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

**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 | Nominal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Solution: - There are a total of 2^3 = 8 possible outcomes when three coins are tossed.

We need to count the number of outcomes that have exactly two heads and one tail. There are three ways this can happen: HHT, HTH, and THH (where H represents heads and T represents tails).

Each of these outcomes has a probability of (1/2)^3 = 1/8 of occurring, since the probability of getting heads or tails on any one toss is 1/2.

Therefore, the probability of getting two heads and one tail is :-

P(Two heads and one tail) = P(HHT) + P(HTH) + P(THH)

= 1/8 + 1/8 + 1/8

= 3/8

So the probability of getting two heads and one tail when three coins are tossed is 3/8, or about 0.375.

**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**

Solution: -

1. The sum of two dice cannot be 1, so the probability of getting a sum of 1 is 0.
2. The possible outcomes of less then or equal to 4 is:

1+2=3

2+1=3

1+1=2

3+1=4

1+3=4

There are 5 possible outcomes out of total (6\*6)=36 possible outcomes when rolling two dice

The probability of obtaining a sum or less then or equal to 4 is : 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?**

Solution:-

To find the probability that none of the balls drawn is blue, the total number of ways to draw 2 balls out of 7, and the number of ways to drawn 2 balls such that none of them is blue.

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

Let S be the sample space

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

=

=21

Let E= Event of drawing 2 ballls, none of which is blue

n(E)= number of ways drawing 2 balls out of (2+3) balls

=5C2

=

=10

P(E)= *=*  Answer

**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

**Solution: -** To calculate the expected number of candies for a randomly selected child, we need to multiply each possible candy count by its corresponding probability, and then add up the products. That is:

Expected number of candies = (1 x 0.015) + (4 x 0.20) + (3 x 0.65) + (5 x 0.005) + (6 x 0.01) + (2 x 0.120)

Expected number of candies = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24

Expected number of candies = 3.085

Therefore, the expected number of candies for a randomly selected child is 3.085.

**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**

**Solution:- Assignments\_1 python notebook**

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?

**Solution**: - Expected value = sum of all values probability of each value

(108+110+123+134+ 135+145+167+187+199) (1/9)

1308/9

145.33 Answer

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution:- Assignments\_1 python notebook**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Solution:- Assignments\_1 python notebook**

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



**Answer:-**

1. Most of chick weight data is between 50 to 100 has high frequency i.e 200.
2. Second largest chick weight is between 100 to 150 that is between frequency of 100 to 150.
3. As much as ChickWeight$Weight is increasing the frequency is gradually decreasing.



**Answer:-** Data which is out of the the boxplot is called outliers. There are 6 data points in this boxplot which is 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?

**Solution: -**

Given:-

Sample Size(n) = 2000

Mean (x̄) = 200 pounds

Standard Deviation(s) = 30 pounds

Confidence Interval (CI) = x̄±z

1. 94% Confidence Interval

CI = 200±1.88x

CI = 200±1.2611

Lower Bound = 198.73

Upper Bound = 201.26

The 94% confidence interval for the population mean weight of adult males in Mexico is (198.73, 201.26) pounds.

1. 96% Confidence Interval

Similarly,

CI = 200±2.05x

CI = 200±1.2611

Lower Bound = 198.62

Upper Bound = 201.37

The 96% confidence interval for the population mean weight of adult males in Mexico is (198.62, 201.37) pounds.

1. 98% Confidence Interval

Similarly,

CI = 200±2.33x

CI = 200±1.2611

Lower Bound = 198.44

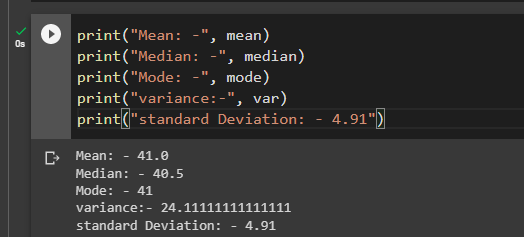
Upper Bound = 201.56

The 96% confidence interval for the population mean weight of adult males in Mexico is (198.44, 201.56) pounds

**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?

**Solution: -** Most of the student performed average is 41 marks

Median indicates that half of the student scored below 40.5 marks and half of the student is scored above the 40.5 marks. And mode indicates that most of the students obtained 41 marks/

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

**Answer: -**

If the mean, median, and mode of a dataset are all the same, it means that the data is evenly distributed around the middle point and has zero skewness . With the same amount of data on both sides.

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

**Answer: -**

When mean is greater than median it means majority of data is on the left side and data is positively skewed.

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

**Answer: -**

When median is greater than mean it means majority of data is on right side and negatively skewed.

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

**Answer**

Positive kurtosis indicates that heavier tails and a higher peak then the normal distribution.

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

**Answer**

Negative kurtosis value indicates distribution has thinner tails and longer tails than the normal distribution.

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



1.What can we say about the distribution of the data?

**Answer: -** The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

**Answer: -**The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

**Answer: -** The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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.

