Problem Statement 1:

You survey households in your area to find the average rent they are paying. Find the

standard deviation from the following data: 1550,1700, 900,850, 1000,950.

In [2]:
```
import statistics as stat

rent =[1550, 1700, 900, 850, 1000, 950]

rent_stdev= stat.stdev(rent)

#Standard deviation for House Hold Rent is
standardDeviation = round(rent_stdev,2)

print("Standard deviation of household rent is $"+str(standardDeviation))
```

```
Standard deviation of household rent is $367.99
```

Problem Statement 1 :

You survey households in your area to find the average rent they are paying. Find the standard deviation from the following data:

1550,1700, 900,850, 1000,950.

Solution:

A : Calculate Mean

The formula to find the sample mean is: = (Σ xi) / n

Σ of xi = (1550 + 1700 + 900 + 850 + 1000 + 950)/6 = $1158.33 6

The Mean is $1158.33

B : Calculate Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma\ (x - \bar{x})^2}{n}}$$

σ = lower case sigma
Σ = capital sigma
$\bar{x}$ = x bar

Where :

s means 'standard deviation'. S means 'the sum of'. X means 'the mean'

Step 1 : Calculate mean

mean = (1550 + 1700 + 900 + 850 + 1000 + 950)/6 = $1158.33

Step 2 : Subtract the mean calculated from step 1 from each value. This gives you the differences:

1550−1158.33 = $391.67

1700−1158.33 = $541.67

900−1158.33 = -$258.33

850−1158.33 = -$308.33

1000−1158.33 = $158.33

950−1158.33 = $208.33

Step 3: Square the differences you found in Step 3:

$391.672 = 153405.3889

$541.672 = 293406.3889

-$258.332 = 66734.3889

-$308.332 = 95067.3889

$158.332 = 25068.3889

In [3]:
```
#Problem 2. Find the variance for the following set of data representing trees in C
alifornia (heights in feet): 3, 21, 98, 203, 17, 9

trees =[3, 21, 98, 203, 17, 9]

trees_variance= stat.variance(trees)

result = round(trees_variance,2)
print("Variance  for trees in California is $" +str(result))
```

```
Variance  for trees in California is $6219.9
```

Variance Formula:

$$\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

σ = **lower case sigma**
$\sum$ = **capital sigma**
$\overline{x}$ = **x bar**

Step 1: Add up the numbers in your given data set. 3 + 21 + 98 + 203 + 17 + 9 = 351

Step 2: Square your answer: 351 × 351 = 123,201 …and divide by the number of items. We have 6 items in our example so: 123,201 / 6 = 20,533.5

Step 3: Square each item in the data set & get the sum of squares 3 × 3 + 21 × 21 + 98 × 98 + 203 × 203 + 17 × 17 + 9 × 9

Add those numbers (the squares) together: 9 + 441 + 9604 + 41209 + 289 + 81 = 51,633

Step 4: Subtract the value calculated in Step 2 from the the value of Step 3. 51,633 – 20,533.5 = 31,099.5

Step 5: Subtract 1 from the number of items in your data set*. For our example: 6 – 1 = 5

Step 6: Divide the number in Step 4 by the number in Step 5. This gives you the variance: 31,099.5 / 5 = 6,219.9

In [4]:
```
#Problem 3. In a class on 100 students, 80 students passed in all subjects, 10 fail
ed in one subject, 7 failed in two subjects and 3 failed in three subjects. Find th
e probability distribution of the variable for number of subjects a student from th
e given class has failed in.

import pandas as pd

#probabilites distributions
zero_sub_failed=80/100
one_sub_failed= 10/100
two_sub_failed=7/100
three_sub_failed=3/100

#dataframe
prob_df =pd.DataFrame({'failed_subject':[0,1,2,3],'probabilities_of_student_failed'
:[zero_sub_failed,one_sub_failed,two_sub_failed,three_sub_failed]})

#probabilities distributions
prob_df
```

Out[4]:

	failed_subject	probabilities_of_student_failed
0	0	0.80
1	1	0.10
2	2	0.07
3	3	0.03

For a random student,

The probability of failing in 0 subjects, P(X=0) = 0.8

The probability of failing in 1 subjects, P(X=1) = 0.1

The probability of failing in 2 subjects, P(X=2) = 0.07

The probability of failing in 3 subjects, P(X=3) = 0.03

The probability distribution can be shown as:

$$\sigma = \sqrt{\frac{\sum (x - \overline{x})^2}{n}}$$

σ = **lower case sigma**
$\sum$ = **capital sigma**
$\overline{x}$ = **x bar**

In [ ]: