ds1_E_coli_limpieza_de_datos

February 9, 2021

Limpieza de datos

[1]:
```
import pandas as pd
import seaborn as sns
import numpy as np
import os
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
%matplotlib inline
from mlxtend.preprocessing import standardize
from scipy import stats
```

1 Declaración de variables

[2]:
```
organismo ="E_coli"
dataset = 1
nombre = ("ds" + str(dataset) + "_" + str(organismo))
nombre2 = (str(organismo)+ " dataset " + str(dataset))
r2 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_filtrar")
r3 = ("Datos/resultados/"+ str(organismo) + "/" + str(nombre) + "/
 ↪transformaciones/sin_atipicos")

nom1 = ("/ds" + str(dataset) + "_AAC_efectores_" + str(organismo) + ".txt")
nom2 = ("/ds" + str(dataset) + "_ACC_hidro_mass_efectores_" + str(organismo) +␣
 ↪".txt")
nom3 = ("/ds" + str(dataset) + "_ACC_mass_efectores_" + str(organismo) + ".txt")
nom4 = ("/ds" + str(dataset) + "_ACC_hidro_efectores_" + str(organismo) + ".
 ↪txt")
nom5 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_efectores_" + str(organismo)␣
 ↪+ ".txt")
nom6 = ("/ds" + str(dataset) + "_PseAAC_mass_efectores_" + str(organismo) + ".
 ↪txt")
nom7 = ("/ds" + str(dataset) + "_PseAAC_hidro_efectores_" + str(organismo) + ".
 ↪txt")
```

```
nom8 = ("/ds" + str(dataset) + "_AAC_no_efectores_" + str(organismo) + ".txt")
nom9 = ("/ds" + str(dataset) + "_ACC_hidro_mass_no_efectores_" + str(organismo)␣
 ↪+ ".txt")
nom10 = ("/ds" + str(dataset) + "_ACC_mass_no_efectores_" + str(organismo) + ".
 ↪txt")
nom11 = ("/ds" + str(dataset) + "_ACC_hidro_no_efectores_" + str(organismo) + ".
 ↪txt")
nom12 = ("/ds" + str(dataset) + "_PseAAC_hidro_mass_no_efectores_" +␣
 ↪str(organismo) + ".txt")
nom13 = ("/ds" + str(dataset) + "_PseAAC_mass_no_efectores_" + str(organismo) +␣
 ↪".txt")
nom14 = ("/ds" + str(dataset) + "_PseAAC_hidro_no_efectores_" + str(organismo)␣
 ↪+ ".txt")

#Efectores
AAC_efec= pd.read_csv(str(r2) + str(nom1), header=None,prefix='X',sep=',')
ACC_hidro_mass_efec = pd.read_csv(str(r2) + str(nom2),␣
 ↪header=None,prefix='X',sep=',')
ACC_mass_efec = pd.read_csv(str(r2) + str(nom3), header=None,prefix='X',sep=',')
ACC_hidro_efec = pd.read_csv(str(r2) + str(nom4),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_mass_efec = pd.read_csv(str(r2) +str(nom5),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_mass_efec = pd.read_csv(str(r2) + str(nom6),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_efec = pd.read_csv(str(r2) + str(nom7),␣
 ↪header=None,prefix='X',sep=',')

#No efectores
AAC_no_efec= pd.read_csv(str(r2) + str(nom8), header=None,prefix='X',sep=',')
ACC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom9),␣
 ↪header=None,prefix='X',sep=',')
ACC_mass_no_efec =pd.read_csv(str(r2) + str(nom10),␣
 ↪header=None,prefix='X',sep=',')
ACC_hidro_no_efec =pd.read_csv(str(r2) + str(nom11),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_mass_no_efec =pd.read_csv(str(r2) + str(nom12),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_mass_no_efec =pd.read_csv(str(r2) + str(nom13),␣
 ↪header=None,prefix='X',sep=',')
PseAAC_hidro_no_efec =pd.read_csv(str(r2) + str(nom14),␣
 ↪header=None,prefix='X',sep=',')
```

2 Composición de aminoácidos (AAC)

[3]:
```
transf = "Composición de aminoácidos (AAC) "
etiq="efectores "
estado = "con valores atípicos.\n"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf) + str(etiq) + " " + str(nombre2) + ", " +str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    #del df['X20']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq) +" dataset "+ str(dataset)+"␣
 ↪"+str(transf)+" "+str(estado))
```

```
efectores
Composición de aminoácidos (AAC) efectores E_coli dataset 1, con valores
atípicos.
Valores del documento csv.

          X0     X1      X2     X3     X4     X5     X6     X7     X8     X9  \
0      9.302  5.426   2.713  8.527  2.326  5.814  4.264  8.527  2.326  8.527
1     12.927  5.366   2.439  4.146  2.683  6.341  4.634  8.049  2.927  5.610
2     14.559  8.429   1.149  3.831  0.000  4.215  4.598  3.448  4.215  3.448
3      1.695  5.085  11.864  6.780  3.390  5.085  1.695  5.085  3.390  1.695
4     12.879  5.303   1.515  2.273  0.909  0.909  3.182  9.394  1.212  6.667
…       …    …       …      …      …      …      …      …      …
4995  11.565  7.029   3.401  5.442  0.680  8.163  4.082  5.669  2.494  6.349
4996  10.924  4.202   3.782  7.143  0.000  7.563  1.261  4.622  0.840  8.403
4997  10.037  5.204   3.532  6.877  0.000  8.736  3.346  6.320  2.045  4.461
4998  10.204  6.122   2.915  7.580  1.458  4.082  4.373  7.289  2.915  2.915
4999  10.448  6.343   2.239  5.597  1.493  4.851  3.358  7.836  3.731  2.985
```

```
      …    X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0     …  5.426  1.550  3.101  2.326  4.264  4.264  1.550  2.326  10.465
1     …  2.683  5.366  3.902  5.366  4.146  4.634  1.951  1.951   6.098
2     …  4.215  4.215  2.299  6.513  4.215  7.280  1.533  3.065   7.663
3     …  3.390  3.390  6.780  3.390  5.085  8.475  0.000  8.475   3.390
4     …  0.758  4.697  3.788  4.697  5.455  5.909  3.182  0.758   7.576
…   …    …    …    …    …    …    …    …    …      …
4995  …  3.855  2.041  2.494  3.628  5.215  5.442  0.454  1.587   7.710
4996  …  8.824  2.521  5.462  3.361  8.824  4.622  0.000  4.202   7.563
4997  …  5.390  2.602  5.019  2.788  5.019  5.762  0.558  2.416   8.736
4998  …  3.790  2.041  5.831  4.373  6.997  5.248  0.875  2.041   8.455
4999  …  3.731  2.239  2.985  3.731  7.836  6.716  2.612  1.493   7.463

             X20
0      efectores
1      efectores
2      efectores
3      efectores
4      efectores
…          …
4995   efectores
4996   efectores
4997   efectores
4998   efectores
4999   efectores

[5000 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      8.947401     5.667111     4.551238     5.298438     1.215573
std       3.150435     2.541771     2.280021     2.030737     1.194666
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.969000     4.000000     2.973750     4.132000     0.395000
50%       8.830000     5.478500     4.219000     5.366000     0.999000
75%      10.777250     7.002250     5.685000     6.538500     1.695000
max      32.075000    31.034000    18.421000    17.431000    12.500000

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      5.585747     4.269495     7.366592     2.248358     5.688120
std       2.435125     1.981494     2.963352     1.344856     2.446714
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	3.715750	2.949250	5.455000	1.342000	4.030750
50%	5.582500	4.093500	7.117000	2.105000	5.431000
75%	7.102500	5.350000	9.045000	2.961250	6.925000
max	20.000000	21.212000	39.706000	15.789000	20.588000

	X10	X11	X12	X13	X14 \
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	9.736381	4.684147	2.635471	3.801719	4.117632
std	3.323317	2.411583	1.334528	1.850007	1.895829
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.477000	2.985000	1.717000	2.597000	2.817750
50%	9.378000	4.255000	2.387000	3.571000	4.107000
75%	11.836250	5.978000	3.235250	4.651000	5.224250
max	33.333000	21.667000	16.216000	16.393000	17.216000

	X15	X16	X17	X18	X19
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	6.710978	5.918568	1.635413	3.176615	6.745052
std	2.385342	2.245609	1.141674	1.651562	2.186234
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	5.123250	4.520000	0.839000	2.020000	5.434000
50%	6.444000	5.682000	1.471000	2.994500	6.737500
75%	8.000000	7.050250	2.273000	4.118000	8.042000
max	17.391000	29.703000	10.309000	12.069000	22.727000

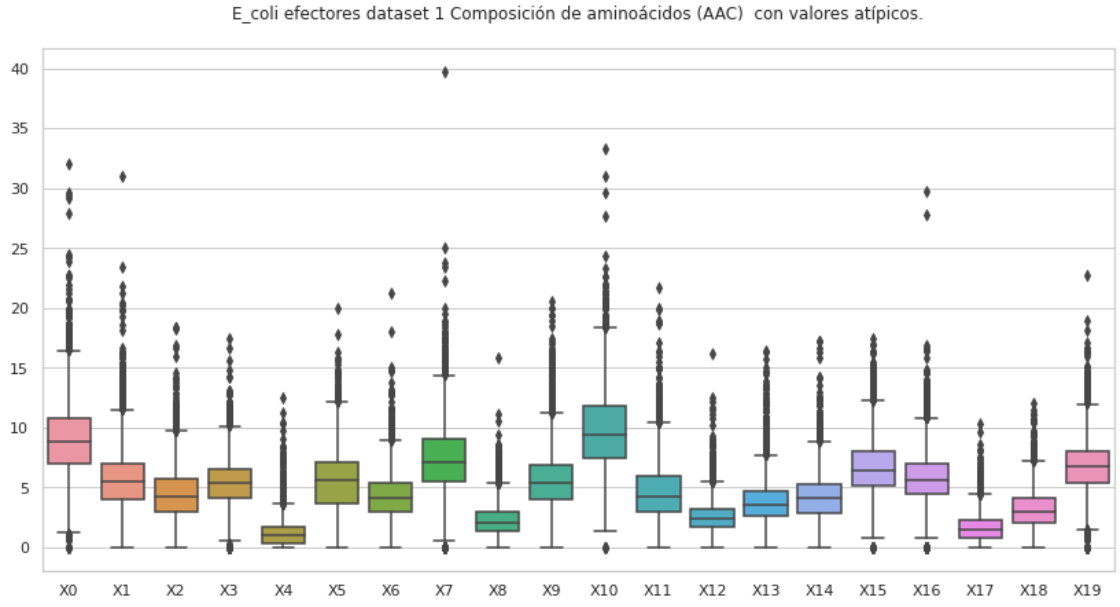
no_efectores
Composición de aminoácidos (AAC) no_efectores E_coli dataset 1, con valores atípicos.
Valores del documento csv.

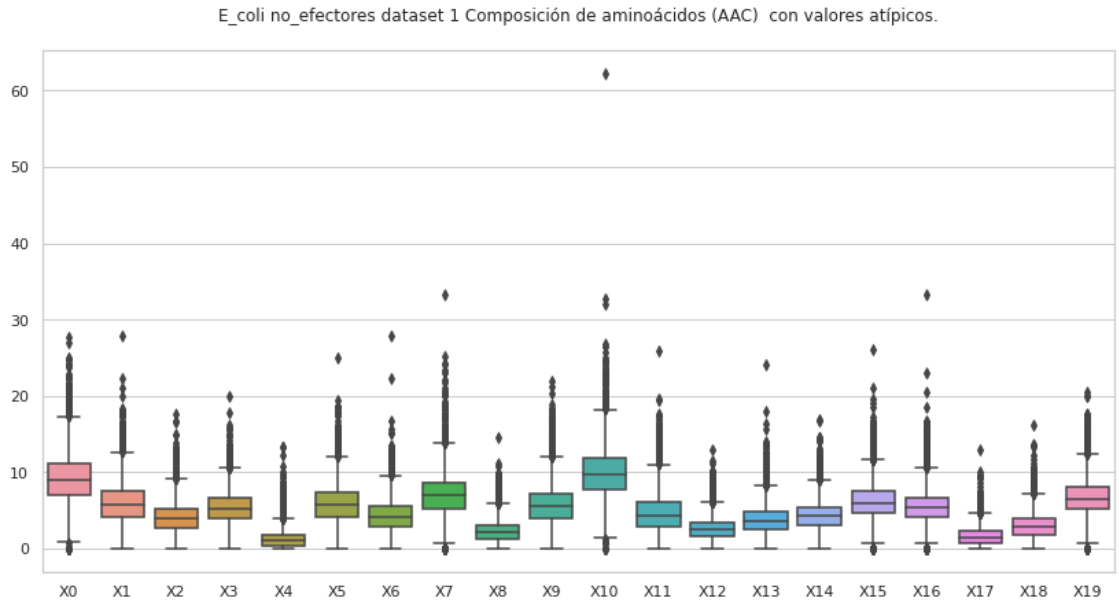
	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	9.944	6.443	4.622	5.322	0.980	6.022	4.622	6.443	2.521
1	9.816	5.521	4.294	6.135	2.761	7.055	3.374	6.748	3.067
2	5.479	3.653	5.479	5.936	0.457	10.502	3.653	9.132	2.740
3	6.250	4.545	5.114	7.386	0.568	4.545	4.545	5.682	2.273
4	7.339	8.257	1.835	4.587	0.917	8.257	3.670	5.505	1.835
…	…	…	…	…	…	…	…	…	…
4995	8.571	4.286	7.143	5.714	0.000	7.143	4.286	8.571	4.286
4996	13.514	3.784	4.324	6.486	1.622	8.108	4.324	7.027	2.703
4997	8.148	12.593	2.222	5.185	0.000	2.963	3.704	11.111	2.222
4998	8.955	8.955	2.985	1.493	0.000	5.970	7.463	10.448	1.493
4999	13.333	3.333	4.444	2.222	0.000	4.444	2.222	10.000	1.111

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	5.882	…	3.501	2.661	4.622	3.081	6.303	6.863	0.840	5.182