**Answer:-**

1. First there are no outliers.
2. Both the boxplot share same median that is approx. in an range between 275 to 250.
3. They are normally distributed with zero to no skewness either at minimum or maximum whisker range.

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)

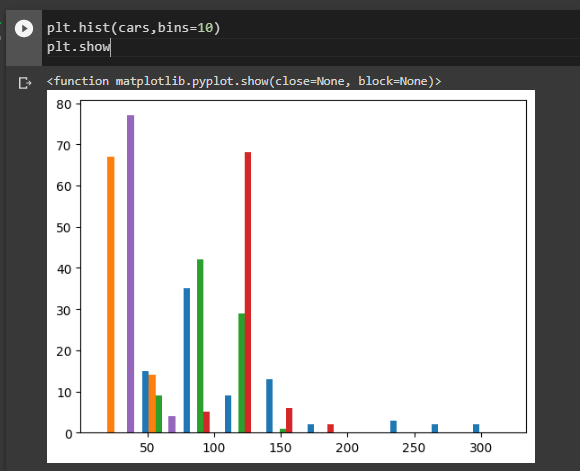
c. P (20<MPG<50)

**Answer:** - python notebook assignment\_1

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

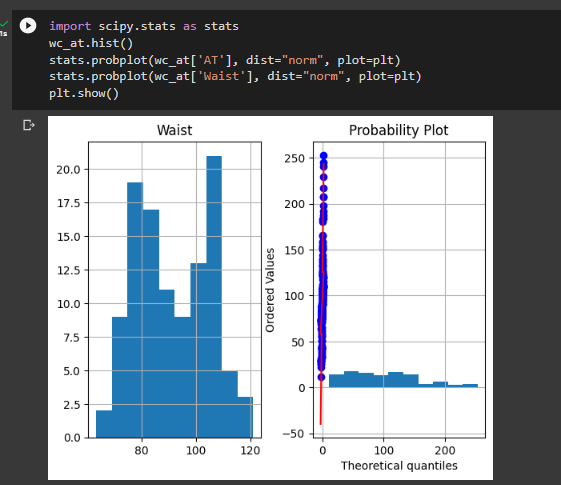
Dataset: Cars.csv

Answer :- MPG of cars is follow normal distribution as per histogram visualization.

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

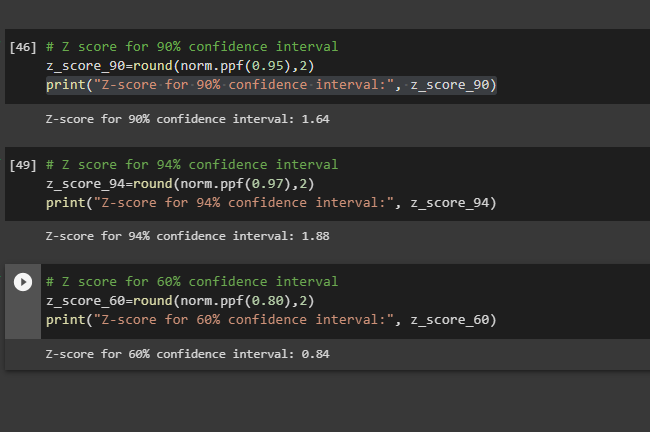
Dataset: wc-at.csv

**Answer: -** Adipose Tissue (AT) and Waist Circumference(Waist) does not follow normal distribution.

****

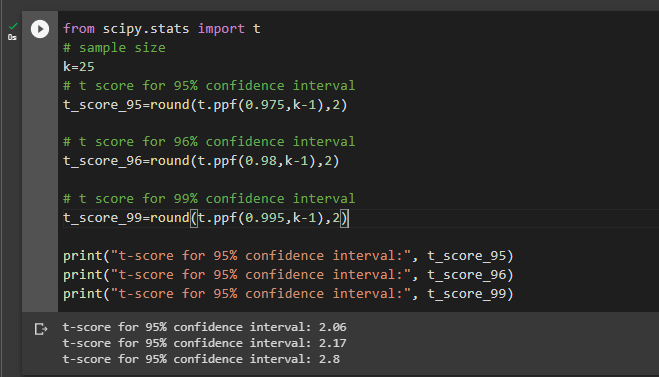
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

**Answer:-**

****

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

**Answer:-**

****

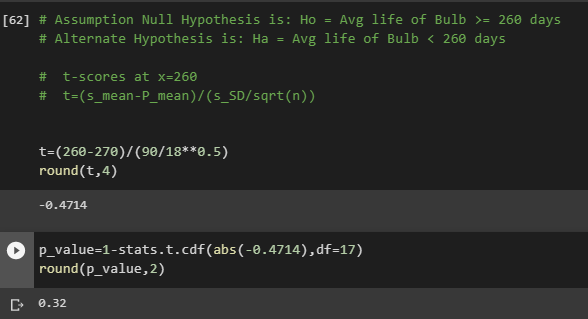
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

**Answer : - 0.32**

****