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

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

Sample set = {HHH,HHT,HTT,TTT,TTH,THH,HTH,THT}

Probability of two heads and one tail = {HHT,THH,HTH}.

=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 2and 3

Sample set ={(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),

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

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

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

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

(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

1. Equal to 1

Favorable sets = 0

Probability = 0.

1. Less than or equal to 1

Favorable sets = 6

Probability = 6/36

= 1/6.

1. Sum is divisible by 2 and 3

Favorable sets = 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?

Probability of drawing blue balls = 2/7.

Probability of no blue balls drawn = 1- 2/7

= 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.

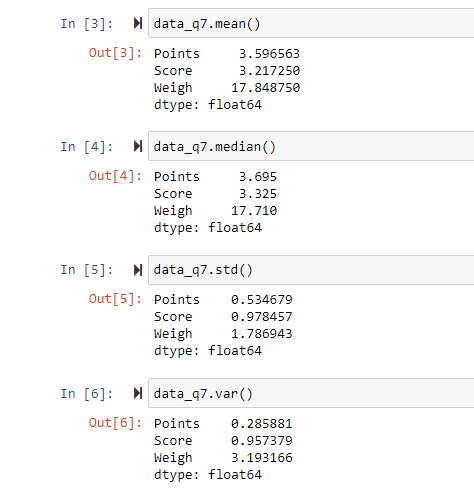
Child B – probability of having 4 candies = 0.20

Probability of count of candies for children = (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.

* Median

Points= 3.695

Score= 3.325

Weigh =17.71

* Mean

Points=3.596

Score=3.217

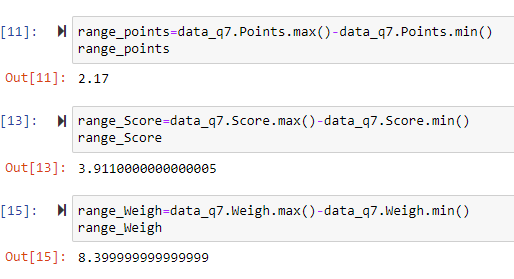
Weigh=17.848

* Mode

Points=3.07, 3.92

Score=3.44

Weigh=18.90, 17.02

* Standard deviation

Points=0.534679

Score=0.978457

Weigh=1.786943

* Variance

Points= 0.285881

Score=0.957379

Weigh=3.193166

* Range = Maximum value – Minimum value

Points=2.17

Score=3.91

Weigh=8.399

**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

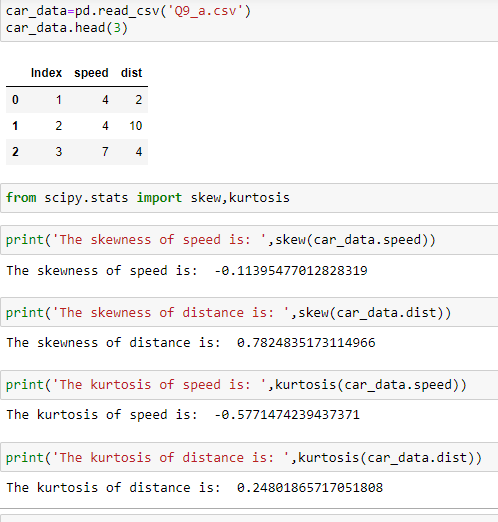
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Expected Value = (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**

