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

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

**Answer:**

Total no. outcomes: **{HHH, HHT, HTT, HTH, TTT, TTH, THH} THT)**

Possible outcomes: **{THH, HTH, HHT}**

**Probability that two heads and one tail obtained: 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:

**Answer:**

**a)**

**Description:**

Since the minimum sum of two dice rolled simultaneously are **1+1=2.** There is no chance of getting the **“sum of rolling two dice simultaneously and getting 1 is not possible”.**

Probability that the sum equal to 1: **Zero**

**b)**

**Description:**

The probable outcome of getting sum less than or equal to 4:

(**1,1), (1,2), (1,3), (2,1), (2,2), (3,1)**

Probability = no. Possible outcome/total no. Outcome

No. Possible outcome = 6

Total no. Outcome = 36

Probability of getting sum less than or equal to 4: 6/36

=1 / 6

=0.166666667

**Probability of getting sum less than or equal to 4: 0.166666667**

**c)**

**Description:**

Possible outcomes of sum rolling two dice are: **2 - 36**

Among which sum which is divisible by **2** and **3** are **6**

Possible outcomes are: **(1,5), (2,4), (3,3), (4,2), (5,1), (6,6)**

No. Possible outcomes: **6**

Total no. Outcome: **36**

**Probability: 6 / 36**

**=1 / 6**

**=** 0.166666667

**Probability of getting sum divisible by 2 and 3: 0.166666667**

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 no. Balls: (**2 red + 3 green + 2 blue) = 7**

Let S be the sample space.

Then, n(S) = number of ways of drawing 2 balls out of 7 = 7C2

= (7 \* 6) / (2 \* 1)

= **21**

Let E be the event of drawing 2 balls, none of which is blue.

Therefore, n(E) = number of ways drawing 2 balls out of (**2 red and 3**  **green) = (**5\*4)/(2\*1)

=10

**P(E) = n(E)/n(S)**

**=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

**Description:**

Expected number of candies for a randomly selected child

=1\*0.015 + 4\*0.02 + 3\*0.065 + 5\*0.005 + 6\*0.01 + 2\*0.12

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

=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 Comment about the values/ Draw some inferences.

In python file

**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?

**Answer:**

**A)** To calculate the expected value of the weight for a randomly chosen patient from the given dataset, we need to sum up all the weights and divide the sum by the total number of patients

Expected value = X\*(probability of x)

The weights of the patient are probability of (x) is 108,110,123,135,145,167,187,199

The expected value E =

X = sum of observation/ No of observations

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

**OUTPUT = 145.33**

The expected value of the weight for a randomly chosen patient is approximately is 145.33 pounds.

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

**Car's speed and distance**

**Use Q9\_a.csv**

In python file

**SP and Weight (WT)**

**Use Q9\_b.csv**

In python file

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



1. It is the graph of the histogram between ChickWeight$weight & frequency

Int the graph we can see the graph is at high peak towards “Right side” so we can get the positive value

When the mean is greater that the median, the distribution is positively skewed (or) Right-skewed, the tail is in the right side of the histogram



INTERFACE: - The box plot shows that data is “negative skewed data”. Since the median of the data is near to the minimum value. And for these data, there are outliers also.

**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.

In Python file

1. What can we say about the student marks?

We can say that highest marks obtained by the student is 53 and the lowest marks obtained is 34.

The total average of the student scores is 41

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

If the mean and median are equal, then it is said to be Normal Distribution, I.e. left side and right side will be symmetric to each other.

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

If the mean is greater than median, ‘then it is said to be positively skewed’, I.e., tail will be more in the right side of the mean as compared to the left of side of the mean.

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

If the median is greater than mean, ‘then it is said to be negatively skewed’, I.e., tail will be more in the left of the mean as compared to the left side of the mean.

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

A positive kurtosis value indicates that the distribution is peaked and possesses thick tails. Leptokurtic distributions have positive kurtosis value. A Leptokurtic distribution will have higher peak and lower than normal distribution.

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

A negative kurtosis value indicates that the distribution is flatter than a normal distribution. Playkurtic distributions have negative kurtosis values. A playkurtic distribution will have lower peak and shorter than normal distribution.

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



What can we say about the distribution of data?

The data shown in the box is left skewed. Since the median is closer to the top of the box and the whisker is shorter on the upper end of the box.

What is the nature of skewness of the data?

It is a negatively skewed or left-skewed data

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

IQR= Q3-Q1

Q1=1

Q3=19

IQR= 18

Q19) Comment on the Boxplot visualizations below?



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

1. Interference:

In the boxplot (1) the data moves towards the “left skewed data”

The variance is also “Less”, compared to the boxplot (2)

1. In the boxplot (2) the data move towards the normal “Normal data”

The data is plotted at the “median”. The variance is also “very high”,

compared to the boxplot(1)

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)

In Python File

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: in python file

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

Dataset: wc-at.csv

Ans: In python file

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval A) We need to calculate the Z-score for 90%confidence interval from the Z-score table

90% confidence interval =1.645 (from Z-score table we will get these values)

60% confidence interval =0.842 (from Z-score table we will get these values)

94% confidence interval for that we need to find Area

Area= 1+ confidence interval/2 =(1 + 94 /100) /2

= 1 + 0.94 / 2

=0.9700

We need to find the 0.9700 in the Z- score table

=1.8+0.08

OUTPUT = 1.88 (approximately)

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

A) Sample size(n)= 25

Degree of freedom =sample size -1

=24

From the t- score table ,we need to find the 95% confidence interval

95 % confidence interval =2.064

99 % confidence interval=2.797

96 % confidence interval=2.171

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 à pt(tscore,df)

df à degrees of freedom

Ans: In python file