|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete data type |
| Results of rolling a dice | Discrete data type |
| Weight of a person | Continuous data type |
| Weight of Gold | Continuous data type |
| Distance between two places | Discrete data type |
| Length of a leaf | Continuous data type |
| Dog's weight | Continuous data type |
| Blue Color | Discrete data type |
| Number of kids | Discrete data type |
| Number of tickets in Indian railways | Discrete data type |
| Number of times married | Discrete data type |
| Gender (Male or Female) | Discrete 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 Data Type |
| High School Class Ranking | Ordinal Data Type |
| Celsius Temperature | Interval Data Type |
| Weight | Ratio Data Type |
| Hair Color | Nominal Data Type |
| Socioeconomic Status | Interval Data Type |
| Fahrenheit Temperature | Interval Data Type |
| Height | Ratio Data Type |
| Type of living accommodation | Nominal Data Type |
| Level of Agreement | Ordinal Data Type |
| IQ(Intelligence Scale) | Interval Data Type |
| Sales Figures | Ratio Data Type |
| Blood Group | Nominal Data Type |
| Time Of Day | Interval Data Type |
| Time on a Clock with Hands | Ordinal Data Type |
| Number of Children | Nominal Data Type |
| Religious Preference | Ordinal Data Type |
| Barometer Pressure | Interval Data Type |
| SAT Scores | Ratio Data Type |
| Years of Education | Nominal Data Type |

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

A) when three coins are tossed the outcomes will be 8

The probability of coming two heads and one tail are 3 that is (HHT,THH,THT)

Probability = Number of favorable outcomes/Total number of outcomes

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

= 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

a) There is no Probability of getting equal to 1 because two dices rolled means.

The outcome will be 2

So, the Probability=0/36

b) The total numbers of outcomes = 36

The Probability of getting less than or equal to 4 is = 6/36

= 1/6 c) Probability=Total number of favorable outcomes/Total number of outcomes

Outcomes = (1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

Probability=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?

A) Total number of balls=7

The random two balls drawn are=7C2

=7C2=7\*6/2\*1=21

The probability that none of the balls drawn blue is =5C2

=5\*4/2\*1

=20/2

=10

The total outcomes=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

A) To calculate the expected number of candies for a randomly selected child, we have to multiple Candies\*Probability=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=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**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data set** | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596 | 3.217 | 17.848 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 0 3.07  1 3.92 | 0 3.44 | 0 17.02  1 18.90 |
| **Standard** **deviation** | 0.534 | 0.978 | 1.786 |
| **Variance** | 0.285 | 0.957 | 3.193 |
| **Range**: **Max**  **Min** | 4.93  2.76 | 5.424  1.513 | 22.900  14.500 |

**A)**

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?

1. The Expected value of the Weight of the patient is

108+110+123+134+135+145+167+187+199/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

|  |  |  |
| --- | --- | --- |
| **Data set** | **Skewness** | **Kurtosis** |
| **Speed** | **-0.117510** | **-0.508994** |
| **Dist** | **0.806895** | **0.405053** |

**A)**

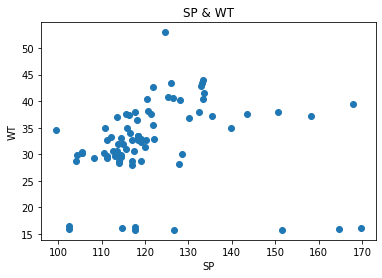
|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
| **Data set** | **Skewness** | **Kurtosis** |
| **SP** | **1.611450** | **2.977329** |
| **Weight(WT)** | **-0.614753** | **0.950291** |

**A)**



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



A) The above Histogram Represents the following of chick Weight&weight. So, on the Y-axis represents the “Frequency” and on X-axis represents the “chick Weight&weight”. The frequency level is from (0-200). The chickWeight$weigh of level is from (0-400).The Highest Frequency occurs at 75 and the lowest one is 375-400.



A)  Median is less than mean right skewed and we have outlier on the upper side of box plot and they is less data points between Quartely 1 and bottom point.

**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?

A)

**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?

A) 1a. Mean=Sum of observations /Number of observations

=738/18

=41

Median=The above data arranged in ascending order, so the n is even (40+41)/2=40.5

Variance=summation of (x-xbar)2/18

=2 Standard deviation=under root (Variance)

=4.91

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

A) If mean and median of the data are equal then there is no skewness and the data is in Normalized.

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

A) The Nature of skewness becomes “Positive”, when mean is greater than median.

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

A) If median is greater than mean, so the nature of skewness will be indicates as “Negative skewness”.

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

A) Positive Kurtosis indicates the value of wider tails and thinner peaks for the data.

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

A) It will indicates the wider peaks and wider tails.

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



What can we say about the distribution of the data?

A) Distribution of the data is not “Normally Distributed”

What is nature of skewness of the data?

A) Negative skewness

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

A) 10-18

Q19) Comment on the below Boxplot visualizations?



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

A) The above two box plots are in Normal Distribution.

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)

1. 0.4074
   1. P(MPG<40)

A)0.7531

C. P (20<MPG<50)

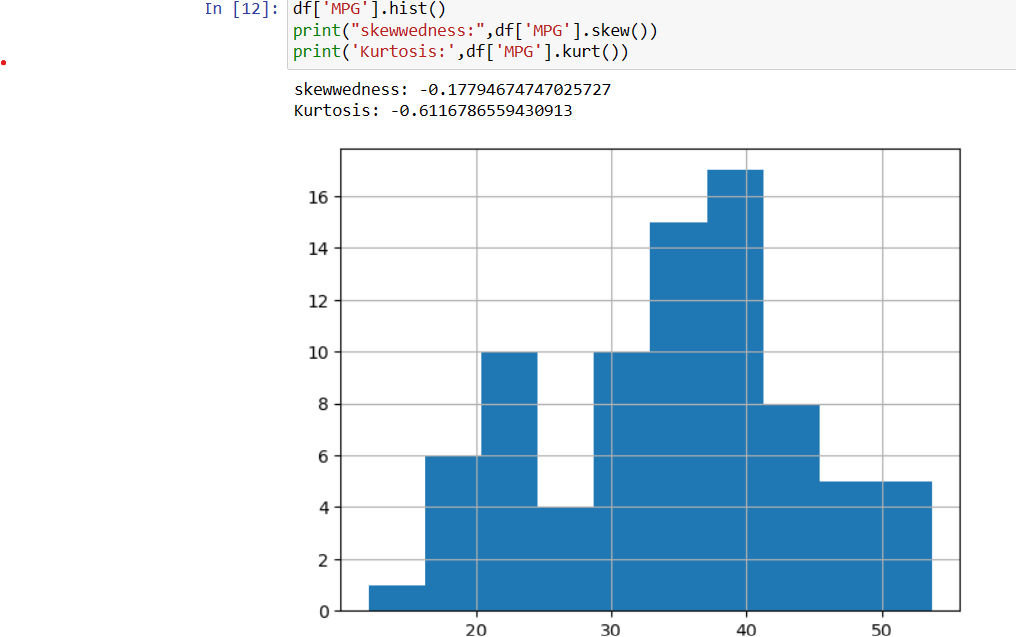
A) 0.0247

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

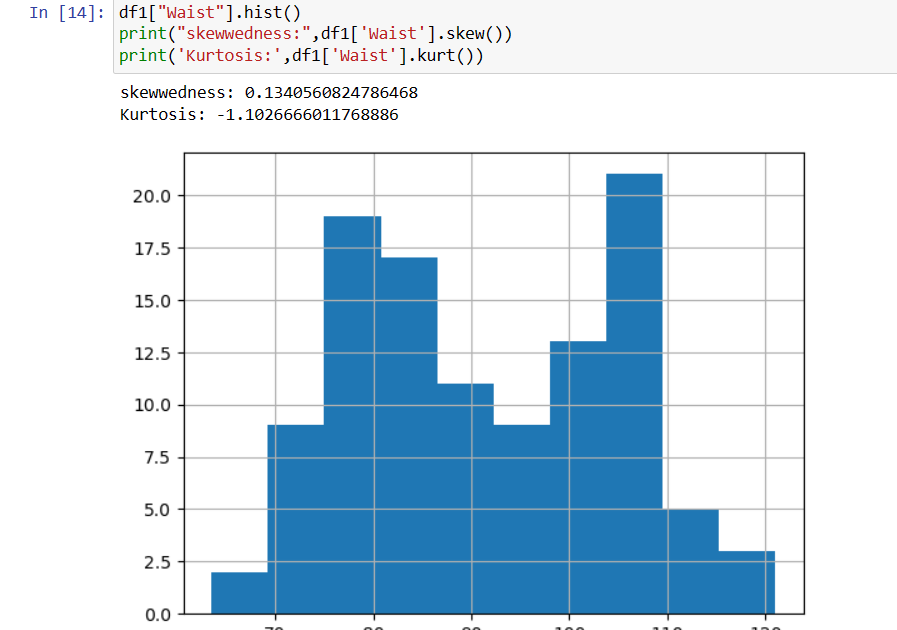
1. The below Histogram represents that the data is Normally Distribution.



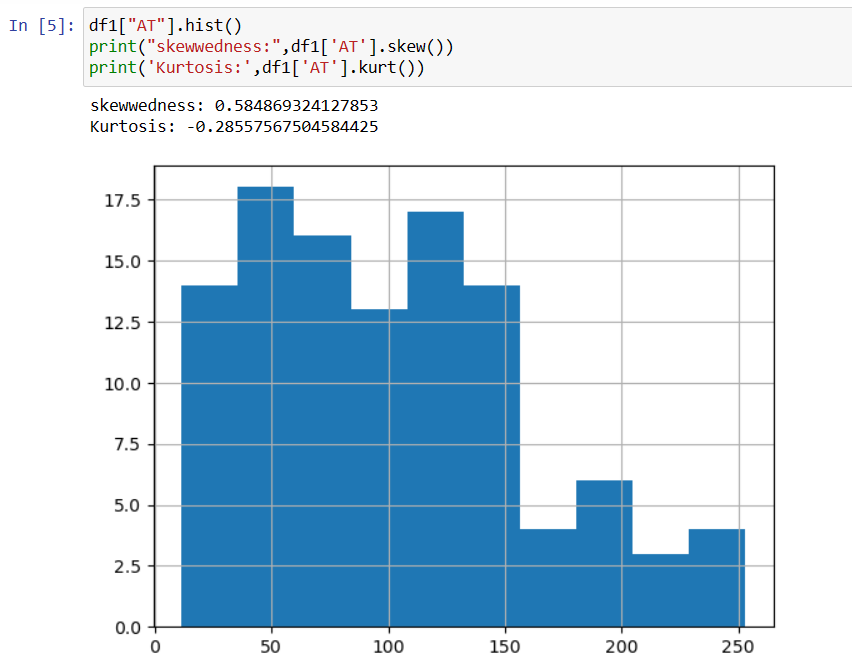
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

1. From the data wc-at.Waist doesn’t follow Normal Distribution it’s show in the below histogram.



* Adipose tissue (AT) also doesn’t follow Normal Distribution



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

A) z-score for 90% confidence interval: 1.6449

z-score for 94% confidence interval: 1.8808

z-score for 60% confidence interval: 0.8416

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

A) For sample size = 25

t-score for 95% confidence interval: 2.0639

t-score for 96% confidence interval: 2.1715

t-score for 99% confidence interval: 2.7969

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

A) The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is: 0.32167253567098364