```
1      5.215  …  4.294  3.067  2.454  5.828   3.988   4.908  1.534  2.761
2      6.393  …  3.653  2.283  6.393  3.196   5.479   5.023  2.283  3.653
3      5.682  …  7.386  3.409  2.841  1.705  10.795   3.977  1.136  3.409
4      4.587  …  5.505  1.835  4.587  5.505   8.257   5.505  0.917  5.505
…      …  …    …    …    …    …    …      …    …    …
4995  10.000  …  4.286  1.429  2.857  7.143   4.286   2.857  0.000  4.286
4996   7.568  …  4.865  1.081  2.703  3.243   5.946   2.703  2.162  2.703
4997   1.481  …  5.185  3.704  5.185  4.444   5.926   5.185  0.000  5.185
4998   1.493  …  4.478  4.478  5.970  1.493   7.463  10.448  0.000  1.493
4999  13.333  …  3.333  2.222  2.222  1.111   5.556  11.111  0.000  2.222

        X19            X20
0     4.762   no_efectores
1     7.362   no_efectores
2     3.653   no_efectores
3     7.955   no_efectores
4     7.339   no_efectores
…     …            …
4995  2.857   no_efectores
4996  8.108   no_efectores
4997  8.889   no_efectores
4998  8.955   no_efectores
4999  8.889   no_efectores

[5000 rows x 21 columns]


Composición de aminoácidos (AAC) no_efectores E_coli dataset 1, con valores
atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      9.126593     5.975243     4.163099     5.260636     1.283059
std       3.441492     2.861889     2.173088     2.224394     1.278754
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       6.946750     4.071750     2.720000     3.900000     0.352750
50%       8.939500     5.714000     3.888500     5.300000     1.091500
75%      11.111000     7.525500     5.298000     6.574000     1.802000
max      27.778000    27.907000    17.699000    20.000000    13.333000

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      5.817490     4.332539     7.111522     2.295351     5.816542
std       2.652878     2.221258     2.860333     1.553002     2.669588
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.134750     2.844000     5.256500     1.223750     4.036750
50%       5.747000     4.167000     6.941000     2.105000     5.506000
```

```
75%       7.326000     5.556000     8.696000     3.125000     7.211000
max      25.000000    27.941000    33.333000    14.634000    21.978000

                X10          X11          X12          X13          X14  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     10.014388     4.751616     2.661510     3.806274     4.340256
std       3.390843     2.687996     1.491687     2.024944     2.003887
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.752000     2.899000     1.653000     2.518000     3.077000
50%       9.756000     4.348000     2.448000     3.571000     4.255000
75%      11.951750     6.173000     3.414250     4.839000     5.443000
max      62.222000    25.926000    12.903000    24.138000    16.858000

                X15          X16          X17          X18          X19
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      6.260959     5.613770     1.606203     3.067200     6.695827
std       2.422484     2.343114     1.264238     1.795012     2.455682
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.762000     4.120000     0.738000     1.835000     5.145750
50%       6.023500     5.389500     1.441000     2.830000     6.516000
75%       7.519500     6.739000     2.296000     4.000000     8.081000
max      26.087000    33.333000    12.903000    16.216000    20.455000
```





2.1 Composición de aminoácidos (AAC), sin valores atípicos

[4]:
```
transf = "Composición de aminoácidos (AAC) "
estado = "sin valores atípicos.\n"
transf2="AAC"

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' +␣
 ↪str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf) + str(etiq) + " " + str(nombre2) + ", " +str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=AAC_efec

    if etiq == "no_efectores":
        df=AAC_no_efec

    del df['X20']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
 ↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
```

```
    df['X20'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)



    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq) +" dataset "+ str(dataset)+"␣
↪"+str(transf))
```

```
efectores
Composición de aminoácidos (AAC) efectores E_coli dataset 1, sin valores
atípicos.
Valores del documento csv.

          X0     X1     X2     X3     X4     X5     X6     X7     X8     X9  \
0      9.302  5.426  2.713  8.527  2.326  5.814  4.264  8.527  2.326  8.527
1     12.927  5.366  2.439  4.146  2.683  6.341  4.634  8.049  2.927  5.610
2     14.559  8.429  1.149  3.831  0.000  4.215  4.598  3.448  4.215  3.448
4     12.879  5.303  1.515  2.273  0.909  0.909  3.182  9.394  1.212  6.667
5     11.250  6.750  3.000  5.750  1.000  5.250  5.750  8.000  3.750  4.750
…        …    …      …      …      …      …      …      …      …      …
4995  11.565  7.029  3.401  5.442  0.680  8.163  4.082  5.669  2.494  6.349
4996  10.924  4.202  3.782  7.143  0.000  7.563  1.261  4.622  0.840  8.403
4997  10.037  5.204  3.532  6.877  0.000  8.736  3.346  6.320  2.045  4.461
4998  10.204  6.122  2.915  7.580  1.458  4.082  4.373  7.289  2.915  2.915
4999  10.448  6.343  2.239  5.597  1.493  4.851  3.358  7.836  3.731  2.985

      …    X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0     …  5.426  1.550  3.101  2.326  4.264  4.264  1.550  2.326  10.465
1     …  2.683  5.366  3.902  5.366  4.146  4.634  1.951  1.951   6.098
2     …  4.215  4.215  2.299  6.513  4.215  7.280  1.533  3.065   7.663
4     …  0.758  4.697  3.788  4.697  5.455  5.909  3.182  0.758   7.576
5     …  3.250  2.500  3.250  4.000  6.000  3.000  2.000  2.000   6.000
…   …    …      …      …      …      …      …      …      …       …
4995  …  3.855  2.041  2.494  3.628  5.215  5.442  0.454  1.587   7.710
```

```
4996  …  8.824  2.521  5.462  3.361  8.824  4.622  0.000  4.202   7.563
4997  …  5.390  2.602  5.019  2.788  5.019  5.762  0.558  2.416   8.736
4998  …  3.790  2.041  5.831  4.373  6.997  5.248  0.875  2.041   8.455
4999  …  3.731  2.239  2.985  3.731  7.836  6.716  2.612  1.493   7.463

            X20
0     efectores
1     efectores
2     efectores
4     efectores
5     efectores
…         …
4995  efectores
4996  efectores
4997  efectores
4998  efectores
4999  efectores

