|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | **Discrete** |
| Results of rolling a dice | **Discrete** |
| Weight of a person | **Continous** |
| Weight of Gold | **Continous** |
| Distance between two places | **Continous** |
| Length of a leaf | **Continous** |
| Dog's weight | **Continous** |
| 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 | **Discrete - Nominal** |
| High School Class Ranking | **Discrete -Nominal** |
| Celsius Temperature | **Continous -Interval** |
| Weight | **Continous -Ratio** |
| Hair Color | **Discrete -Ratio** |
| Socioeconomic Status | **Continous -Interval** |
| Fahrenheit Temperature | **Continous -Ratio** |
| Height | **Continous - Ratio** |
| Type of living accommodation | **Discrete -Ordinal** |
| Level of Agreement | **Discrete -Interval** |
| IQ(Intelligence Scale) | **Discrete -Interval** |
| Sales Figures | **Discrete -Interval** |
| Blood Group | **Discrete - Ratio** |
| Time Of Day | **Continous -Interval** |
| Time on a Clock with Hands | **Continous - Interval** |
| Number of Children | **Discrete -Interval** |
| Religious Preference | **Discrete -Ratio** |
| Barometer Pressure | **Discrete - Interval** |
| SAT Scores | **Discrete -Ratio** |
| Years of Education | **Discrete -Nominal** |

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

ANS:- No. of sample (S)={H H H,H H T,H T H,H T T,T H H,T H T,T T H,T T T}

n(S)=8

The probability that two heads and one tail are obtained

No. of event(E)={H T H,T H H,H H T}

n(E)=3

P=n(E)/n(S)

3/8 (ans) .

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

ANS:-One dice rolled then possibility are {1,2,3,4,5,6}

two dice rolled then possibility are 6\*6=36

No. of sample space (S)=36 ,n(S)=36

1. Equal to 1

n(E)=0

n(S)=36

P(E)=0/36

=0

1. Less than or equal to 4

E={(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

n(E)=6

n(S)=36

P(E)=6/36

=1/6

1. Sum is divisible by 2 and 3

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

n(E)=6

n(S)=36

P(E)= 6/36

=1/6

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 no. of balls = (2+3+5 ) = 7

S = sample space

n(S) = no . Drawing the 2 balls out of 7

7C2 = 7\*6/2\*1 = 42/ 2 = 21

E = Event of the 2 balls none of the blue

n(E) = no . Of drawing 2 balls out of the (2+3) balls

5C2 = 5\*4/2\*1 = 20/ 2 = 10

P(E) = n(E) *n(S) = 10*/ 21 (ans ) .

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 no .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+0.025+0.06+0.24

= 3.09 (ans) .

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?

ANS :the weights (X) of the patients at a clinic (in pounds) are 108,110,123,134,135,145,167,187,199 .

one of the patients is chosen at random probability value

There are 9 patients

probabilities of each points = 1/9

E(x1)= 108, p(x1) = 1/9

E(x2) = 110, p(x2)=1/9

E(x3)= 123,p (x3)= 1/9

E(x4)= 134,p(x4)= 1/9

E(x5)= 135,p(x5)=1/9

E(x6)=145,p(x6)= 1/9

E(x7)=167,p(x7)= 1/9

E(x8)=187,p(x8)=1/9

E(x9)=199,p(x9)=1/9

Expected value =sum(p(x).E(x)

=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\*(1308)

=145.333 (ans) .

**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**

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



Ans : Right side skewed or positively skewked .

The data point are concentred in the range 50-100 with frequency 200 and least range of the weight is 400 .



Ans :median is less than mean and outlier on the upper side of box plot and there is less data points between bottom point.

The interface for this boxplot is positively 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?

**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?

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

Ans: symmetrical /Normalized skewness

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

Ans :right skewed /positively skewed

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

Ans :left skewed/ negatively skewed

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

Ans :The positive kurtosis value indicates distribution is peaked and thick tails .the data is narmally distribution and kurtosis value is 0 .

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

Ans :The negative kurtosis value indicates that a distribution is flat and has thin tails then the normal distribution .

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



What can we say about the distribution of the data?

ANS: The data is distributed in de -assignmemed format .

What is nature of skewness of the data?

ANS: left side skewed

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

ANS : Q3-Q1

=18-10

=8 is IQR .

Q19) Comment on the below Boxplot visualizations?



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

Ans : from the distribution of data for Boxplot 1 with respect to Boxplot 2

by observing both of the plots whiskers level is high in boxplot 2.

Mean and Median are equal .Hence with distribution is symmetrical .

The box plot 1 design with range 3 ,and the second one range is 1.5 .

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)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

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

ANS:- for 90% confidence interval

we have the significance level at 5%(two tailed)

5%= 0.05

z = 0.05 from the z-table wii be

z=1.645

for 94% confidence interval

we have the significance level at 3% (two tailed)

3% = 0.03

z= 0.03 from the z-table will be

z=1.880

for 60% confidence interval

we have the significance level is 20%(two tailed)

20%=0.20

z= 0.2 from table will be

z=0.253 (ans) .

|  |
| --- |
|  |

|  |
| --- |
|  |

|  |
| --- |
|  |

|  |
| --- |
|  |

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

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