|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | Interval |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Ratio |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans:

Total outcomes = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

A = {HHH, HHT, HTH, THH}

B = {THH, THT, TTH, TTT}

P(A) = P(HHH)+P(HHT)+P(HTH)+P(THH)

= 1/8+1/8+1/8+1/8 = 4/8 = ½

P(B) = P(THH)+P(THT)+P(TTH)+P(TTT)

= 1/8+1/8+1/8+1/8 = 4/8 = ½

P(AᴖB) = P(THH) = 1/8 (ans)

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:

Total outcomes = 36

1. P(x1+x2 = 1)

As the no of favourable outcome is 0

Then, P = 0/36 = 0 (ans)

1. P(x1+x2 <= 4)

Here the no of favourable outcome is 6

Then, P = 6/36 = 1/6 (ans)

1. Here Only the no 6 is divisible by 2 & 3

So, P(x1+x2 = 6)

The no of favourable outcome is 5

Then, P = 5/36 (ans)

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+2 = 7

Total no of outcomes = 7C2 = 7! / 2! = 7\*6 / 2\*1 = 42/2 = 21

As per given condition no blue balls should be drawn, so we have to minus no of blue balls from total no of balls, which is 7-2 = 5.

Total Favourable outcomes = 5C2 = 10

So the Probability = 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 number 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 = 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**

**Ans.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596 | 3.217 | 17.848 |
| **Median** | 3.695 | 3.325 | 17.710 |
| **Mode** | 3.07, 3.92 | 3.44 | 17.02, 18.90 |
| **Variance** | 0.285 | 0.957 | 3.193 |
| **Standard Deviation** | 0.534 | 0.978 | 1.786 |
| **Range** | 2.170 | 3.911 | 8.400 |

Note: In this data mean value are very closer for both “Points” and “Score”.

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.

X = {108, 110, 123, 134, 135, 145, 167, 187, 199}

P(X) = 1/9 for all.

So, Expected value = (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)

= 145.333 (ans)

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

Skewness for Speed = -0.11395477, So it’s negatively skewed distribution.

Skewness for Distance = 0.78248352, So It’s positively skewed distribution.

Kurtosis for Speed = -0.57714742, So it’s negative kurtosis.

Kurtosis for Distance = 0.24801866, So it’s positive kurtosis.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

Skewness for SP = 1.58145368, So it’s positively skewed distribution.

Skewness for WT = -0.60330993, So It’s negatively skewed distribution.

Kurtosis for SP = 2.72352149, So it’s positive kurtosis.

Kurtosis for WT = 0.81946588, So it’s positive kurtosis.

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



Ans:

* The most range of weight is found in between 50-100 with frequency 200 and least the least range is 350-400 with frequency around 0-10.
* Here we can notice a long tail towards right, so it’s positively skewed distribution.



Ans:

* The data is positively skewed distributed.
* There are some outliers on the upper side.
* In this data the median is less than the mean.
* The Upper quartile is larger than the lower quartile.

**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:**

Mean = 200, Std = 30, n = 2000

Degrees of freedom = 2000-1 = 1999

* CI for 94% = 1.882

So the range is [198.738, 201.262]

* CI for 98% = 2.328

So the range is [198.439, 201.561]

* CI for 96% = 2.055

So the range is [198.622, 201.378]

**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:**

1. Mean = 41

Median = 40.5

Variance = 24.11

Standard deviation = 4.91

1. Maximum students are scores marks between 38-42 and if we calculate skewness then it’s positives skewed distributed.

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

Ans: No skewness or it’s symmetrical.

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

Ans: Positively skewed distribution.

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

Ans: Negatively skewed distribution.

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

Ans: It indicates that the mean is more and variance is less.

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

Ans: It indicates that the mean is less and variance is more.

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



What can we say about the distribution of the data?

Ans: The data is not normally distributed.

What is nature of skewness of the data?

Ans: Negatively skewed distribution.

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:

* In the given data Boxplot 2 is widely distributed than the Boxplot 1.
* The median for both Boxplots are equal.
* The data distributed in both the Boxplots are symmetrical.
* In the Boxplot 2 it’s hard to make a prediction as the data range varies high.

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)

Ans: Mean(MPG) = 34.422

Std(MPG) = 9.131

Using this code in python

“stats.norm.sf(38,34.422,9.131)”

P(MPG>38) = 0.347 = 34.7% (ans)

* 1. P(MPG<40)

Ans:

Mean(MPG) = 34.422

Std(MPG) = 9.131

Using this code in python

“stats.norm.cdf(40,34.422,9.131)”

P(MPG<40) = 0.729 = 72.9% (ans)

* 1. P (20<MPG<50)

Ans:

Using this code in python

1-stats.norm.cdf(20,34.422,9.131)-stats.norm.sf(50,34.422,9.131)

P(20<MPG<50) = 0.898 = 89.8 % (ans)

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 is Normally Distributed.

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

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) is Normaly Distributed and also Circumference(Waist) is Normaly Distributed.

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

Ans: Z scores for 90% CI = 1.644

Z scores for 94% CI = 1.880

Z scores for 60% CI = 0.841

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

Ans: t scores for 95% CI = 2.063

t scores for 96% CI = 2.171

t scores for 99% CI = 2.796

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:

Sample mean = 260

Population mean = 270

Std = 90

n = 18

So, as per formula t = -0.471

Degrees of freedom = 18-1 = 17

Using code in python – stats.t.cdf(-0.471,17) = 0.3218 = 32.88%

So there is a 32.88% probability that 18 randomly selected bulbs would have an average life of no more than 260 days.