|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete-Countable |
| Results of rolling a dice | Discrete-Nominal |
| Weight of a person | Continues-Interval |
| Weight of Gold | Continues-Interval |
| Distance between two places | Continues-Interval |
| Length of a leaf | Continues-Interval |
| Dog's weight | Continues-Interval |
| Blue Color | Discrete-Nominal |
| Number of kids | Discrete-Countable |
| Number of tickets in Indian railways | Discrete-Countable |
| Number of times married | Discrete-Countable |
| Gender (Male or Female) | Discrete-Nominal |

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

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

Answer:-

The possible outcomes if 3 coins are tossed

HHH, **HHT**, **HTH**, HTT, **THH**, THT, TTT, TTH

The possible outcomes that 2 heads and 1tail obtained are 3 as highlighted above

P(x)= 3/8 = 0.375 =37.5%

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

Answer:-

a)prob (x=0) = 0/36

b)prob(x<=4) = 6/36 = 0.166 = 16.6%

c)prob(x=6) + prob(x=12) = (5+1)/36 = 0.166 = 16.6%

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?

Answer:-

Total possible ways that 2 balls at random drawn from 7 balls is 7C2 ways =21

Total possible ways that 2 ball picked at random that is red/green is 5C2 ways = 10

Probability that two balls picked at random from those balls that is not blue is

= 10/21 = 0.4761 = 47.6%

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

Answer:-The expected number of candies for a randomly selected student is

 1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

= 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**

Answer:-

Mean=[ 3.596563 , 3.217250 , 17.848750], Median=[ 3.695 , 3.325 , 17.710], Mode=[[3.07,3.92] , 3.44 , [17.02,18.90]], Variance=[ 0.285881, 0.957379, 3.193166], Standard Deviation=[ 0.534679 , 0.978457 , 1.786943], and Range=[2.17 , 3.911 , 8.400]

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?

Answer:-

Expected Values are calculated as follows

 =  (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

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer:-

* The skewness value for car speed is **-0.11750** so it is said to be negatively skewed . Kurtosis value for car speed is **-0.5089** so it is said to be there is no peakness in the data
* The skewness value for distance is **0.80689** so it is said to be positively skewed . Kurtosis value for distance is **0.40505** so it is said to be there is peakness in the data

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answer:-

* The skewness value for SP is **1.6114** so it is said to be positively skewed

Kurtosis value for SP is **2.9773** so it is said to be lepto kurtosis

* The skewness value for WT is **-0.6147** so it is said to be negatively skewed

Kurtosis value for WT is **0.9502** so it is said to be lepto kurtosis

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



The histogram is positively skewed means most of the chicks weight between 0 to 200

The frequency of the chicks having weight 50 to 100 is high i.e 200 and the chicks having weight above 350 are not eligible.



From this boxplot we can identify many outliers are present in the data and we can observe that will be positively skewed because 50% of the data is present in the bottom of boxplot

**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?

**A)**The confidence interval for 94%,96%,98% is (198.62, 201.37)

**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.

Mean=41.000000, median=40.5, variance=25.529412, standard deviation= 5.052664

What can we say about the student marks?

We can say that majority of the students scored around 41marks with a standard deviation of 5.05

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

If mean= median =mode then the nature of the skewness is symmetric

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

1. positively skewed

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

1. Negatively skewed

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

A)It should have peakness of the data

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

A)It doesn’t have any peakness

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



What can we say about the distribution of the data?

A)50% of the data is lies between 10 to 18

What is nature of skewness of the data?

A)Negatively skewed

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

1. 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.

1. we observe that the spread of the data 1 is very less as compared to 2

Upper whisker and lower whisker is very less for 1 compared to 2

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

A)P(MPG>38) =0.3475 = 34.75%

B)P(MPG<40) = 0.6930 = 69.3%

C) P(20<MPG<50) = 0.8988 = 89.88%

Q 21) Check whether the data follows normal distribution

a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:-

MGP of cars follow normal distribution according to Shapiro test

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

Dataset: wc-at.csv

Answer:-

Waist Circumference and AT does not follow normal distribution according to the Shapiro test

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

90%-->1.6449

94%-->1.8808

60%-->0.8416

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

95%--> 2.0639

96%-->2.1715

99%-->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

