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

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

Solution: S = {HHH, TTT, HTT, THT, TTH, THH, HTH, HHT}

n(S) = 8

Let A be event that contain two heads and one tail

A = {THH, HTH, HHT}

n(A) = 3

P(A) = n(A) / n(S)

= 3/8 = **0.375**

**Therefore, the probability of getting two heads and one tail is 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

Ans: n(S) = 36

a) Let A be the event that probability of getting sum is 1

A = { }

n(A)= 0

P(A) = n(A) / n(S)

= 0

**Therefore, the probability of getting sum one is 0.**

b) Let B be the event that probability of getting sum is less than or equal to 4

B = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

n(B)= 6

P(B) = n(B) / n(S)

= =

**Therefore, the probability of getting sum less than or equal to 4 is 1/6.**

c) Let C be the event that probability of getting sum is divisible by 2 and 3

C = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

n(C)= 6

P(C) = n(C) / n(S)

= =

**Therefore, the probability of getting sum divisible by 2 and 3is 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 balls = 2 + 3+ 2 = 7 balls

n(S) = number of ways of drawing two balls out of 7 balls

= = 21

Let A be the event that none of them drawn is blue ball

n(A) = = 10

P(A) = n(A) / n(S)

=

**Therefore, the probability that none of balls is blue.**

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 value E(x) =

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

= 3.09

**Therefore, the Expected number of candies for a randomly selected child is 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.

Ans:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard  deviation | Range |
| Points | 3.60 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.32 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.4 |

For Points, mean< median so it’s left skewed. Average points are 3.60.

For Score, mean< median so it’s left skewed.

For Weigh, mean> median so it’s right skewed.

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:**

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

= 1308/9

= 145.33 pounds

**Therefore, the average value of the weight of that patient is 145.33 pounds.**

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

Cars speed and distance

Use Q9\_a.csv

**Ans:** (Values are showing in jupyter notebook)

Skewness:

|  |  |
| --- | --- |
| Speed | -0.117510 |
| Distance | 0.806895 |

Here skewness of speed < 0 so distribution is left skewed and skewness of distance is greater than 0 so distribution is right skewed.

Kurtosis:

|  |  |
| --- | --- |
| Speed | -0.508994 |
| Distance | 0.405053 |

Here kurtosis for car speed is less than 0 so its platy kurtic and kurtosis for distance is grater than 0 so its lepto kurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:** (Values are showing in jupyter notebook)

Skewness:

|  |  |
| --- | --- |
| SP | 1.611450 |
| WT | -0.614753 |

Here skewness of SP > 0 so distribution is right skewed and skewness of Weight is less than 0 so distribution is left skewed.

Kurtosis:

|  |  |
| --- | --- |
| SP | 2.977329 |
| WT | 0.950291 |

Here kurtosis for SP is greater than 0 so its lepto kurtic and kurtosis for WT is less than 0 so its platy kurtic.

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



Here from above Histogram of ChickWeight$weight I observed that Distribution is right skewed. Maximum frquency for ChickWeight$weight occurs in interval [50, 100] which is 200. And outliers lies in [350, 400].



Here from above boxplot, I observed that distribution is right skewed as median is near to lower quartile. There are many outliers above upper extreme.

**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:** (Values are showing in jupyter notebook)

For confidence interval 94%, range is (143.54415570565965, 256.45584429434035)

For confidence interval 98%, range is (130.15355671679083, 269.84644328320917)

For confidence interval 96%, range is (138.34730111522666, 261.6526988847733)

**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?

Ans: Mean of score is 41 that’s means most of the students got the 41 marks.

Median for score is 40.5 which near to average value.

Variance of score is 24.11 and standard deviation is 4.91

Mean is greater than median so its right skewed which means students got very less score in the tests. (Values are showing in jupyter notebook)

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

**Ans:** The nature of skewness when mean = median is symmetric.

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

**Ans:** The nature of skewness when mean > median is positive skewed.

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

**Ans:** The nature of skewness when mean < median is negative skewed.

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

**Ans:** It indicates that distribution is peaked and possesses thick tails. Its called as lepto kurtic.

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

**Ans:** It indicates that distribution has lighter tail than normal distribution. Its also called as platy Kurtic.

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



What can we say about the distribution of the data?

**Ans:** Median of the data is 15.5 approximately. Q1 is 10 and Q3 is 18.Distribution of data is negative skewed. Outlier lies in left side of data.

What is nature of skewness of the data?

**Ans:** The nature of skewness is negative skewed.

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

**Ans:** IQR = upper quartile(Q3) – lower quartile(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:** I observed from above boxplots, there are no outliers for both boxplots. Also, both boxplots having same median value which is approximately 264. Graphs are normally distributed so there is no skewed occurs.

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) = **0.3476**
  2. P(MPG<40) = **0.7293**
  3. P (20<MPG<50) = **0.8989**

(Values are showing in jupyter notebook)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

**Ans :** Its not a normal distribution as mean is not equal to median also in histogram, its clearly shows that its not a normal distribution.

(Values are showing in jupyter notebook)

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

Dataset: wc-at.csv

**Ans :** (Values are showing in jupyter notebook)

For Adipose Tissue (AT), It’s not a normal distribution. It’s clearly shows in density plot as well as histogram. Also, mean = 91.08 which is not equal to median.

For Waist Circumference (Waist), Its not a normal distribution as its right skewed. Also, mean is not equal to median.

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

**Ans:** The Z score of 90% confidence interval is 1.6449

TheZ score of 94% confidence interval is 1.8808

TheZ score of 60% confidence interval is 0.5244

(Values are showing in jupyter notebook)

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

**Ans:** The t score of 95% confidence interval is 2.06

Thet score of 96% confidence interval is 2.17

Thet score of 99% confidence interval is 2.8

(Values are showing in jupyter notebook)

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

**Ans:** n = 18, = 260, s = 90, = 270

(Calculation in jupyter notebook)

Therefore, probability for average life of no more than 260 days is 0.32.