|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Ratio |
| Results of rolling a dice | Ratio |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Ratio |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Ratio |
| Number of tickets in Indian railways | Ratio |
| Number of times married | Ratio |
| 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 | 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 | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Answer: P(two heads and one tail)= 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

Anwers:

P(Equal to 1)= 0

P(Less than or equal to 4)= 6/36 = 1/6

P(Sum is divisible by 2 and 3) = 6/36 = 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?

Answer: P(none of the balls drawn is blue)= 5/7

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

Answer:

|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | weighted Average |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.2 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.12 | 0.24 |
| Total | 21 | 1 | 3.09 |

Expected number of candies for a randomly selected child = 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.

**Use Q7.csv file**

**Done in Jupyter notbeook**

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  
Expected value of the randomly selected patient is 145.3333

**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**

**Done in Jupyter notbeook**

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



It has positive skewness.



Data has some outliers

**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?

Solved in notebook

**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. Mean = 41.0
   2. Standard deviation = 5.0526
   3. Variance = 25.529
2. What can we say about the student marks?
   1. Average marks that students score is 41
   2. Most of the students have score between 34 to 41
   3. Only three students have score more than or equal to 45
   4. 56 is the highest marks
   5. 34 is the lowest marks

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

Zero sckewness

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

Positive Skewness

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

Negative Skewness

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

Data is centred around the mean and there are less values at tails

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

A distribution with a negative kurtosis value indicates 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?

Most of the values are distributed between 10 to 18

There are some data points between 2 to 10

What is nature of skewness of the data?

Date as negative skewness

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

IQR = 8

Q19) Comment on the below Boxplot visualizations?



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

1. In BQ1 date is distributed approximately between 235-287 where as in BQ2 data is approximately between 200-337.
2. Both data has BQ has the same mean

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)=34.75%
  2. P(MPG<40)=72.93%

c. P (20<MPG<50)=89.88%

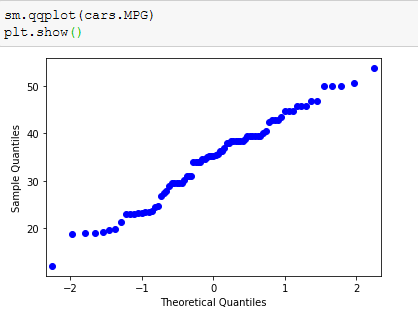
solved in python

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

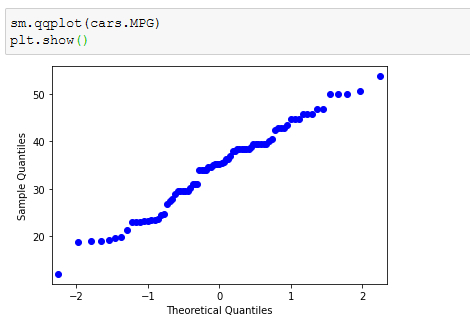
Data follows the normal distribution

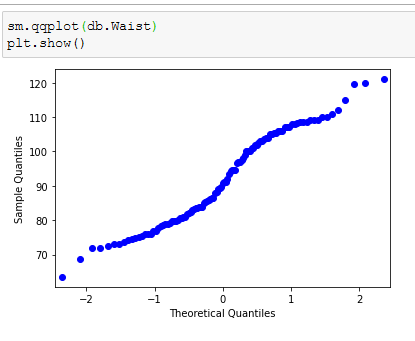


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

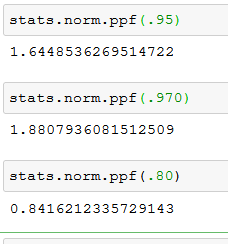
Dataset: wc-at.csv

Data follows the normal distribution





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



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

**Solved in notbook**

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