[4276 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4276.000000  4276.000000  4276.000000  4276.000000  4276.000000
mean      9.036647     5.612884     4.547577     5.391152     1.114203
std       2.814727     2.127889     2.065838     1.759758     0.917953
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.194000     4.144750     3.053000     4.335000     0.427000
50%       8.947000     5.525000     4.238000     5.466500     0.986000
75%      10.811000     6.949000     5.637250     6.541000     1.619000
max      18.333000    13.287000    11.268000    11.307000     4.762000

                X5           X6           X7           X8           X9  \
count  4276.000000  4276.000000  4276.000000  4276.000000  4276.000000
mean      5.646005     4.304722     7.489703     2.235789     5.579283
std       2.199749     1.767777     2.684318     1.163465     2.103142
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.020000     3.119500     5.668000     1.399000     4.077000
50%       5.701000     4.156000     7.263000     2.115500     5.395000
75%       7.102000     5.350250     9.091000     2.941000     6.796000
max      12.782000    10.135000    16.250000     6.280000    12.953000

               X10          X11          X12          X13          X14  \
count  4276.000000  4276.000000  4276.000000  4276.000000  4276.000000
mean      9.750458     4.591335     2.538961     3.698220     4.145017
```

std	3.028652	2.123230	1.106456	1.516195	1.662302
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.571500	3.015000	1.734000	2.642750	2.973000
50%	9.448000	4.238500	2.354500	3.546000	4.167000
75%	11.818000	5.854250	3.125000	4.523000	5.224000
max	19.394000	11.765000	6.620000	9.343000	9.778000

	X15	X16	X17	X18	X19
count	4276.000000	4276.000000	4276.00000	4276.000000	4276.000000
mean	6.724088	5.979605	1.61435	3.182652	6.817392
std	2.210610	1.962765	0.98633	1.471689	1.887385
min	0.000000	0.000000	0.00000	0.000000	0.694000
25%	5.191000	4.669750	0.89575	2.094750	5.602500
50%	6.466000	5.763500	1.47800	3.019000	6.808500
75%	7.937750	7.056750	2.21600	4.086250	8.040000
max	13.725000	12.543000	4.95900	8.108000	13.281000

no_efectores
Composición de aminoácidos (AAC) no_efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	\
0	9.944	6.443	4.622	5.322	0.980	6.022	4.622	6.443	2.521	
1	9.816	5.521	4.294	6.135	2.761	7.055	3.374	6.748	3.067	
2	5.479	3.653	5.479	5.936	0.457	10.502	3.653	9.132	2.740	
3	6.250	4.545	5.114	7.386	0.568	4.545	4.545	5.682	2.273	
4	7.339	8.257	1.835	4.587	0.917	8.257	3.670	5.505	1.835	
…	…	…	…	…	…	…	…	…	…	
4995	8.571	4.286	7.143	5.714	0.000	7.143	4.286	8.571	4.286	
4996	13.514	3.784	4.324	6.486	1.622	8.108	4.324	7.027	2.703	
4997	8.148	12.593	2.222	5.185	0.000	2.963	3.704	11.111	2.222	
4998	8.955	8.955	2.985	1.493	0.000	5.970	7.463	10.448	1.493	
4999	13.333	3.333	4.444	2.222	0.000	4.444	2.222	10.000	1.111	

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18	\
0	5.882	…	3.501	2.661	4.622	3.081	6.303	6.863	0.840	5.182	
1	5.215	…	4.294	3.067	2.454	5.828	3.988	4.908	1.534	2.761	
2	6.393	…	3.653	2.283	6.393	3.196	5.479	5.023	2.283	3.653	
3	5.682	…	7.386	3.409	2.841	1.705	10.795	3.977	1.136	3.409	
4	4.587	…	5.505	1.835	4.587	5.505	8.257	5.505	0.917	5.505	
…	…	…	…	…	…	…	…	…	…	…	
4995	10.000	…	4.286	1.429	2.857	7.143	4.286	2.857	0.000	4.286	
4996	7.568	…	4.865	1.081	2.703	3.243	5.946	2.703	2.162	2.703	
4997	1.481	…	5.185	3.704	5.185	4.444	5.926	5.185	0.000	5.185	
4998	1.493	…	4.478	4.478	5.970	1.493	7.463	10.448	0.000	1.493	

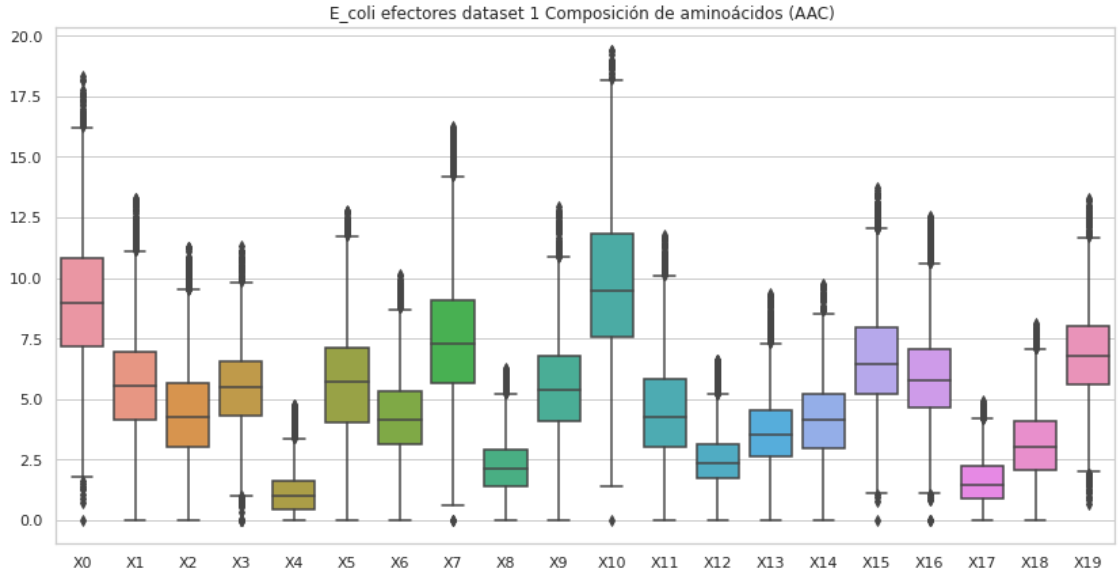
```
4999  13.333  …  3.333  2.222  2.222  1.111   5.556  11.111  0.000  2.222

        X19           X20
0     4.762  no_efectores
1     7.362  no_efectores
2     3.653  no_efectores
3     7.955  no_efectores
4     7.339  no_efectores
…     …           …
4995  2.857  no_efectores
4996  8.108  no_efectores
4997  8.889  no_efectores
4998  8.955  no_efectores
4999  8.889  no_efectores

[4181 rows x 21 columns]
```

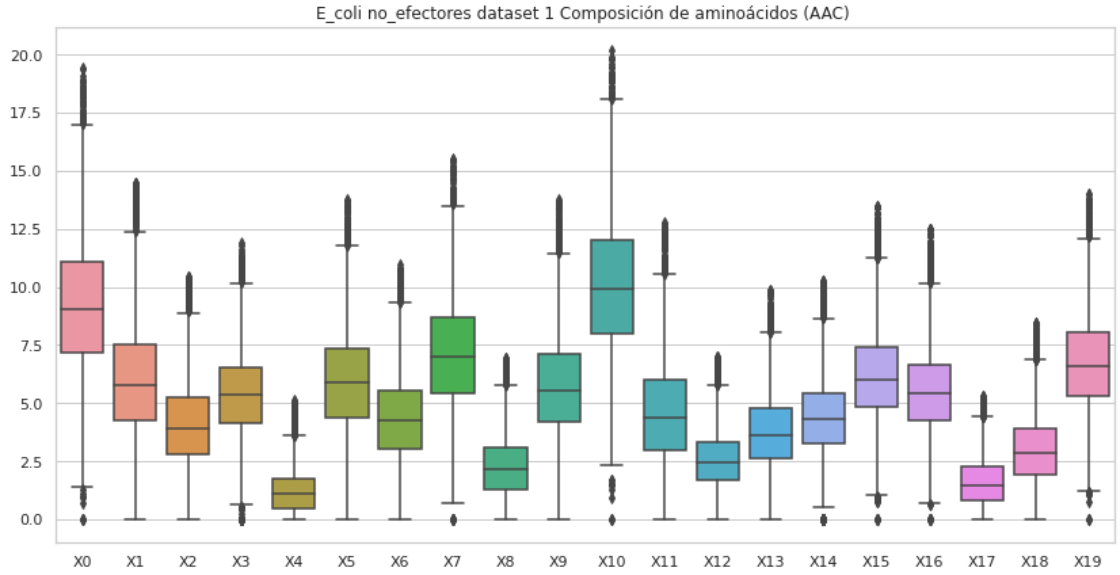
Composición de aminoácidos (AAC) no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4181.00000  4181.000000  4181.000000  4181.000000  4181.000000
mean      9.21118     6.001000     4.123324     5.335465     1.195560
std       3.07128     2.525378     1.924003     1.951540     0.980236
min       0.00000     0.000000     0.000000     0.000000     0.000000
25%       7.19000     4.245000     2.786000     4.145000     0.457000
50%       9.04900     5.764000     3.883000     5.382000     1.087000
75%      11.11100     7.500000     5.231000     6.557000     1.724000
max      19.40300    14.483000    10.417000    11.905000     5.109000

                X5           X6           X7           X8           X9  \
count  4181.000000  4181.000000  4181.000000  4181.000000  4181.000000
mean      5.925465     4.385929     7.106071     2.285668     5.789039
std       2.349632     1.948389     2.464433     1.352278     2.319878
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.396000     3.030000     5.435000     1.309000     4.202000
50%       5.882000     4.271000     7.000000     2.151000     5.537000
75%       7.345000     5.560000     8.667000     3.093000     7.107000
max      13.750000    10.959000    15.544000     6.944000    13.793000

               X10          X11          X12          X13          X14  \
count  4181.000000  4181.000000  4181.000000  4181.000000  4181.000000
mean     10.113286     4.669280     2.592499     3.784619     4.373652
std       2.982179     2.369549     1.264125     1.740172     1.741357
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       7.979000     2.985000     1.695000     2.597000     3.261000
```

50%	9.929000	4.348000	2.439000	3.599000	4.322000
75%	12.030000	6.024000	3.333000	4.776000	5.423000
max	20.181000	12.766000	6.977000	9.865000	10.256000

