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

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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Three coins are tossed at once.

The possible outcomes will be (HHH,TTT,HHT,HTH,THH,TTH,THT,HTT).

So, favorable outcomes = HHT,THH,HTH.

NO. of favourable outcomes = 3

Hence the required probability is PE = no. of favourable outcomes/total no. of outcomes

Probability = 3/8

Hence, when 3 coins are tossed then the probability of 2h and 1 tail is 3/8.

窗体底端

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

The probability of getting a sum of 1 is 0

1. Less than or equal to 4

Probability of sum less than or equal to 4 = 1/6

1. Sum is divisible by 2 and 3

Probability of sum divisible by 2 and 3 = 5/36

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?

Total number of balls = 2 + 3 + 2 = 7

Number of ways of drawing 2 balls out of 7 = 7C2 = (7 × 6) / (2 × 1) = 42/2 = 21

Number of balls other than blue = 5

Number of ways of drawing 2 balls out of 5 = 5C2 = (5 × 4) / (2 × 1) = 20/2 = 10

So the required probability = 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

Expected Number of Candies (E[X]) = Σ [Candy count of child \* Probability of that child]

E[X] = (1 \* 0.015) + (4 \* 0.20) + (3 \* 0.65) + (5 \* 0.005) + (6 \* 0.01) + (2 \* 0.120)

E[X] = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

E[X] = 3.095

So, the expected number of candies for a randomly selected child is 3.095

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**

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?

To calculate the expected value (mean) of the weights of the patients at the clinic, you need to sum up all the weights and then divide by the number of patients. Here are the weights given:

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

Calculating the expected value (E[X])

E[X] = (Σx) / n

E[X] is the expected value.

Σx represents the sum of all values in the dataset.

n is the number of values in the dataset.

Now, sum up all the weights:

Σx = 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199

Σx = 1098

calculating the expected value:

E[X] = Σx / n

E[X] = 1098 / 9

E[X] = 122

So, the expected value of the weight of a patient chosen at random is 122 pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**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?

import numpy as np

import pandas as pd

from scipy import stats

from scipy.stats import norm

# Avg. weight of Adult in Mexico with 94% CI

stats**.**norm**.**interval(0.94,200,30**/**(2000**\*\***0.5))

(198.738325292158, 201.261674707842)

# Avg. weight of Adult in Mexico with 98% CI

stats**.**norm**.**interval(0.98,200,30**/**(2000**\*\***0.5))

(198.43943840429978, 201.56056159570022)

# Avg. weight of Adult in Mexico with 96% CI

stats**.**norm**.**interval(0.96,200,30**/**(2000**\*\***0.5))

(198.62230334813333, 201.37769665186667)

**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?

arrange the scores in ascending order:

34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56

Mean (Average): Mean = (Sum of all scores) / (Total number of scores)

Mean = (34 + 36 + 36 + 38 + 38 + 39 + 39 + 40 + 40 + 41 + 41 + 41 + 41 + 42 + 42 + 45 + 49 + 56) / 18 Mean = 676 / 18

Mean ≈ 37.56

Median: The median is the middle value in the sorted list. Since we have 18 scores, the median will be the average of the 9th and 10th scores because there are 9 scores below and 9 scores above these positions.

Median = (40 + 40) / 2

Median = 40

Variance = [(34 - 37.56)^2 + (36 - 37.56)^2 + ... + (56 - 37.56)^2] / 18

Take the average: variance ≈ 74.67

Standard Deviation:

The standard deviation is the square root of the variance.

Standard Deviation ≈ √74.67 ≈ 8.65

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

Zero skew, if the mean and median are equal.

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

If the mean is greater than the median, the distribution is positively skewed

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

 If the mean is greater than the median, the distribution is positively skewed

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

Positive values of kurtosis indicate that distribution is peaked and possesses thick tails

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

 That the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Q19) Comment on the below Boxplot visualizations?



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

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

For 90% confidence interval:

We have the significance level at 5 % ( as it is a two tailed test)

that is:

α = 5 % = 0.05

z at α = 0.05 from the z table will be:

z = 1.645.

For 94 % confidence interval, we get:

We have the significance level at 3 % ( as it is a two tailed test)

that is:

α = 3 % = 0.03

z at α = 0.03 from the z table will be:

z = 1.555.

For 60 % confidence interval, we get:

We have the significance level at 20 % ( as it is a two tailed test)

that is:

α =20 % = 0.2

z at α = 0.2 from the z table will be:

z = 0.253

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

1. For a 95% confidence interval:

The confidence interval extends from α/2 = 0.025 to 1 - α/2 = 0.975.

Using a t-table or calculator with 24 degrees of freedom, the t-score for the 95% confidence interval is approximately 2.064.

1. For a 96% confidence interval:

The confidence interval extends from α/2 = 0.02 to 1 - α/2 = 0.98.

Using a t-table or calculator with 24 degrees of freedom, the t-score for the 96% confidence in

approximately 2.171.

1. For a 99% confidence interval:

The confidence interval extends from α/2 = 0.005 to 1 - α/2 = 0.995.

Using a t-table or calculator with 24 degrees of freedom, the t-score for the 99% confidence interval is approximately 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

Calculating the t-score:

t\_score = sample mean - population mean/standard error

= 260-270/(90/√18)

Calculating the degrees of freedom (df), which is equal to the sample size minus 1:

df = 18-1

= 17

Now, can use the pt() function in R to find the probability that the t-score is less than or equal to the calculated t-score:

# Calculating t-score

t\_score <- (260 - 270) / (90 / sqrt(18))

# Calculate probability using pt() function

probability <- pt(t\_score, df)

# Print the probability

Probability

When we run this R code, it will give you the probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the CEO's claim were true.