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

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

Total no. of outcomes when 3 coins are tossed = 8

Obtained outcomes of 2head and 1 tail =HHT,THH,HTH

Probability = obtained outcome/Total outcome = 3/8.

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: a)Total outcome = 6\*6=36

Probability = Favourable outcome/Total outcome=1/36

b) Less than or equal to 4

Total outcome =6\*6=36

Possible outcome: (1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

Probability= No. of favourable outcome to get sum less than or equal to 4/Total outcome=6/36=1/6

1. Sum is divisible by 2 and 3

The only possible sum divisible by 2 and 3 =6

Probability of getting sum of 6 is (3,3)

So,probability of getting sum that is divisible by 2 and 3 is 1/36.

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: {R,R,G,G,G,B,B}

Probability of 1st ball not blue=5/7

Probability of 2nd ball not blue =4/6

The probability that none of the balls drawn is blue= 5/7\*4/6=20/42=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: Let candies count be X, Probability of candies count be p(x)

Expected no. of candies = Sum(X\*p(x))

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

=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.

**Use Q7.csv file**

**Ans: Points: mean=3.596563 ,median= 3.695 ,mode=3.92 , Varience=0.2858814, Std deviation=0.5346787**

**Score: mean= 3.21725 ,median= 3.325 ,mode= 3.44 , Varience= 0.957379 , Std deviation=0.9784574**

**Weight: mean= 17.84875 ,median= 17.71 ,mode= 17.02 , Varience= 3.193166 , Std deviation=16943.78**

**Inference: mean values for point and score are close to each other**

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: Sum wt. of patients = 1308

Total no. of patients = 9

Expected Value of the Weight of that patient=Sum of wt.of patient / Total no. of patients

=1308.9/9

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness for speed is -0.11395

Here the speed is -ve which means that the graph is left skewed i.e the length of the left tail is longer.

Kurtosis for speed is 2.42285

Here the curve is peaked.

Skewness for distance is 0.7824

Here the value of skewness is +ve which means that the graph is right skewed i.e the length of the right tail is longer.

Kurtosis for speed is 3.2480

Here the curve is peaked. More than normal distribution.

**SP and Weight(WT)**

**Use Q9\_b.csv:** Skewness for SP is 1.581454

Here the value of skewness is +ve which means that the graph is right skewed i.e the length of the right tail is longer.

Kurtosis for SP is 2.9773

Here the curve is peaked, More than normal distribution as value of normally distributed data is 0.

Skewness for WT is -0.60331

Here the value of skewness is -ve which means that the graph is left skewed i.e the length of the left tail is longer.

Kurtosis for WT is 0.950291

Here the curve is slightly peaked, a little more than the normal distribution.

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

Here we can see that the major Chick weights fall in the category of 50-100g(measures in x) as the maximum which is 200.The minimum weights have frequency if less than or equal to 5. The plot is Right skewed which show that there is lesser concentration of chick weights in the 300-400gram category .The expected value should be above 46.45.



Ans: Median is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point

**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: 1.- CI = 94% (μ0 - 1. 04 < x < μ0 +1. 04)

2.- CI = 98% (μ0 - 2.05 < x < μ0 +2.05)

3.- CI = 96% (μ0 - 1.75< x < μ0 + 1.75)

Sample size n = 3000000

Sample mean x = 200

Standard deviation s = 30

From z-table values of z(c):

CI 94 % Confidential level α = 6 % α = 0,06 z(c) = 1.55

CI 98 % Confidential level α = 2 % α = 0,02 z(c) = 2.05

CI 96 % Confidential level α = 4 % α = 0,04 z(c) = 1.75

MOE = z(c) \* σ/√n

1.-MOE = 1.55\* 30 / √2000 MOE = 1.04

2.-MOE = 2.05\*30/√2000 MOE = 1.38

3.-MOE = 1.75\*30/√2000 MOE = 1.17

Then CI

1. CI = 94% (μ0 - MOE < x < μ0 - MOE) CI = (μ0 – 1.04 < x < μ0 +1.04)

2. CI = 98 % CI = (μ0 – 2.05 < x < μ0 + 2.05)

3. CI = 96 % CI = (μ0 – 1.75 < x < μ0 + 1.75)

**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.

Ans: Mean = 41

  Median = 40.5

 Variance = 5.05

Standard Deviation = 25.529

1. What can we say about the student marks?

Ans: This class contains students that have average performance. Most of the students in the class are having an average percentage of 65 and there are only a few students securing value above 90%.

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

Ans: Their will be a symmetrical skewness as the data is normally distributed on both side, when mean and median of data are equal.

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

Ans: It will show us positive skewness when mean is greater than median as outliers are present on right side.

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

Ans: It will show us negative skewness when mean is less than median as outliers are present on left side.

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

Ans: Positive kurtosis show sharp peak have more extreme values in the distribution than normal distribution

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

Ans: The negative kurtosis shows flat top and have fewer extreme values in the distribution than normal distribution

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



What can we say about the distribution of the data?

Ans: The boxplot shows median= 15.2

What is nature of skewness of the data?

Ans: Negative skewness

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

Ans: Q1=10 , Q3= 18

IQR= Q3-Q1 = 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: Median lines of both of the box plots lie within and overlap between two boxes. The Short box mean their data points spread consistently over around the center values. And Taller box shows more variable data and both boxes have no outliers.

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: Probability of (MPG>38) = 0.4074074074074074

Probability of (MPG<40) = 0.7530864197530864

Probability of (20<MPG>50) = 0.8518518518518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: The data does not follow 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: The data does not follow normal distribution

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

Ans: The Z scores of 90% confidence interval = 1.645

The Z scores of 94% confidence interval = 1.8807

The Z scores of 60% confidence interval = 0.85

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

Ans: The T scores of 95% confidence interval = 2.064

The T scores of 96% confidence interval = 2.085

The T scores of 99% confidence interval = 2.797

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: The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 0.471