	X15	X16	X17	X18	X19
count	4181.000000	4181.000000	4181.000000	4181.000000	4181.00000
mean	6.214377	5.560315	1.580205	3.056645	6.69648
std	2.109966	1.977944	1.086461	1.587359	2.14930
min	0.000000	0.000000	0.000000	0.000000	0.00000
25%	4.839000	4.260000	0.820000	1.942000	5.29800
50%	6.034000	5.398000	1.449000	2.857000	6.56400
75%	7.407000	6.627000	2.260000	3.922000	8.02800
max	13.452000	12.500000	5.319000	8.434000	14.01300





3 Composición de pseudo aminoácidos (PseAAC) hidro_mass

[5]:
```
#hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro_mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    #del df['X83']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores E_coli dataset
1, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.045200	0.011300	0.041434	0.028250	0.015067	0.041434	0.011300	
1	0.064510	0.013389	0.020692	0.031646	0.019475	0.040166	0.014606	
2	0.091344	0.000000	0.024038	0.026442	0.014423	0.021634	0.026442	
3	0.048033	0.096065	0.192131	0.144098	0.192131	0.144098	0.096065	
4	0.025321	0.001787	0.004468	0.001787	0.007447	0.018469	0.002383	
…	…	…	…	…	…	…	…	
4995	0.052964	0.003116	0.024924	0.037386	0.011424	0.025963	0.011424	
4996	0.043545	0.000000	0.028472	0.030147	0.021773	0.018423	0.003350	
4997	0.053306	0.000000	0.036524	0.046396	0.026653	0.033563	0.010859	
4998	0.053765	0.007681	0.039939	0.021506	0.030723	0.038403	0.015361	
4999	0.050486	0.007212	0.027046	0.023440	0.014425	0.037865	0.018031	

	X7	X8	X9	…	X74	X75	X76	\
0	0.041434	0.026367	0.033900	…	-0.011852	-0.003246	0.020068	
1	0.027995	0.013389	0.043818	…	-0.008206	-0.006394	0.015215	
2	0.021634	0.026442	0.069710	…	0.021551	0.018045	-0.015204	
3	0.048033	0.096065	0.336229	…	-0.319554	-0.135437	-0.081256	
4	0.013107	0.001489	0.037236	…	0.011702	0.003193	0.016512	
…	…	…	…	…	…	…	…	
4995	0.029078	0.017655	0.058156	…	0.015248	0.024914	0.025931	
4996	0.033496	0.035171	0.023447	…	-0.012895	-0.006741	0.013362	
4997	0.023691	0.028627	0.059229	…	0.003540	0.012698	0.020761	
4998	0.015361	0.019970	0.055301	…	0.001695	0.018552	0.049042	
4999	0.014425	0.018031	0.059502	…	0.014920	0.008888	0.036817	

	X77	X78	X79	X80	X81	X82	X83
0	0.018190	0.015577	0.038584	0.028168	0.009180	0.047022	efectores
1	-0.004059	-0.014764	0.036190	0.007852	-0.000704	0.026010	efectores
2	0.005217	0.001941	0.021966	-0.015517	-0.029748	0.017434	efectores
3	-0.131643	-0.184049	0.080941	0.341261	-0.163305	0.145854	efectores
4	0.019998	0.004710	0.016537	0.013478	0.001854	0.014946	efectores
…	…	…	…	…	…	…	
4995	0.015953	0.002756	0.027128	0.013843	0.011870	0.013407	efectores
4996	0.042277	0.039718	-0.003367	0.004456	0.011392	0.015079	efectores
4997	-0.006376	0.001709	0.021238	0.000518	-0.005744	0.023917	efectores

```
4998  0.000113 -0.012108  0.024367  0.009208 -0.011859  0.027308  efectores
4999  0.011170 -0.001495  0.043895  0.002571 -0.002621  0.026541  efectores

[5000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.047404	0.008049	0.030166	0.034140	0.022845	
std	0.048624	0.017799	0.035651	0.062364	0.041001	
min	-0.933934	-0.166480	-0.860839	-0.810496	-0.648397	
25%	0.031100	0.001478	0.017416	0.015100	0.010792	
50%	0.042630	0.004712	0.027508	0.028935	0.017083	
75%	0.056576	0.009715	0.038283	0.044681	0.027028	
max	1.326359	0.663179	1.224938	3.674814	1.326359	

	X5	X6	X7	X8	X9	…	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	…	
mean	0.038535	0.014535	0.033177	0.028829	0.057636	…	
std	0.077660	0.040083	0.066924	0.078118	0.116392	…	
min	-1.134695	-0.332960	-1.134695	-0.665920	-1.634384	…	
25%	0.026050	0.005070	0.016659	0.012733	0.029417	…	
50%	0.035744	0.010395	0.025645	0.021374	0.046085	…	
75%	0.045424	0.017798	0.038559	0.034787	0.069760	…	
max	4.899752	2.449876	3.674814	4.899752	6.124691	…	

	X73	X74	X75	X76	X77	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.016291	0.003082	0.003094	0.014830	-0.000042	
std	0.118099	0.091827	0.067376	0.073350	0.068413	
min	-7.517625	-1.293403	-3.370023	-3.221425	-2.250656	
25%	0.006862	-0.009457	-0.005381	0.004493	-0.010534	
50%	0.018730	0.004104	0.003456	0.018362	0.001872	
75%	0.030741	0.015990	0.013572	0.028874	0.013565	
max	1.246506	4.744352	1.373594	2.084603	1.557752	

	X78	X79	X80	X81	X82
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	0.001647	0.018302	-0.000445	0.002453	0.014406
std	0.062114	0.099242	0.188250	0.077345	0.127197
min	-2.260451	-1.497417	-11.884098	-2.937708	-7.715546
25%	-0.006049	0.007166	-0.011525	-0.006750	0.006185
50%	0.003598	0.018806	0.002541	0.002576	0.018277
75%	0.013393	0.029472	0.014639	0.012978	0.029470

```
max       1.154733     5.035852     2.370572     1.898779     1.451150

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores E_coli
dataset 1, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.086962	0.008574	0.046543	0.052667	0.040419	0.056342	0.022047	
1	0.056433	0.015872	0.035270	0.040561	0.014108	0.038798	0.017635	
2	0.031904	0.002659	0.034562	0.061149	0.037221	0.053173	0.015952	
3	0.037219	0.003384	0.043986	0.027068	0.016918	0.033835	0.013534	
4	0.044735	0.005592	0.027960	0.050327	0.027960	0.033551	0.011184	
…	…	…	…	…	…	…	…	
4995	0.047675	0.000000	0.031783	0.039729	0.015892	0.047675	0.023837	
4996	0.058937	0.007072	0.028290	0.035362	0.011787	0.030647	0.011787	
4997	0.038361	0.000000	0.024412	0.013950	0.024412	0.052311	0.010462	
4998	0.051716	0.000000	0.008619	0.034477	0.034477	0.060335	0.008619	
4999	0.026888	0.000000	0.004481	0.008963	0.004481	0.020166	0.002241	

	X7	X8	X9	…	X74	X75	X76	\
0	0.051442	0.030620	0.082063	…	0.025759	-0.012842	0.018303	
1	0.029980	0.024689	0.056433	…	0.000831	0.024385	0.032282	
2	0.037221	0.021269	0.063807	…	0.001751	-0.030175	0.033240	
3	0.033835	0.043986	0.064287	…	-0.002959	0.027884	-0.025142	
4	0.027960	0.033551	0.050327	…	-0.074230	-0.036206	-0.012170	
…	…	…	…	…	…	…	…	
4995	0.055621	0.023837	0.055621	…	-0.032732	-0.031095	-0.008924	
4996	0.033005	0.021217	0.030647	…	0.017492	0.004633	0.022221	
4997	0.006975	0.024412	0.031387	…	0.028207	0.003213	0.026990	
4998	0.008619	0.025858	0.034477	…	0.013151	0.020171	0.006889	
4999	0.026888	0.006722	0.017926	…	-0.010914	-0.010184	0.010544	

	X77	X78	X79	X80	X81	X82	X83
0	0.011259	0.006807	-0.007402	0.006795	0.000572	0.016165	no_efectores
1	0.019009	0.006265	0.047636	0.024499	0.013085	0.014052	no_efectores
2	0.000621	0.022056	0.037254	0.037652	0.027530	0.046316	no_efectores
3	-0.026518	-0.003399	0.045201	0.006863	0.024850	0.031627	no_efectores
4	0.004873	-0.018783	-0.020137	0.009259	0.029253	0.002517	no_efectores
…	…	…	…	…	…	…	
4995	0.010499	-0.001878	0.030026	0.015800	0.076642	0.026401	no_efectores
4996	-0.008209	-0.022378	-0.016801	0.012457	-0.002190	0.017950	no_efectores
4997	-0.010521	0.023489	0.018119	0.002861	0.011365	-0.008047	no_efectores
4998	-0.039370	-0.048876	0.070014	-0.000036	0.030935	0.046062	no_efectores

```
4999  0.003793 -0.006053  0.012871 -0.001430  0.001663  0.015715  no_efectores

[5000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores E_coli
dataset 1, con valores atípicos.
Estadísticas.

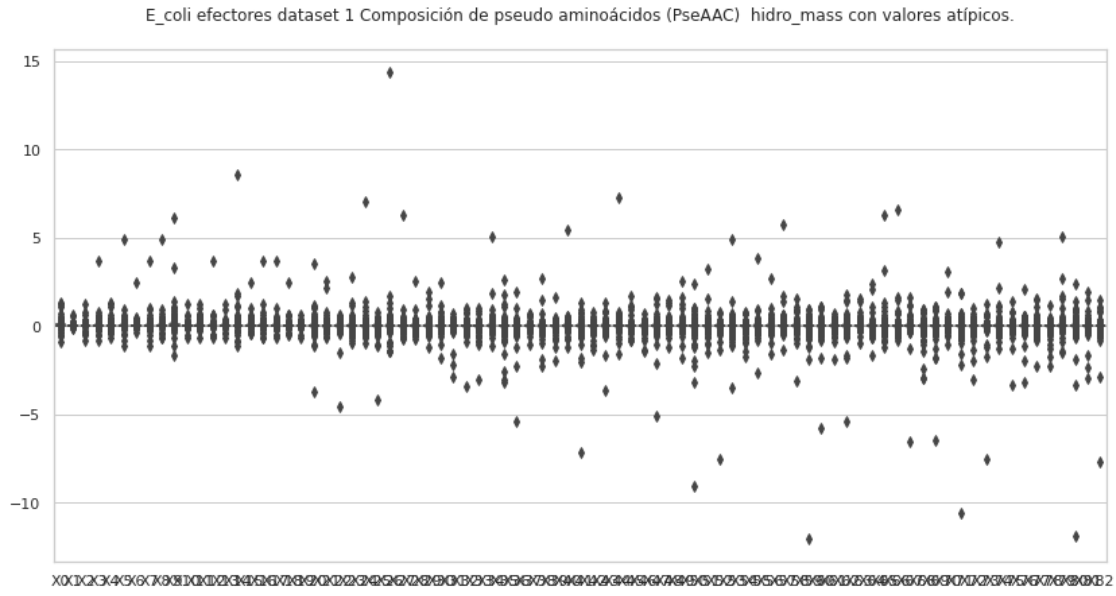
	X0	X1	X2	X3	X4	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.049311	0.007209	0.031965	0.035273	0.022876	
std	0.083654	0.037755	0.041953	0.077786	0.043589	
min	-3.445671	-1.647762	-1.722835	-4.019949	-1.632004	
25%	0.031944	0.001302	0.016839	0.017900	0.010716	
50%	0.045051	0.005031	0.028167	0.031209	0.017951	
75%	0.060521	0.010498	0.042033	0.048247	0.028913	
max	1.387563	0.272409	0.925042	0.925042	1.233389	

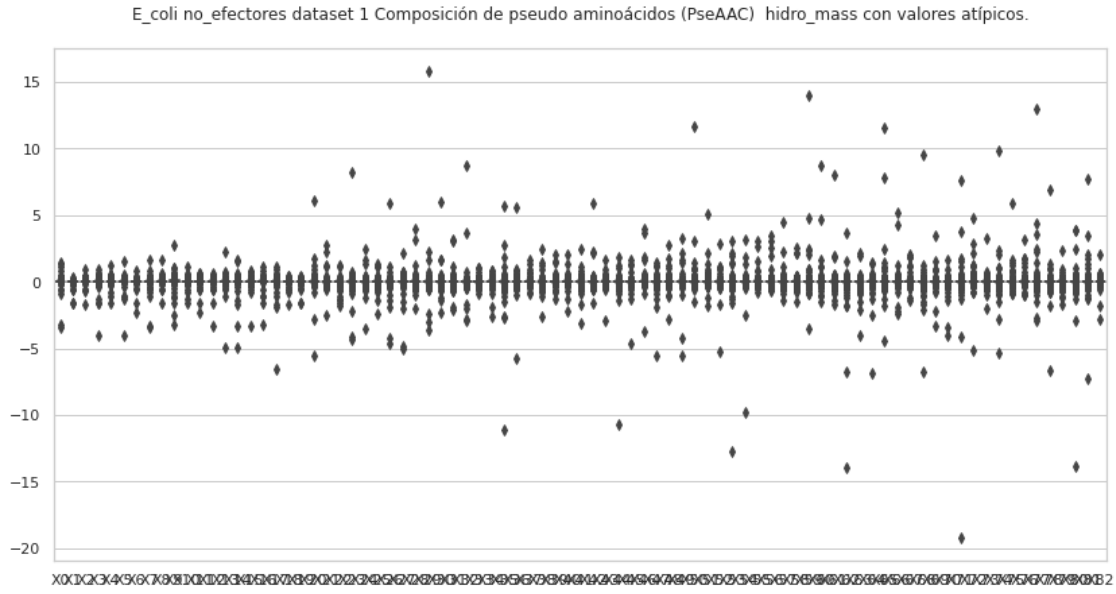
	X5	X6	X7	X8	X9	…	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	…	
mean	0.037849	0.014323	0.032914	0.029096	0.057800	…	
std	0.073824	0.047112	0.082257	0.045788	0.093113	…	
min	-4.080010	-2.297114	-3.445671	-1.647762	-3.264008	…	
25%	0.024485	0.004852	0.017372	0.012340	0.032900	…	
50%	0.035003	0.010848	0.027472	0.022430	0.049889	…	
75%	0.047241	0.019457	0.042430	0.037661	0.074144	…	
max	1.541737	0.784603	1.634451	1.646647	2.744412	…	

	X73	X74	X75	X76	X77	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.014811	0.001144	0.005820	0.016150	0.004723	
std	0.070561	0.179025	0.103593	0.075806	0.220744	
min	-2.025771	-5.320100	-0.887609	-0.978009	-2.947529	
25%	0.002235	-0.013768	-0.008420	0.001873	-0.012595	
50%	0.017369	0.003200	0.003331	0.016656	0.003538	
75%	0.030172	0.016934	0.017268	0.030038	0.017830	
max	3.291455	9.840829	5.915962	3.179144	12.923401	

	X78	X79	X80	X81	X82
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	0.003114	0.014777	-0.000406	0.005471	0.014265
std	0.147375	0.065318	0.224332	0.170752	0.069576
min	-6.656670	-1.742069	-13.829176	-7.234831	-2.838828
25%	-0.007612	0.003307	-0.012240	-0.007463	0.001740
50%	0.003662	0.017550	0.004223	0.003984	0.016813
75%	0.015653	0.030045	0.017596	0.016409	0.029991
max	6.901324	2.314508	3.895082	7.674034	2.061680

[8 rows x 83 columns]





3.1 Composición de pseudo aminoácidos (PseAAC) hidro_mass, sin valores atípicos

[6]:
```
#hidro_mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_mass_no_efec

    del df['X83']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X83'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)


    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
```

```
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass efectores E_coli dataset
1, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.045200	0.011300	0.041434	0.028250	0.015067	0.041434	0.011300	
1	0.064510	0.013389	0.020692	0.031646	0.019475	0.040166	0.014606	
2	0.091344	0.000000	0.024038	0.026442	0.014423	0.021634	0.026442	
4	0.025321	0.001787	0.004468	0.001787	0.007447	0.018469	0.002383	
5	0.060009	0.005334	0.030671	0.028004	0.017336	0.042673	0.020003	
…	…	…	…	…	…	…	…	
4995	0.052964	0.003116	0.024924	0.037386	0.011424	0.025963	0.011424	
4996	0.043545	0.000000	0.028472	0.030147	0.021773	0.018423	0.003350	
4997	0.053306	0.000000	0.036524	0.046396	0.026653	0.033563	0.010859	
4998	0.053765	0.007681	0.039939	0.021506	0.030723	0.038403	0.015361	
4999	0.050486	0.007212	0.027046	0.023440	0.014425	0.037865	0.018031	

	X7	X8	X9	…	X74	X75	X76	\
0	0.041434	0.026367	0.033900	…	-0.011852	-0.003246	0.020068	
1	0.027995	0.013389	0.043818	…	-0.008206	-0.006394	0.015215	
2	0.021634	0.026442	0.069710	…	0.021551	0.018045	-0.015204	
4	0.013107	0.001489	0.037236	…	0.011702	0.003193	0.016512	
5	0.025337	0.017336	0.068010	…	0.008664	-0.012196	-0.015447	
…	…	…	…	…	…	…	…	
4995	0.029078	0.017655	0.058156	…	0.015248	0.024914	0.025931	
4996	0.033496	0.035171	0.023447	…	-0.012895	-0.006741	0.013362	
4997	0.023691	0.028627	0.059229	…	0.003540	0.012698	0.020761	
4998	0.015361	0.019970	0.055301	…	0.001695	0.018552	0.049042	
4999	0.014425	0.018031	0.059502	…	0.014920	0.008888	0.036817	

	X77	X78	X79	X80	X81	X82	X83
0	0.018190	0.015577	0.038584	0.028168	0.009180	0.047022	efectores
1	-0.004059	-0.014764	0.036190	0.007852	-0.000704	0.026010	efectores
2	0.005217	0.001941	0.021966	-0.015517	-0.029748	0.017434	efectores
4	0.019998	0.004710	0.016537	0.013478	0.001854	0.014946	efectores
5	-0.012897	-0.013691	0.029463	-0.016415	-0.012756	0.016399	efectores
…	…	…	…	…	…	…	
4995	0.015953	0.002756	0.027128	0.013843	0.011870	0.013407	efectores
4996	0.042277	0.039718	-0.003367	0.004456	0.011392	0.015079	efectores
4997	-0.006376	0.001709	0.021238	0.000518	-0.005744	0.023917	efectores
4998	0.000113	-0.012108	0.024367	0.009208	-0.011859	0.027308	efectores
4999	0.011170	-0.001495	0.043895	0.002571	-0.002621	0.026541	efectores

[4798 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4798.000000	4798.000000	4798.000000	4798.000000	4798.000000	
mean	0.044714	0.006670	0.028661	0.031382	0.020170	
std	0.020088	0.007577	0.016609	0.021408	0.014193	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.030801	0.001488	0.017337	0.014777	0.010598	
50%	0.041987	0.004557	0.027203	0.028174	0.016752	
75%	0.055195	0.009089	0.037300	0.043150	0.025497	
max	0.184878	0.061046	0.136277	0.172541	0.143060	

	X5	X6	X7	X8	X9	…	\
count	4798.000000	4798.000000	4798.000000	4798.000000	4798.000000	…	
mean	0.036191	0.012624	0.029750	0.025221	0.051479	…	
std	0.015273	0.010755	0.019856	0.018216	0.030450	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.025866	0.005039	0.016426	0.012524	0.028857	…	
50%	0.035293	0.010119	0.025034	0.020626	0.044684	…	
75%	0.044522	0.016906	0.036481	0.033153	0.066852	…	
max	0.121897	0.094040	0.186715	0.114890	0.246503	…	

	X73	X74	X75	X76	X77	\
count	4798.000000	4798.000000	4798.000000	4798.000000	4798.000000	
mean	0.017881	0.002472	0.004042	0.016036	0.001023	
std	0.022324	0.028978	0.022979	0.022076	0.026636	
min	-0.117941	-0.254613	-0.143816	-0.137934	-0.194540	
25%	0.007471	-0.008646	-0.004911	0.005236	-0.009613	
50%	0.018738	0.004182	0.003463	0.018442	0.002013	
75%	0.030206	0.015631	0.013117	0.028557	0.013237	
max	0.184482	0.219951	0.182753	0.162290	0.188055	

	X78	X79	X80	X81	X82
count	4798.000000	4798.000000	4798.000000	4798.000000	4798.000000
mean	0.003725	0.017732	0.001677	0.003040	0.016690
std	0.022222	0.022009	0.027968	0.022922	0.023092
min	-0.146568	-0.128841	-0.253427	-0.165876	-0.145724
25%	-0.005416	0.007804	-0.010882	-0.006242	0.006911
50%	0.003660	0.018867	0.002580	0.002593	0.018419
75%	0.013059	0.029240	0.014173	0.012656	0.029097
max	0.181813	0.226724	0.221641	0.192726	0.226339

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores E_coli
dataset 1, sin valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.086962  0.008574  0.046543  0.052667  0.040419  0.056342  0.022047
1     0.056433  0.015872  0.035270  0.040561  0.014108  0.038798  0.017635
2     0.031904  0.002659  0.034562  0.061149  0.037221  0.053173  0.015952
3     0.037219  0.003384  0.043986  0.027068  0.016918  0.033835  0.013534
4     0.044735  0.005592  0.027960  0.050327  0.027960  0.033551  0.011184
…          …         …         …         …         …         …         …
4995  0.047675  0.000000  0.031783  0.039729  0.015892  0.047675  0.023837
4996  0.058937  0.007072  0.028290  0.035362  0.011787  0.030647  0.011787
4997  0.038361  0.000000  0.024412  0.013950  0.024412  0.052311  0.010462
4998  0.051716  0.000000  0.008619  0.034477  0.034477  0.060335  0.008619
4999  0.026888  0.000000  0.004481  0.008963  0.004481  0.020166  0.002241

            X7        X8        X9  …       X74       X75       X76  \
0     0.051442  0.030620  0.082063  …  0.025759 -0.012842  0.018303
1     0.029980  0.024689  0.056433  …  0.000831  0.024385  0.032282
2     0.037221  0.021269  0.063807  …  0.001751 -0.030175  0.033240
3     0.033835  0.043986  0.064287  … -0.002959  0.027884 -0.025142
4     0.027960  0.033551  0.050327  … -0.074230 -0.036206 -0.012170
…          …         …         …  …         …         …         …
4995  0.055621  0.023837  0.055621  … -0.032732 -0.031095 -0.008924
4996  0.033005  0.021217  0.030647  …  0.017492  0.004633  0.022221
4997  0.006975  0.024412  0.031387  …  0.028207  0.003213  0.026990
4998  0.008619  0.025858  0.034477  …  0.013151  0.020171  0.006889
4999  0.026888  0.006722  0.017926  … -0.010914 -0.010184  0.010544

            X77       X78       X79       X80       X81       X82           X83
0     0.011259  0.006807 -0.007402  0.006795  0.000572  0.016165  no_efectores
1     0.019009  0.006265  0.047636  0.024499  0.013085  0.014052  no_efectores
2     0.000621  0.022056  0.037254  0.037652  0.027530  0.046316  no_efectores
3    -0.026518 -0.003399  0.045201  0.006863  0.024850  0.031627  no_efectores
4     0.004873 -0.018783 -0.020137  0.009259  0.029253  0.002517  no_efectores
…          …         …         …         …         …         …         …
4995  0.010499 -0.001878  0.030026  0.015800  0.076642  0.026401  no_efectores
4996 -0.008209 -0.022378 -0.016801  0.012457 -0.002190  0.017950  no_efectores
4997 -0.010521  0.023489  0.018119  0.002861  0.011365 -0.008047  no_efectores
4998 -0.039370 -0.048876  0.070014 -0.000036  0.030935  0.046062  no_efectores
4999  0.003793 -0.006053  0.012871 -0.001430  0.001663  0.015715  no_efectores

[4831 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores E_coli
dataset 1, sin valores atípicos.
Estadísticas.

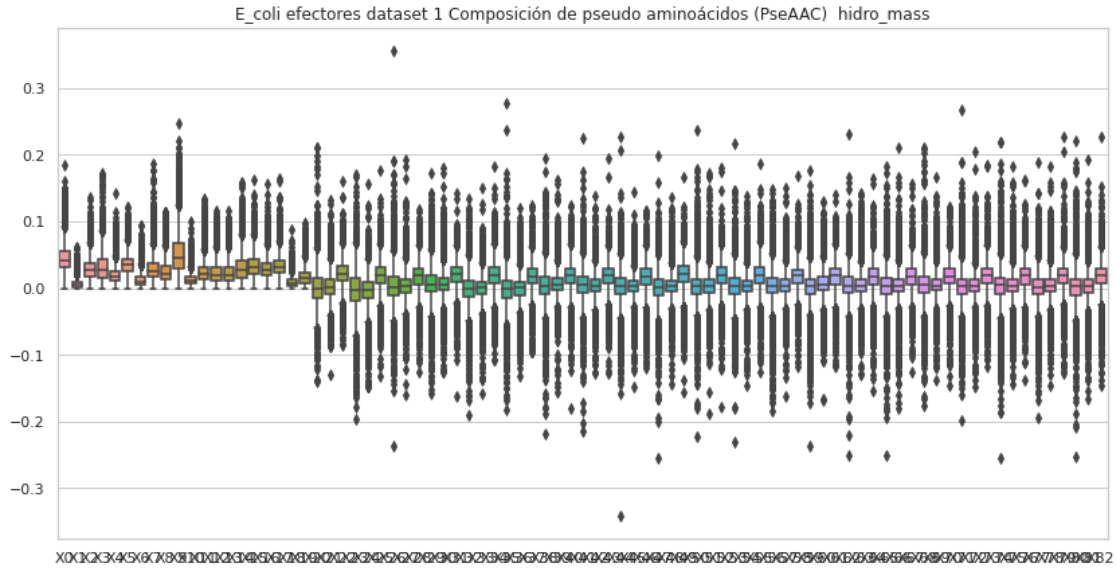
	X0	X1	X2	X3	X4	\
count	4831.000000	4831.000000	4831.000000	4831.000000	4831.000000	
mean	0.048205	0.007463	0.030628	0.034460	0.021457	
std	0.024080	0.009031	0.019933	0.024098	0.016356	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.031677	0.001372	0.016670	0.017516	0.010565	
50%	0.044515	0.004985	0.027767	0.030545	0.017596	
75%	0.059428	0.010029	0.040769	0.046460	0.027881	
max	0.216461	0.105327	0.136595	0.218551	0.147457	

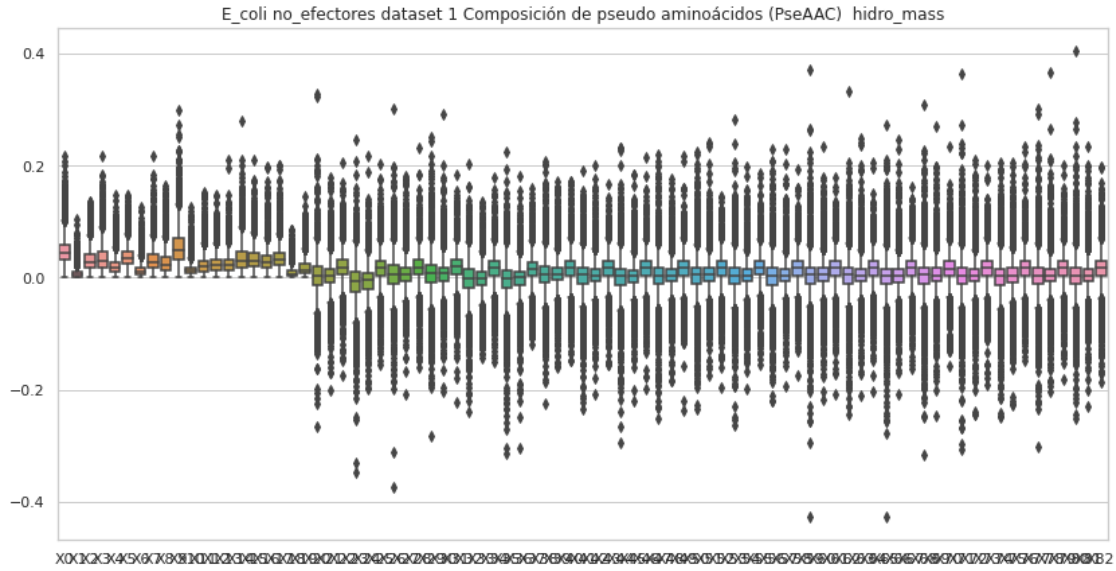
	X5	X6	X7	X8	X9	…	\
count	4831.000000	4831.000000	4831.000000	4831.000000	4831.000000	…	
mean	0.037122	0.013836	0.032128	0.027381	0.055652	…	
std	0.018410	0.013224	0.022079	0.021535	0.032328	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.024399	0.004833	0.017197	0.012164	0.032531	…	
50%	0.034595	0.010511	0.027064	0.021983	0.048794	…	
75%	0.045997	0.018737	0.040746	0.036230	0.071536	…	
max	0.148250	0.125980	0.218551	0.164454	0.299125	…	

	X73	X74	X75	X76	X77	\
count	4831.000000	4831.000000	4831.000000	4831.000000	4831.000000	
mean	0.015613	0.001050	0.003889	0.015534	0.002175	
std	0.027626	0.035562	0.030946	0.027277	0.035662	
min	-0.192448	-0.248808	-0.228503	-0.147231	-0.301054	
25%	0.003063	-0.013003	-0.007958	0.002844	-0.011755	
50%	0.017487	0.003259	0.003332	0.016804	0.003612	
75%	0.029881	0.016566	0.016639	0.029914	0.017421	
max	0.200085	0.205061	0.204494	0.234749	0.301312	

	X78	X79	X80	X81	X82
count	4831.000000	4831.000000	4831.000000	4831.000000	4831.000000
mean	0.004301	0.015807	0.001668	0.004628	0.014962
std	0.029867	0.027445	0.035662	0.029653	0.029464
min	-0.203875	-0.180709	-0.251168	-0.228709	-0.186360
25%	-0.006976	0.003969	-0.011437	-0.006962	0.002644
50%	0.003742	0.017621	0.004205	0.004014	0.016947
75%	0.015430	0.029815	0.017271	0.016016	0.029704
max	0.365763	0.201230	0.404784	0.235082	0.198711

[8 rows x 83 columns]


E_coli efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3


E_coli no_efectores dataset 1 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.4
0.2
0.0
−0.2
−0.4

4 Composición de pseudo aminoácidos (PseAAC) mass

[7]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

    #del df['X41']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  mass efectores E_coli dataset 1, con
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.047936  0.011984  0.043941  0.029960  0.015979  0.043941  0.011984
1     0.062270  0.012924  0.019974  0.030548  0.018799  0.038772  0.014099
2     0.100480  0.000000  0.026442  0.029086  0.015865  0.023798  0.029086
3     0.018143  0.036285  0.072570  0.054428  0.072570  0.054428  0.036285
4     0.046641  0.003292  0.008231  0.003292  0.013718  0.034021  0.004390
…          …         …         …         …         …         …         …
4995  0.057432  0.003378  0.027027  0.040540  0.012387  0.028153  0.012387
4996  0.065817  0.000000  0.043034  0.045566  0.032909  0.027846  0.005063
```

```
4997  0.057601  0.000000  0.039467  0.050134  0.028800  0.036267  0.011734
4998  0.051205  0.007315  0.038038  0.020482  0.029260  0.036575  0.014630
4999  0.052694  0.007528  0.028229  0.024465  0.015055  0.039520  0.018819

            X7        X8        X9  …       X32       X33       X34  \
0     0.043941  0.027963  0.035952  …  0.028108  0.004038  0.006488
1     0.027023  0.012924  0.042297  …  0.021838  0.021875  0.014707
2     0.023798  0.029086  0.076682  …  0.012381 -0.002042  0.031318
3     0.018143  0.036285  0.126998  … -0.024012  0.074452 -0.030741
4     0.024144  0.002744  0.068590  …  0.016617  0.036259  0.025357
…        …         …        …  …        …         …         …
4995  0.031531  0.019144  0.063063  …  0.044883  0.024445  0.026537
4996  0.050629  0.053160  0.035440  …  0.007278  0.018517  0.026783
4997  0.025600  0.030934  0.064001  …  0.032127  0.034771  0.022234
4998  0.014630  0.019019  0.052668  …  0.020530  0.030604  0.036961
4999  0.015055  0.018819  0.062104  …  0.027684  0.020413  0.019308

           X35       X36       X37       X38       X39       X40         X41
0     0.040213  0.027445  0.018954  0.021282  0.040919  0.049867  efectores
1     0.036204  0.024230  0.011190  0.014686  0.034933  0.025107  efectores
2    -0.019645  0.039036  0.005794 -0.016725  0.024163  0.019177  efectores
3     0.029378 -0.078490  0.044220 -0.030691  0.030573  0.055091  efectores
4     0.038589  0.020729  0.030860  0.030416  0.030462  0.027532  efectores
…        …         …         …         …         …         …         …
4995  0.033619  0.018343  0.034216  0.028119  0.029417  0.014538  efectores
4996  0.028688  0.035398  0.007782  0.020196 -0.005090  0.022792  efectores
4997  0.009945  0.022562  0.016895  0.022434  0.022949  0.025844  efectores
4998  0.028561  0.002798  0.000846  0.046707  0.023207  0.026008  efectores
4999  0.026286  0.007448  0.024344  0.038427  0.045814  0.027701  efectores

[5000 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.051468     0.008205     0.032523     0.036407     0.024658
std       0.025263     0.010433     0.019763     0.025470     0.019750
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038204     0.001776     0.020169     0.017566     0.012999
50%       0.048988     0.005307     0.029849     0.032213     0.019954
75%       0.060996     0.010855     0.041423     0.048291     0.030874
max       0.890832     0.230235     0.461115     0.345353     0.345353

                X5           X6           X7           X8           X9  …  \
```

```
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000  …
mean      0.040992     0.014559     0.035963     0.030625     0.061253  …
std       0.016193     0.012849     0.026368     0.025487     0.037546  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.031880     0.006304     0.019637     0.014367     0.035876  …
50%       0.039334     0.011826     0.029892     0.023813     0.055291  …
75%       0.047527     0.019243     0.044277     0.038923     0.078492  …
max       0.345353     0.222708     0.668124     0.445416     0.890832  …

               X31          X32          X33          X34          X35  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.019444     0.017265     0.017055     0.016015     0.016654
std       0.035383     0.039383     0.041428     0.038471     0.034296
min      -1.288468    -1.712925    -2.036664    -1.248528    -0.880373
25%       0.008011     0.007406     0.006323     0.006362     0.006247
50%       0.022953     0.021430     0.022243     0.020797     0.020635
75%       0.034516     0.031631     0.032925     0.032058     0.032152
max       0.205013     0.967655     0.192977     0.298287     0.353871

               X36          X37          X38          X39          X40
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.016739     0.018611     0.017007     0.018472     0.017167
std       0.036151     0.036098     0.037250     0.033512     0.036347
min      -1.099307    -0.986417    -0.988358    -0.718289    -1.041274
25%       0.007025     0.008252     0.005698     0.008866     0.007483
50%       0.020878     0.022183     0.021578     0.021653     0.021423
75%       0.032391     0.033591     0.032658     0.032575     0.032969
max       0.280212     0.327126     1.092382     0.262388     0.358639

[8 rows x 41 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  mass no_efectores E_coli dataset 1,
con valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.068881  0.006791  0.036866  0.041717  0.032015  0.044627  0.017463
1     0.058057  0.016328  0.036286  0.041728  0.014514  0.039914  0.018143
2     0.037989  0.003166  0.041155  0.072812  0.044320  0.063315  0.018994
3     0.036165  0.003288  0.042740  0.026302  0.016438  0.032877  0.013151
4     0.045906  0.005738  0.028691  0.051644  0.028691  0.034429  0.011476
…        …         …         …         …         …         …         …
4995  0.047880  0.000000  0.031920  0.039900  0.015960  0.047880  0.023940
4996  0.061295  0.007355  0.029422  0.036777  0.012259  0.031873  0.012259
4997  0.046192  0.000000  0.029395  0.016797  0.029395  0.062989  0.012598
```

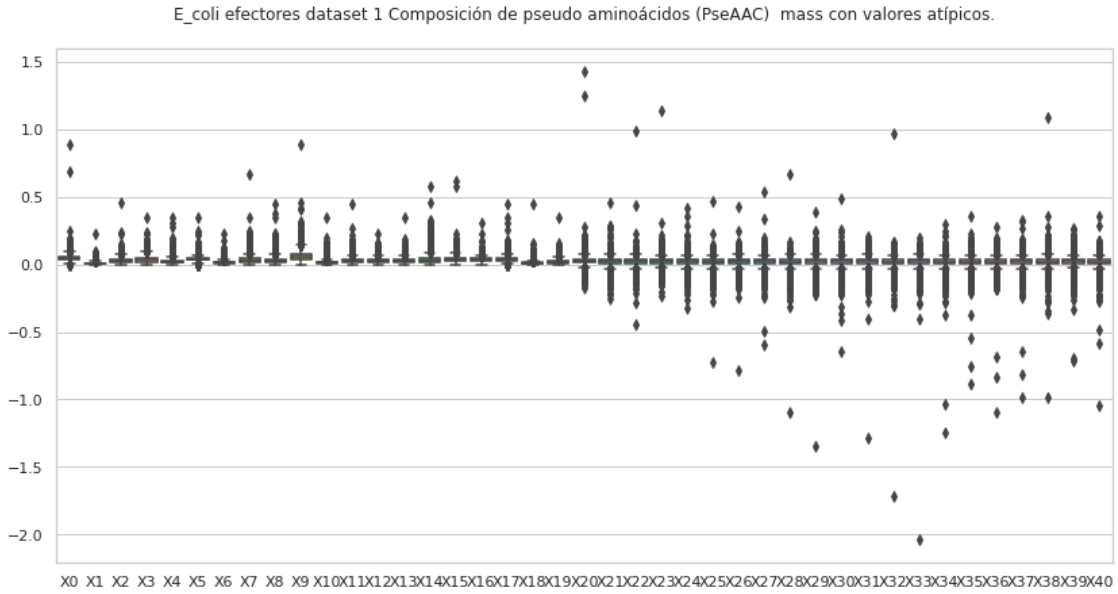
```
4998  0.050044  0.000000  0.008341  0.033363  0.033363  0.058384  0.008341
4999  0.037750  0.000000  0.006292  0.012583  0.006292  0.028313  0.003146

            X7        X8        X9  …       X32       X33       X34  \
0     0.040747  0.024254  0.065000  …  0.021796  0.003569  0.024928
1     0.030843  0.025400  0.058057  …  0.037758  0.016559  0.013925
2     0.044320  0.025326  0.075978  … -0.009578 -0.049752 -0.028086
3     0.032877  0.042740  0.062466  …  0.011103 -0.008341  0.008188
4     0.028691  0.034429  0.051644  …  0.000169  0.043277  0.073306
…          …         …       …  …         …         …
4995  0.055860  0.023940  0.055860  …  0.062893  0.016005 -0.007957
4996  0.034325  0.022066  0.031873  …  0.025929  0.007894  0.027633
4997  0.008399  0.029395  0.037794  …  0.038005 -0.020582  0.010328
4998  0.008341  0.025022  0.033363  …  0.054453  0.040053  0.019190
4999  0.037750  0.009438  0.025167  …  0.035924  0.047054  0.033626

           X35       X36       X37       X38       X39       X40           X41
0     0.002210  0.016530  0.027903  0.014497 -0.005863  0.012804  no_efectores
1     0.004807  0.025425  0.025757  0.033211  0.049007  0.014456  no_efectores
2     0.017733  0.028743  0.024973  0.039580  0.044360  0.055151  no_efectores
3     0.017738  0.003033  0.015507 -0.024430  0.043921  0.030731  no_efectores
4     0.063060  0.047587 -0.005800 -0.012489 -0.020664  0.002583  no_efectores
…          …         …         …         …         …         …           …
4995  0.008000  0.030342  0.044794 -0.008962  0.030155  0.026515  no_efectores
4996  0.021248  0.049276 -0.011532  0.023110 -0.017473  0.018668  no_efectores
4997  0.020624  0.012640  0.042657  0.032500  0.021818 -0.009689  no_efectores
4998  0.045823 -0.004871 -0.010252  0.006666  0.067750  0.044573  no_efectores
4999  0.025576  0.030986  0.017969  0.014803  0.018070  0.022063  no_efectores

[5000 rows x 42 columns]


Composición de pseudo aminoácidos (PseAAC)  mass no_efectores E_coli dataset 1,
con valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.054805     0.008639     0.034692     0.039891     0.025612
std       0.023471     0.010485     0.021778     0.028286     0.019723
min      -0.103318    -0.051659     0.000000     0.000000     0.000000
25%       0.040713     0.001738     0.019973     0.020653     0.012948
50%       0.051470     0.006108     0.031421     0.034320     0.020994
75%       0.064607     0.011719     0.044925     0.052795     0.033096
max       0.275020     0.148375     0.235778     0.331815     0.221107

                X5           X6           X7           X8           X9  …  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000  …
```

