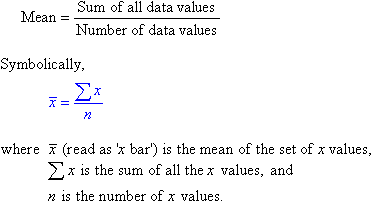
**STEP 1. Find out measures of central tendency for single and multi-attributes data using Statistical analysis**

**Mean :**Mean is also known as average of all the numbers in the data set which is calculated by below equation.

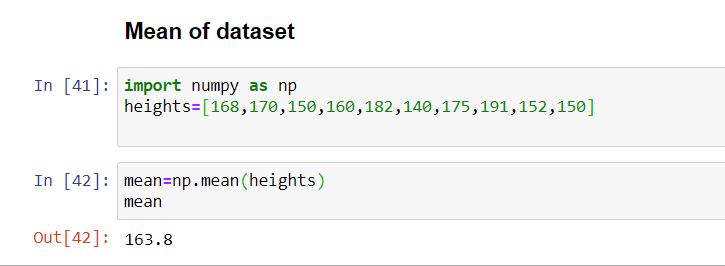


The following data consist of heights of persons.

heights=[168,170,150,160,182,140,175,191,152,150]

**Compute the Mean of the Height .**

**=(**168+170+150+160+182+140+175+191+152+150)/10=163.8



**Median :**

Median is mid value in this ordered data set.

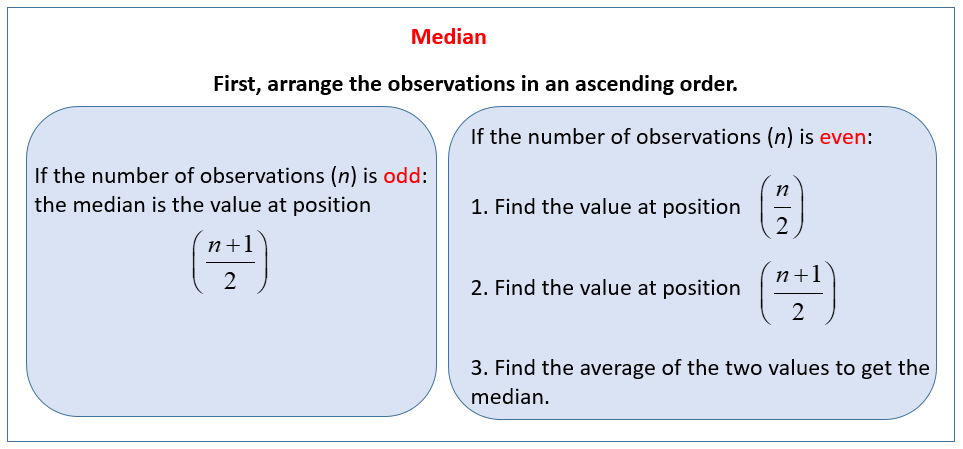
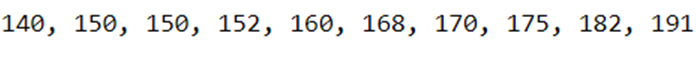
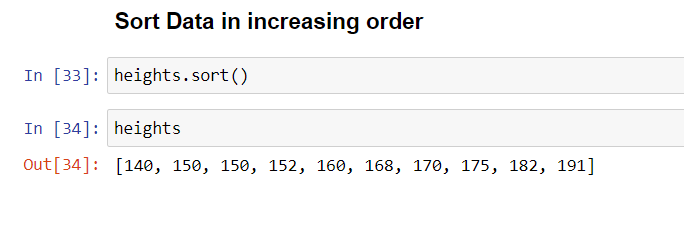


image:source unknown

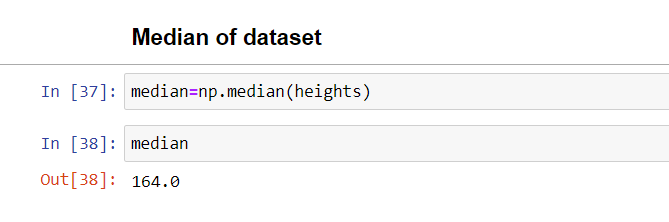
Arrange the data in the increasing order and then find the mid value.



