Exemplo de aplicação componentes principais 3.1 - Dados relativos às empresas

In [43]:
```
import numpy as np
import scipy as sp

data = np.genfromtxt('dados_brutos_cap3_mingoti.csv',delimiter= ';', skip_header = 1, usecols= (1,2,3))

print('Importing DataSet \n Gross P  Net G Heritage')
print(data)
```

```
Importing DataSet
 Gross P  Net G Heritage
[[ 9893.   564. 17689.]
 [ 8776.   389. 17359.]
 [13572.  1103. 18597.]
 [ 6455.   743.  8745.]
 [ 5129.   203. 14397.]
 [ 5432.   215.  3467.]
 [ 3807.   385.  4679.]
 [ 3423.   187.  6754.]
 [ 3708.   127.  2275.]
 [ 3294.   297.  6754.]
 [ 5433.   432.  5589.]
 [ 6287.   451.  8972.]]
```

In [39]:
```
print('Descriptive Stats \n')
print('Means',np.mean(data, 0),'\n')
print('Standard Deviations',np.std(data, 0),'\n')
print('Minimums',np.min(data, 0),'\n')
print('Maximums',np.max(data, 0),)
```

```
Descriptive Stats

Means [6267.41666667  424.66666667 9606.41666667]

Standard Deviations [2958.83838295  264.41203116 5616.1170076 ]

Minimums [3294.  127. 2275.]

Maximums [13572.  1103. 18597.]
```

In [40]:
```
covMatrix = np.cov(data, rowvar = False)
print('Covariance Matrix \n', covMatrix)
```

```
Covariance Matrix
 [[ 9550608.62878788   706121.06060606 14978232.53787879]
 [  706121.06060606    76269.51515152   933915.06060606]
 [14978232.53787879   933915.06060606 34408112.99242424]]
```

In [41]:
```
eigVal, eigVec = np.linalg.eig(covMatrix)
print('Eigenvalues :', eigVal)
print('\n Eigvectors :\n', eigVec)
```

```
Eigenvalues : [4.14743915e+07 2.53950713e+06 2.10925256e+04]

 Eigvectors :
[[ 0.42509725  0.8997068   0.09909593]
 [ 0.02766083  0.09651661 -0.99494694]
 [ 0.90472493 -0.42569029 -0.01614231]]
```

In [42]:
```
percentageExplainedByComponent = eigVal / np.sum(eigVal)
print(percentageExplainedByComponent)
```

```
[9.41850797e-01 5.76702086e-02 4.78994659e-04]
```

In [54]:
```
varArray = np.diag(covMatrix)

Component_Variable_correlation_matrix = eigVec * np.sqrt(eigVal) / np.sqrt(varArray[:, np.newaxis])

print('Correlation between components and original variables: \n    Component 1     Component 2     Component 3\n',Component_Variable_correlation_matrix)
```

```
Correlation between components and original variables:
    Component 1     Component 2     Component 3
 [[ 8.85855291e-01  4.63938267e-01  4.65698236e-03]
 [ 6.45029692e-01  5.56930874e-01 -5.23225283e-01]
 [ 9.93290182e-01 -1.15647974e-01 -3.99668177e-04]]
```

In [58]:
```
firstComponentScores = data @ eigVec[:,0]

print('First Component Scores: \n', firstComponentScores)
```

```
First Component Scores:
 [20224.76716464 19446.53366209 22625.09937764 10676.37429662
 15211.26381976  5451.75669095  5862.20261865  7570.79266768
  3638.02275626  7518.99781347  7378.01049745 10802.25355641]
```

In [64]:
```
print('Ranking: \n')
print('1st Place: ', np.max(firstComponentScores) , ' Company №:', np.where(firstComponentScores == np.max(firstComponentScores))[0] )
print('Last Place: ', np.min(firstComponentScores), ' Company №:', np.where(firstComponentScores == np.min(firstComponentScores))[0] )
```

```
Ranking:

1st Place:  22625.09937764474  Company №: [2]
Last Place:  3638.0227562603063  Company №: [8]
```