**Skewness**

Speed = -0.1139

Distance = 0.7824

**Kurtosis**

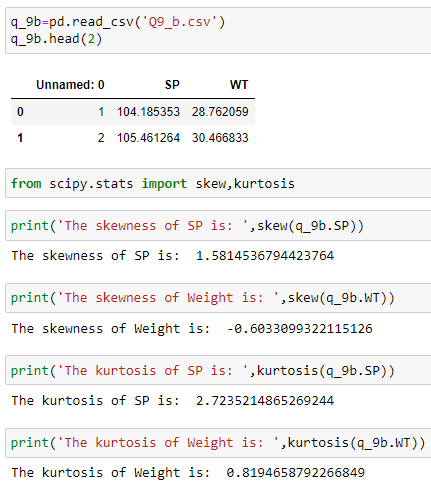
Speed = -0.5771

Distance = 0.2480

* The skewness of speed is negative, it signifies that the data are negatively skewed or left skewed
* Mean<median<mode
* Negative kurtosis signifies that the curve is flat than the normal curve with thin tails and the data is called as platykurtic data
* The skewness of distance is positive, it signifies that the data is positively skewed or right skewed
* Mean>median>mode
* Positive kurtosis signifies that the curve is peaked than the normal curve with thick tails and the data is called as Leptokurtic data

**SP and Weight(WT)**

**Use Q9\_b.csv**

** Skewness**

SP = 1.58145

WT = -0.60333

**Kurtosis**

SP = 2.72352

WT = 0.81946

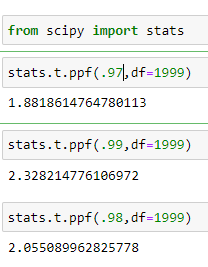
**Q10) Draw inferences about the following boxplot & histogram**



* The above histogram shows that the values are positively skewed or right skewed since more number of data are present on the left side
* The skewness values is greater than zero
* Thin tail is present towards the right side
* mean>median>mode



* The boxplot signifies that greater density of data present towards left side
* The data are right or positively skewed
* The data containes outlier in the positive direction, the dots signifies the presence of 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?

t94 = 1.882 `

t98 = 2.328

t96 = 2.055

CI for 94% confidence

= 200-1.882\*(30/√2000) to 200+1.882\*(30/√2000)

=198.73 to 201.26

CI for 98% confidence

= 200-2.328\*(30/√2000) to 200+2.328\*(30/√2000)

=198.43 to 201.56

CI for 96% confidence

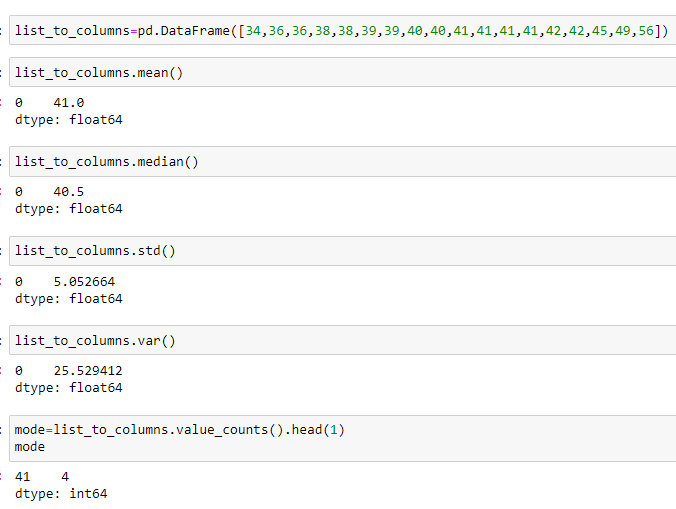
= 200-2.055\*(30/√2000) to 200+2.055\*(30/√2000)

=198.62 to 201.378

**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.

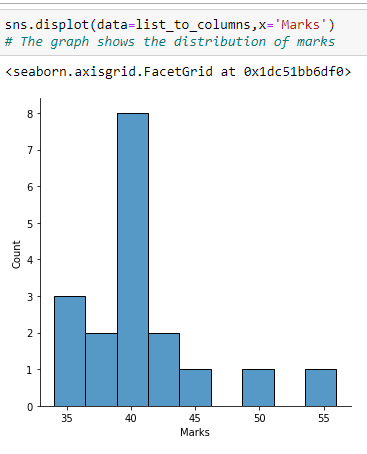
Mean =41

Median =40.5

Mode =41

Std deviation =5.053

Variance =25.53

1. What can we say about the student marks?

* Marks of 18 students are given.
* The average mark scored by students is 41.0
* The maximum and Minimum marks are 56 and 34 respectively
* The marks are arranged in ascending order
* The graph shows that, most of the students score marks in between 35 to 45

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

When mean, median of data are equal, then the nature of skewness is Zero

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 less than the median, the distribution is negatively skewed

Q16) What does positive kurtosis value indicates for adata?

Positive values of kurtosis indicate that **a distribution is peaked and possess thick 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?

* The values are not normally distributed
* The median value is 15
* The upper and lower quartile values are 18 and 10 respectively
* The values are left skewed

What is nature of skewness of the data?

Here the data are skewed towards left side as there is longer whisker on left side.

What will be the IQR of the data (approximately)?   
 IQR = Q3-Q1

= 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.

* Both the boxplot shows some symmetry means that they are normally distributed.
* There is no outliers.
* Median is around 262.5.

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

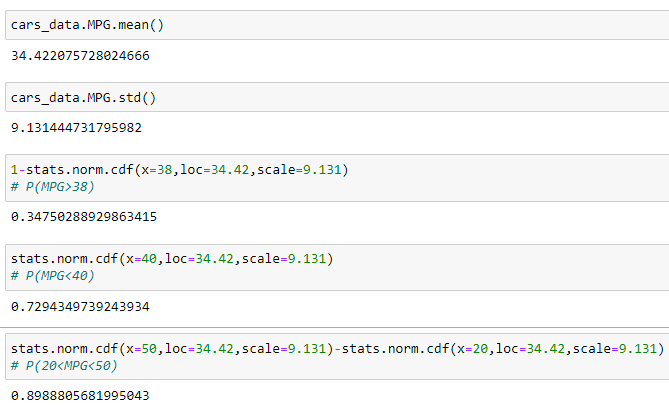
Data \_set: Cars.csv

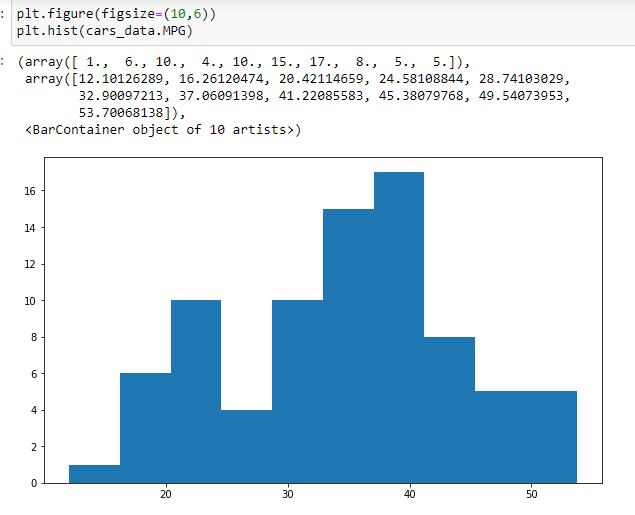
Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

c. P (20<MPG<50)



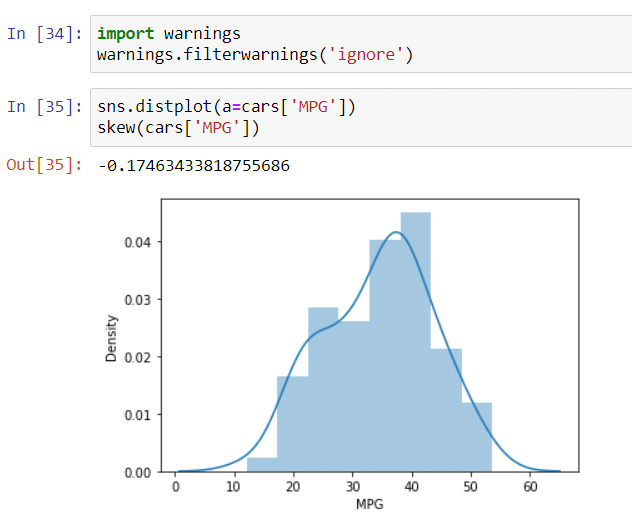


1. P(MPG>38) = 0.3475
2. P(MPG<40) = 0.729

c. P (20<MPG<50) = 0.899

Q 21) Check whether the data follows normal distribution

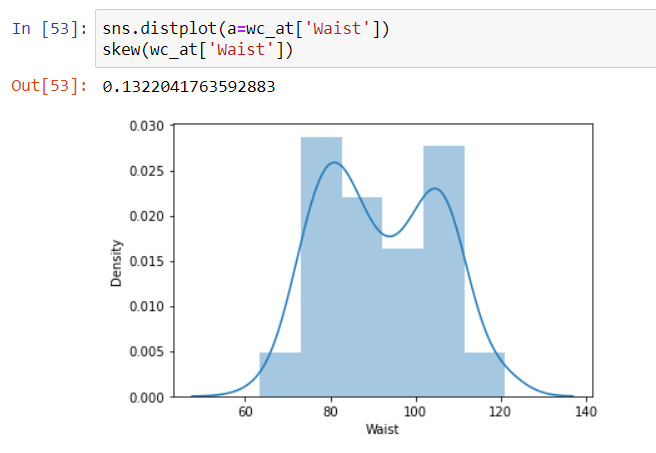
1. Check whether the MPG of Cars follows Normal Distribution

 Dataset: Cars.csv

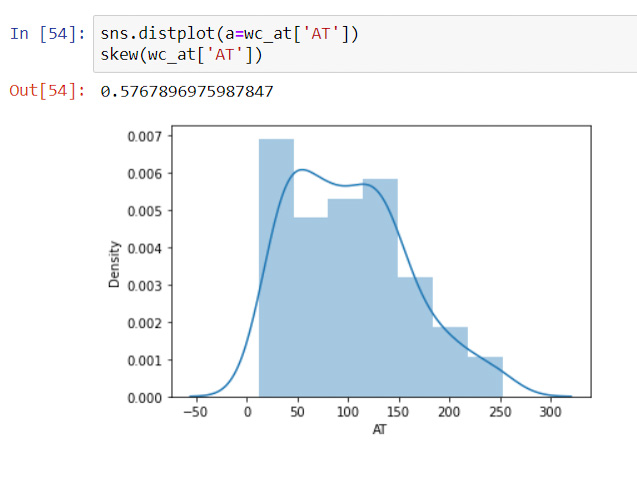
* Skewness around -0.5 to 0.5 is acceptable, here skewness of MPG is -0.17 hence it follows normal distribution.

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

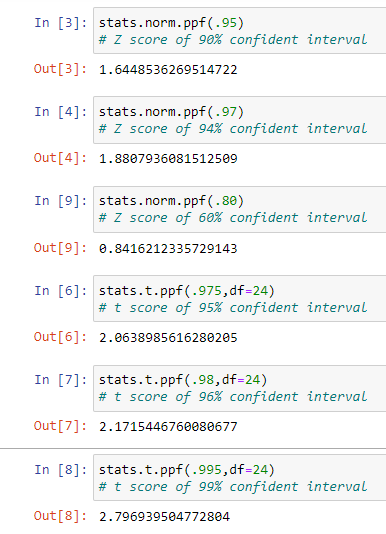
Dataset: wc-at.csv



Skewness around -0.5 to 0.5 is acceptable, here skewness of Waist is0.13 hence it follows normal distribution.



Skewness around -0.5 to 0.5 is acceptable, here skewness of AT is0.57 hence it follows normal distribution.

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

Z90 = 1.645

Z94 = 1.88

Z60 = 0.842

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

t95= 2.064

t96 = 2.172

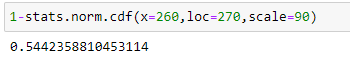
t99 = 2.797.

Q 24**)**A Government companyclaims 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



There is 54.4% probability that the randomly selected bulbs would have an average life of no more than 260days.