Problem Statement

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 [1]:
```
from math import sqrt
import statistics as stat



def find_mean(arr):
    return sum(arr)/len(arr)

def find_variance(arr):
    variance = 0.0
    mean = find_mean(arr)
    variance = sum((x - mean)**2 for x in arr) / (len(arr)-1)
    return variance

def find_std_dev(arr):
    return find_variance(arr)**0.5

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

print("Calculation using custom method")
print ("Mean " + str(find_mean(arr)))
print ("variance " + str(find_variance(arr)))
print("std_dev " + str(find_std_dev(arr)))

print("Calculation using statistics library")

print(str("Mean " + str(stat.mean(arr))))
print(str("Variance " + str(stat.variance(arr))))
print(str("std_dev " + str(stat.stdev(arr))))
```

```
Calculation using custom method
Mean 1158.3333333333333
variance 135416.6666666667
std_dev 367.99003609699366
Calculation using statistics library
Mean 1158.3333333333333
Variance 135416.66666666666
std_dev 367.9900360969936
```

Problem Statement 2:

Find the variance for the following set of data representing trees in California (heights in feet):

3, 21, 98, 203, 17, 9

In [2]:
```
arr = [3, 21, 98, 203, 17, 9]

print("Calculation using custom method")
print ("Mean " + str(find_mean(arr)))
print ("variance " + str(find_variance(arr)))
print("std_dev " + str(find_std_dev(arr)))

print("Calculation using statistics library")

print(str("Mean " + str(stat.mean(arr))))
print(str("Variance " + str(stat.variance(arr))))
print(str("std_dev " + str(stat.stdev(arr))))
```

```
Calculation using custom method
Mean 58.5
variance 6219.9
std_dev 78.86634263106157
Calculation using statistics library
Mean 58.5
Variance 6219.9
std_dev 78.86634263106157
```

Problem Statement 3:

In a class on 100 students, 80 students passed in all subjects, 10 failed in one subject, 7 failed in two subjects and 3 failed in three subjects. Find the probability distribution of the variable for number of subjects a student from the given class has failed in.

In [3]:
```
# Probabability distribution for the studuent of a class.
import pandas as pd

COLUMN_NAMES = ('Failed_In_Subject','P(Failed_In_Subject)', 'No. of Students')
df = pd.DataFrame(columns=COLUMN_NAMES)

df['Failed_In_Subject'] =[0,1, 2, 3]
df['P(Failed_In_Subject)'] =[80/100,10/100,7/100,3/100]
df['No. of Students'] =[80, 10, 7, 3]
df
```

Out[3]:

	Failed_In_Subject	P(Failed_In_Subject)	No. of Students
0	0	0.80	80
1	1	0.10	10
2	2	0.07	7
3	3	0.03	3