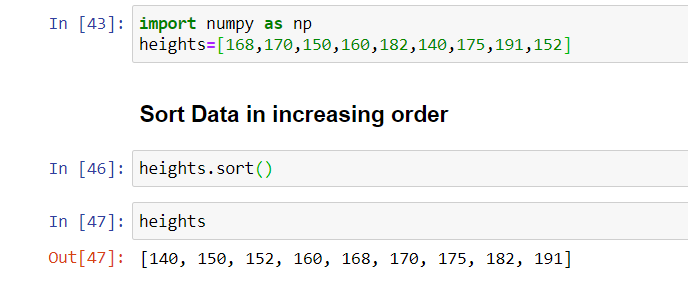
Median =160+168/2=164

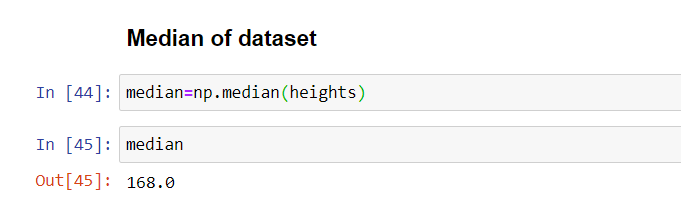


If we have even number of values in the data set then median is sum of mid two numbers divided by 2



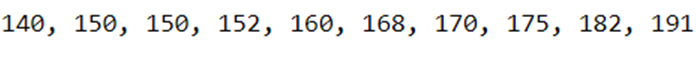
In we have odd number in the data set like below we have 9 heights the median will be 5th number value.

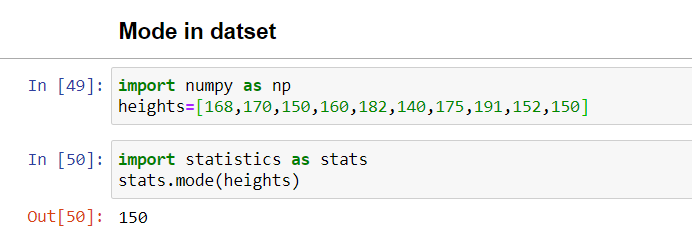




**Mode :**

Mode is the number which occur most often in the data set.Here 150 is occurring twice so this is our mode.

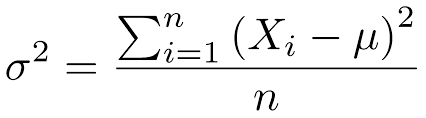




**Variance :**

Variance is the numerical values that describe the variability of the observations from its arithmetic mean and denoted by sigma-squared(σ2 )

Variance measure how far individuals in the group are spread out, in the set of data from the mean.



Where

Xi : Elements in the data set

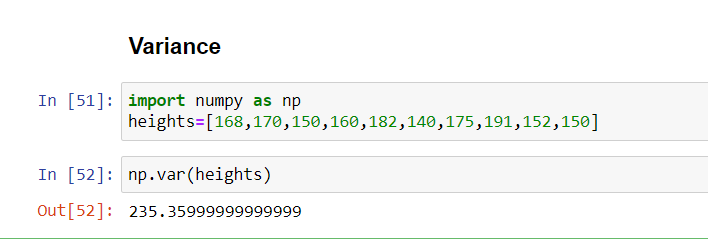
mu : the population mean

=[the population mean](https://www.google.com/search?sxsrf=ALeKk01Xep1wN5s8k6zttsrM8iSxc3DJxQ%3A1590843384306&q=Mean&stick=H4sIAAAAAAAAAOPgE-LQz9U3MCs3MlICs0zKCky0tLKTrfRTU0qTE0sy8_P00_KLcktzEq2gtEJmbmJ6qkJiXnF5atEjRmNugZc_7glLaU1ac_IaowoXV3BGfrlrXklmSaWQGBcblMUjxcUFt4BnESuLb2piHgBE6iKyfwAAAA&sa=X&ved=2ahUKEwiWhsK20dvpAhWCY80KHawnBxsQ24YFMAF6BAgEEAI)

Step 1: This formula says that take each element from dataset(population) and subtract from mean of data set.Later sum all the values.

Step 2: Take the sum in Step 1 and divide by total number of elements.

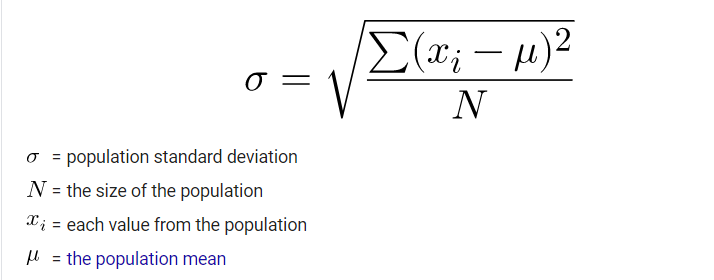
Square in the above formula will nullify the effect of negative sign(-)

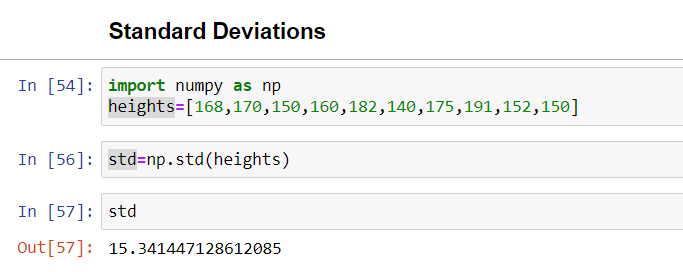


**Standard Deviation :**

It is a measure of dispersion of observation within dataset relative to their mean.It is square root of the variance and denoted by Sigma (σ) .

Standard deviation is expressed in the same unit as the values in the dataset so it measure how much observations of the data set differs from its mean.





**Interquartile of Data**

**import numpy as np**

**#define array of data**

**data = np.array([14, 19, 20, 22, 24, 26, 27, 30, 30, 31, 36, 38, 44, 47])**

**#calculate interquartile range**

**q3, q1 = np.percentile(data, [75 ,25])**

**iqr = q3 - q1**

**#display interquartile range**

**iqr**

**12.25**

**import numpy as np**

**import pandas as pd**

**#create data frame**

**df = pd.DataFrame({'rating': [90, 85, 82, 88, 94, 90, 76, 75, 87, 86],**

**'points': [25, 20, 14, 16, 27, 20, 12, 15, 14, 19],**

**'assists': [5, 7, 7, 8, 5, 7, 6, 9, 9, 5],**

**'rebounds': [11, 8, 10, 6, 6, 9, 6, 10, 10, 7]})**

**#define function to calculate interquartile range**

**def find\_iqr(x):**

**return np.subtract(\*np.percentile(x, [75, 25]))**

**#calculate IQR for 'rating' and 'points' columns**

**df[['rating', 'points']].apply(find\_iqr)**

**rating 6.75**

**points 5.75**

**dtype: float64**

**#calculate IQR for all columns**

**df.apply(find\_iqr)**

**rating 6.75**

**points 5.75**

**assists 2.50**

**rebounds 3.75**

**dtype: float64**

**STEP 2:**

Write a programming code with a procedure to compute mean mode and

median, variance, standard deviation, interquartile of a given sample

data without using readymade function.

