|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | DISCRETE |
| Results of rolling some 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 | CONTINUOUS |
| 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 | ORDINAL |
| 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 | INTERVAL |
| Religious Preference | NOMINAL |
| Barometer Pressure | RATIO |
| SAT Scores | RATIO |
| Years of Education | RATIO |

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

Ans: HHH, HHT, HTH, THH, TTH, TTT, HTT, THT

3/8 = 0.375

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1,1 | 2, 1 | 3, 1 | 4, 1 | 5, 1 | 6, 1 |
| 1,2 | 2, 2 | 3, 2 | 4, 2 | 5, 2 | 6, 2 |
| 1,3 | 2, 3 | 3, 3 | 4, 3 | 5, 3 | 6, 3 |
| 1,4 | 2, 4 | 3, 4 | 4, 4 | 5, 4 | 6, 4 |
| 1,5 | 2, 5 | 3, 5 | 4, 5 | 5, 5 | 6, 5 |
| 1,6 | 2, 6 | 3, 6 | 4, 6 | 5, 6 | 6, 6 |

Possible outcomes:

Summation of the values:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUM | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

1. Equal to 1

Ans: 0

1. Less than or equal to 4

Ans: (1,1) (1,2) (1,3) (2, 1) (2, 2) (3,1)

6/36 = 0.167

1. Sum is divisible by 2 and 3.

Ans: (1,5) (2,4) (3,3) (4,2) (5,1) (6,6)

6/36 = 0.167

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: 5C2 / 7C2 =10/21

Q6) Calculate the Expected number of candies for a randomly selected child.

Below are the probabilities of the count of candies for children (ignoring the nature of the child-Generalized view)

Child A – the probability of having 1 candy = 0.015.

Child B – the probability of having 4 candies = 0.20

|  |  |  |
| --- | --- | --- |
| 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 |

Ans: Expected number of candies for a randomly selected child is the weighted average of the data.

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | X\*P(X) |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.06 |
| C | 3 | 0.65 | 0.045 |
| D | 5 | 0.005 | 0.075 |
| E | 6 | 0.01 | 0.09 |
| F | 2 | 0.12 | 0.03 |
|  |  | Weighted average | 0.315 |

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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.596563 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.276948 | 0.927461 | 0.927461 |
| Standard Deviation | 0.526258 | 0.963048 | 1.758801 |
| Range | 2.17 | 3.911 | 8.4 |

Use Q7.csv file

Ans:

* The mean median and mode of the data are not equal.
* Hence, we can conclude that they are not forming a normal distribution for all three data.
* The standard deviation of these data shows it is more

scattered.

Q8) Calculate the 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: (108+110+123+134+135+145+167+187+199)/9 = 145.3

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

Cars speed and distance

Use Q9\_a.csv.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Index | speed | dist |
| Skewness | -3.77627E-17 | -0.11751 | 0.806895 |
| Kurtosis | -1.2 | -0.50899 | 0.405053 |

SP and Weight (WT)

Use Q9\_b.csv.

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.61E+00 | -0.61475 |
| Kurtosis | 2.977329 | 0.950291 |

Q10) Draw inferences about the following boxplot & histogram



Ans: Histogram:

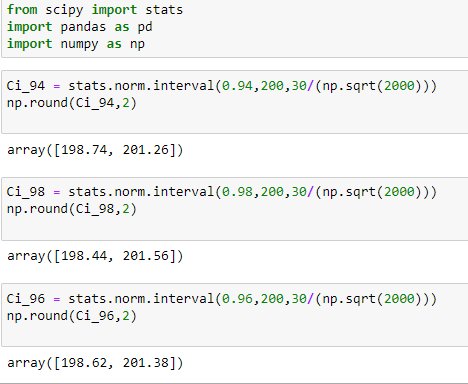
* The mean is btw 50 to 100.
* The data is positively skewed.
* The kurtosis is positive.

Box plot:

* There are few outliers in the data.
* The outliers are above the upper fence.

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?

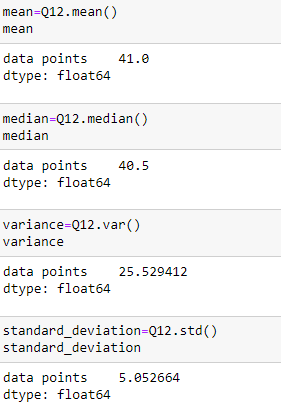
Ans:



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.



1. What can we say about the student marks?

* It is not normally distributed as mean and median are not equal. However, the difference is 0.5.
* The data points are deviated from the mean by 5 points.

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

The data is normally distributed as the mean and median are equal. Which means skewness is zero.

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

The data will be right-skewed when the mean > median.

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

The data will be left-skewed when the mean > median.

Q16) What does positive kurtosis value indicate for data?

The positive kurtosis value indicates that the data has a sharp peak.

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

The negative kurtosis value indicates that the data has a wider peak.

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



What can we say about the distribution of the data?

It is a continuous distribution. And not normally distributed.

What is the nature of the skewness of the data?

The data is left-skewed or negatively skewed.

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

As per the data approximately the Q3 is 18 and Q1 is 10 so, IQR=Q3-Q1=8.

Q19) Comment on the below Boxplot visualizations?



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

* The median of the data is equal, whereas the IQR differs.
* The range of the data is also less with 1 than the 2.
* The data is normally distributed.
* Data does not have any outliers.

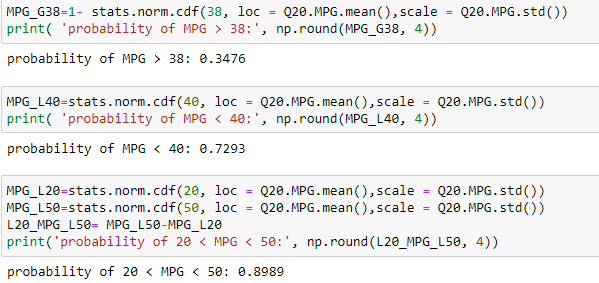
Q 20) Calculate probability from the given dataset for the below cases.

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

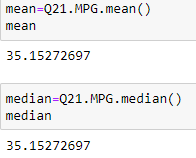
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows the normal distribution.

1. Check whether the MPG of Cars follows Normal Distribution.

Dataset: Cars.csv

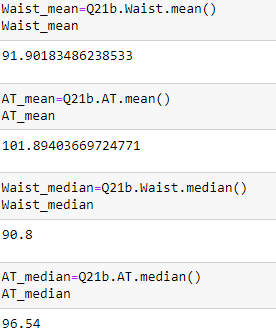
Ans:

Yes, the data is normally distributed as the mean and median are equal.

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

Dataset: wc-at.csv

Ans:



The Adipose Tissue (AT) and Waist Circumference (Waist) from the wc-at data set do not follow Normal Distribution as the mean and median of the data are not equal.

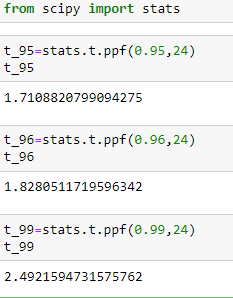
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.

Ans:

|  |  |
| --- | --- |
| Confidence interval | Z scores |
| 90% | 1.645 |
| 94% | 1.88 |
| 60% | 0.84 |

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

Ans:



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:

