval: 2.0638985616280205

T Score for 96 % Confidence Interval: 2.1715446760080677

T Score for 99 % Confidence Interval: 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

Ans: Probability for 18 bulbs with avrage life 260 days: 0.828384478201844