1. # Python program to print mean of elements

# list of elements to calculate mean

n\_num = [168,170,150,160,182,140,175,191,152,150]

n = len(n\_num)

get\_sum = sum(n\_num)

mean = get\_sum / n

print("Mean / Average is: " + str(mean))

**2)# Find Median of data observations**

n\_num = [168,170,150,160,182,140,175,191,152,150]

n = len(n\_num)

n\_num.sort()

if n % 2 == 0:

    median1 = n\_num[n//2]

    median2 = n\_num[n//2 - 1]

    median = (median1 + median2)/2

else:

    median = n\_num[n//2]

print("Median is: " + str(median))

1. **Find Mode of Data without Readymade Functions:**

# Python program to print mode of elements from collections import Counter

# list of elements to calculate mode

n\_num = [168,170,150,160,182,140,175,191,152,150]

n = len(n\_num)

 data = Counter(n\_num)

get\_mode = dict(data)

mode = [k for k, v in get\_mode.items() if v == max(list(data.values()))]

 if len(mode) == n:

    get\_mode = "No mode found"

else:

    get\_mode = "Mode is / are: "

+ ', '.join(map(str, mode))

 print(get\_mode)

1. **Find standard Deviations and variance :**

**def** calculate\_std\_dev**(**lst**):**

mean **=** sum**(**lst**)** **/** len**(**lst**)**

variance **=** sum**((**xi **-** mean**)** **\*\*** **2** **for** xi **in** lst**)** **/** len**(**lst**)**

std\_dev **=** variance **\*\*** **0.5**

**return** std\_dev

STEP 3:

Compute and analyze the group data to find out frequency, mean, mode and median, standard deviation, variance, interquartile of data?

EXAMPLE:

**Find the mean, mode, and median for the following data,**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Class** | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | Total |
| **Frequency** | 8 | 16 | 36 | 34 | 6 | 100 |

**Solution:**

*We have,*

| **Class** | **Mid Value *x*i** | **Frequency fi** | **Cumulative Frequency** | **fi .*x*i** | x \*x | f \* x\*x |
| --- | --- | --- | --- | --- | --- | --- |
| 0-10 | 5 | 8 | 8 | 40 | 25 | 200 |
| 10-20 | 15 | 16 | 24 | 240 | 225 | 3600 |
| 20-30 | 25 | 36 | 60 | 900 | 625 | 22500 |
| 30-40 | 35 | 34 | 94 | 1190 | 1225 | 41650 |
| 40-50 | 45 | 6 | 100 | 270 | 2025 | 12150 |
|  |  | ∑fi=100 |  | ∑fi. xi=2640 | ∑ 4125 | ∑ 80100 |

***Mean = ∑(fi.xi)/∑f***

*= 2640/100*

*= 26.4*

*Here, N = 100 ⇒ N / 2 = 50.*

*Cumulative frequency just greater than 50 is 60 and corresponding class is 20-30.*

*Thus, the median class is 20-30.*

*Hence, l = 20, h = 10, f = 36, c =  c. f. of preceding class = 24 and N/2=50*

***Median, Me = l + h{(N/2 – cf)/f}***

*= 20+10{(50-24)/36}*

*Median = 27.2.*

***Mode = 3(median) – 2(mean)****= (3 × 27.2 – 2 × 26.4) = 28.8.*

***Standard Deviation and Variance of Group Data*** *:*

*Variance= (80100 / 100) – (2640/100)2*

*=801-696.96= 104.04*

*Standard Deviation= 10.2*

**Program:**

**import** pandas **as** pd *# Load pandas library*

Let’s also create some example data in Python:

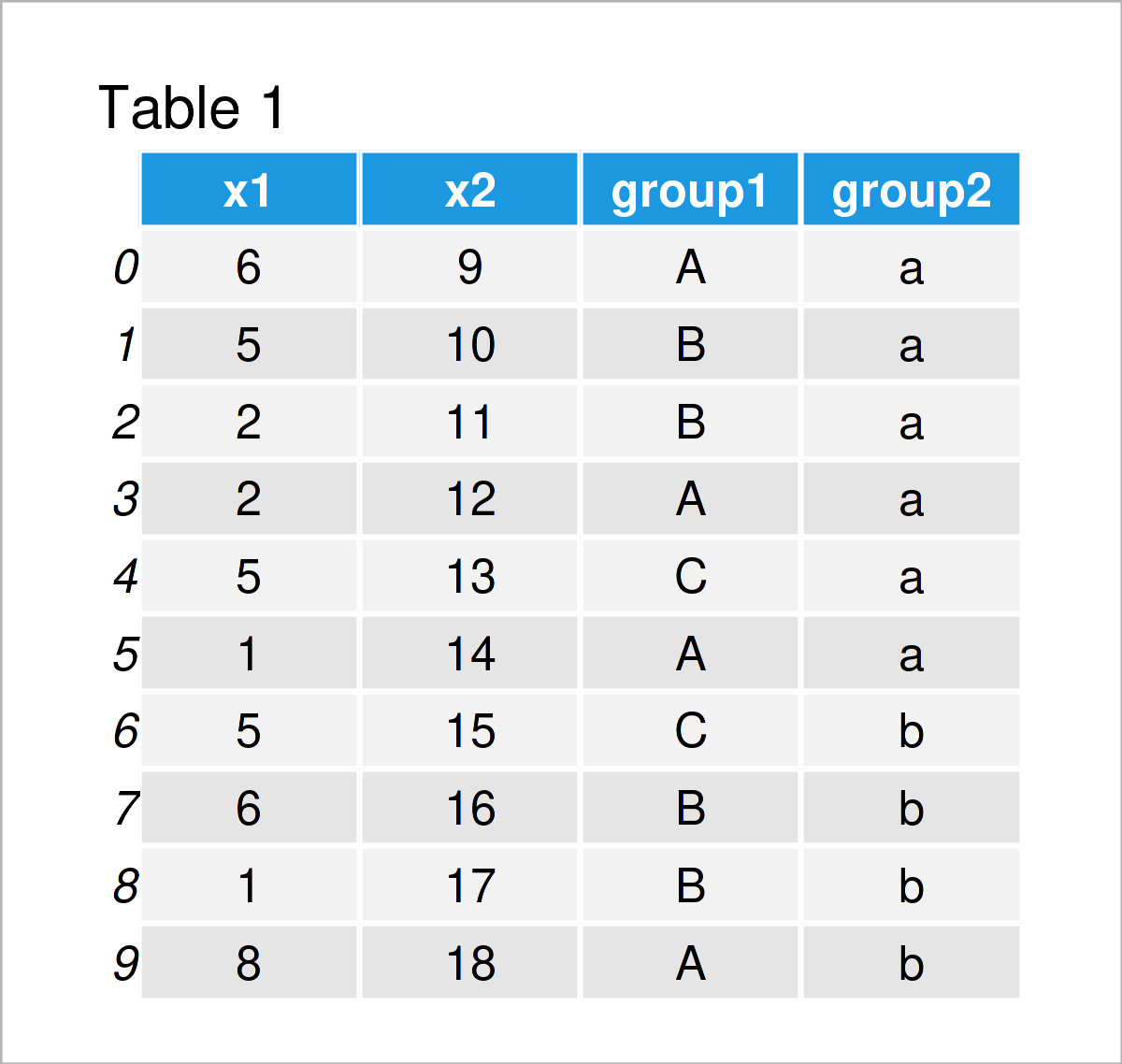
data = pd.DataFrame({'x1':[6, 5, 2, 2, 5, 1, 5, 6, 1, 8], *# Create pandas DataFrame*

'x2':range(9, 19),

'group1':['A', 'B', 'B', 'A', 'C', 'A', 'C', 'B', 'B', 'A'],

'group2':['a', 'a', 'a', 'a', 'a', 'a', 'b', 'b', 'b', 'b']})

**print**(data) *# Print pandas DataFrame*



**print**(data.groupby('group').mean())

**print**(data.groupby('group1').median())

**print**(data['x1'].mode()) # mode of one attribute

**print**(data.groupby('group1').std())

**print**(data.groupby('group1').var())

**q75, q25 = np.percentile(data['x1'], [75 ,25])**

**iqr = q75 - q25**

**print(“IQR”,iqr)**