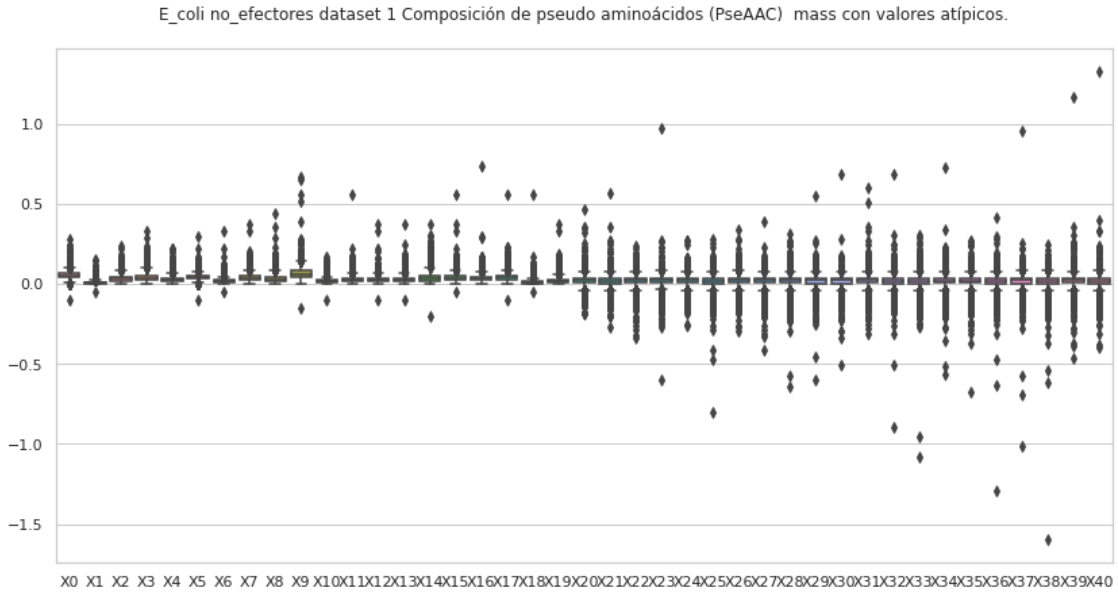
mean	0.042233	0.015894	0.038025	0.032521	0.065126	…
std	0.018150	0.015370	0.025456	0.027175	0.036417	…
min	-0.103318	-0.051659	0.000000	0.000000	-0.154977	…
25%	0.031659	0.006266	0.021357	0.014702	0.041314	…
50%	0.039620	0.012444	0.032173	0.025148	0.060283	…
75%	0.049490	0.021398	0.047586	0.042351	0.081795	…
max	0.294191	0.331815	0.368683	0.441287	0.663629	…

	X31	X32	X33	X34	X35	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.017599	0.015995	0.014935	0.016300	0.016483	
std	0.035765	0.037256	0.040292	0.037497	0.036927	
min	-0.313382	-0.895119	-1.082683	-0.565021	-0.673473	
25%	0.004942	0.003403	0.002180	0.004025	0.003862	
50%	0.021449	0.019810	0.020071	0.020208	0.020511	
75%	0.034360	0.033205	0.032701	0.033336	0.034187	
max	0.601493	0.687009	0.300154	0.722749	0.266490	

	X36	X37	X38	X39	X40
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	0.015773	0.016123	0.015041	0.016856	0.016105
std	0.040463	0.041350	0.042881	0.040296	0.042144
min	-1.289509	-1.016113	-1.596988	-0.467222	-0.394243
25%	0.002314	0.002918	0.002016	0.003925	0.002129
50%	0.019421	0.020799	0.019839	0.020878	0.020219
75%	0.032854	0.034375	0.033468	0.033609	0.033619
max	0.413963	0.955918	0.247882	1.165430	1.324941

[8 rows x 41 columns]





4.1 Composición de pseudo aminoácidos (PseAAC) mass, sin valores atípicos

[8]:
```
#mass
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "mass"
```

```
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))

    if etiq == "efectores":
        df=PseAAC_mass_efec

    if etiq == "no_efectores":
        df=PseAAC_mass_no_efec

    del df['X41']
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X41'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp))
```

```
Composición de pseudo aminoácidos (PseAAC)  mass efectores E_coli dataset 1, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.047936  0.011984  0.043941  0.029960  0.015979  0.043941  0.011984
1    0.062270  0.012924  0.019974  0.030548  0.018799  0.038772  0.014099
2    0.100480  0.000000  0.026442  0.029086  0.015865  0.023798  0.029086
4    0.046641  0.003292  0.008231  0.003292  0.013718  0.034021  0.004390
```

```
5     0.058407  0.005192  0.029852  0.027256  0.016873  0.041534  0.019469
…        …         …         …         …         …         …         …
4995  0.057432  0.003378  0.027027  0.040540  0.012387  0.028153  0.012387
4996  0.065817  0.000000  0.043034  0.045566  0.032909  0.027846  0.005063
4997  0.057601  0.000000  0.039467  0.050134  0.028800  0.036267  0.011734
4998  0.051205  0.007315  0.038038  0.020482  0.029260  0.036575  0.014630
4999  0.052694  0.007528  0.028229  0.024465  0.015055  0.039520  0.018819

            X7        X8        X9  …       X32       X33       X34  \
0     0.043941  0.027963  0.035952  …  0.028108  0.004038  0.006488
1     0.027023  0.012924  0.042297  …  0.021838  0.021875  0.014707
2     0.023798  0.029086  0.076682  …  0.012381 -0.002042  0.031318
4     0.024144  0.002744  0.068590  …  0.016617  0.036259  0.025357
5     0.024661  0.016873  0.066194  …  0.012805  0.038121  0.033590
…        …         …         …  …       …         …         …
4995  0.031531  0.019144  0.063063  …  0.044883  0.024445  0.026537
4996  0.050629  0.053160  0.035440  …  0.007278  0.018517  0.026783
4997  0.025600  0.030934  0.064001  …  0.032127  0.034771  0.022234
4998  0.014630  0.019019  0.052668  …  0.020530  0.030604  0.036961
4999  0.015055  0.018819  0.062104  …  0.027684  0.020413  0.019308

           X35       X36       X37       X38       X39       X40        X41
0     0.040213  0.027445  0.018954  0.021282  0.040919  0.049867  efectores
1     0.036204  0.024230  0.011190  0.014686  0.034933  0.025107  efectores
2    -0.019645  0.039036  0.005794 -0.016725  0.024163  0.019177  efectores
4     0.038589  0.020729  0.030860  0.030416  0.030462  0.027532  efectores
5     0.002965  0.006047  0.033578 -0.015035  0.028677  0.015961  efectores
…        …         …         …         …         …         …         …
4995  0.033619  0.018343  0.034216  0.028119  0.029417  0.014538  efectores
4996  0.028688  0.035398  0.007782  0.020196 -0.005090  0.022792  efectores
4997  0.009945  0.022562  0.016895  0.022434  0.022949  0.025844  efectores
4998  0.028561  0.002798  0.000846  0.046707  0.023207  0.026008  efectores
4999  0.026286  0.007448  0.024344  0.038427  0.045814  0.027701  efectores

[4372 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4372.000000  4372.000000  4372.000000  4372.000000  4372.000000
mean      0.049179     0.006548     0.030233     0.032857     0.021266
std       0.016520     0.006341     0.014650     0.020222     0.012914
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.038005     0.001785     0.019793     0.016576     0.012209
50%       0.048228     0.004944     0.029097     0.030197     0.018827
```

```
75%       0.059135     0.009413     0.039152     0.044316     0.028085
max       0.124394     0.036862     0.091524     0.111765     0.083234

                X5           X6           X7           X8           X9  …  \
count  4372.000000  4372.000000  4372.000000  4372.000000  4372.000000  …
mean      0.039508     0.012811     0.031478     0.026101     0.055557  …
std       0.011670     0.008991     0.017567     0.017158     0.027699  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.031891     0.006075     0.018521     0.013668     0.034065  …
50%       0.038749     0.011117     0.028302     0.021896     0.052145  …
75%       0.046059     0.017502     0.039334     0.033954     0.072431  …
max       0.087169     0.052706     0.110778     0.100907     0.172558  …

               X31          X32          X33          X34          X35  \
count  4372.000000  4372.000000  4372.000000  4372.000000  4372.000000
mean      0.022496     0.020192     0.020818     0.020100     0.019675
std       0.020727     0.018108     0.019881     0.019416     0.019205
min      -0.085086    -0.090874    -0.089251    -0.094902    -0.070947
25%       0.011575     0.010716     0.010643     0.009693     0.009232
50%       0.024209     0.022312     0.023534     0.022076     0.021615
75%       0.034516     0.031601     0.033099     0.032356     0.032377
max       0.107487     0.103458     0.140028     0.123675     0.113848

               X36          X37          X38          X39          X40
count  4372.000000  4372.000000  4372.000000  4372.000000  4372.000000
mean      0.019803     0.021045     0.019622     0.020922     0.020466
std       0.019787     0.019491     0.020434     0.019371     0.020063
min      -0.078106    -0.089210    -0.079400    -0.081227    -0.086183
25%       0.009718     0.011044     0.009489     0.011718     0.010608
50%       0.021911     0.022932     0.022904     0.022410     0.022544
75%       0.032379     0.033369     0.032647     0.032392     0.032865
max       0.107902     0.093614     0.124871     0.116365     0.115541

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

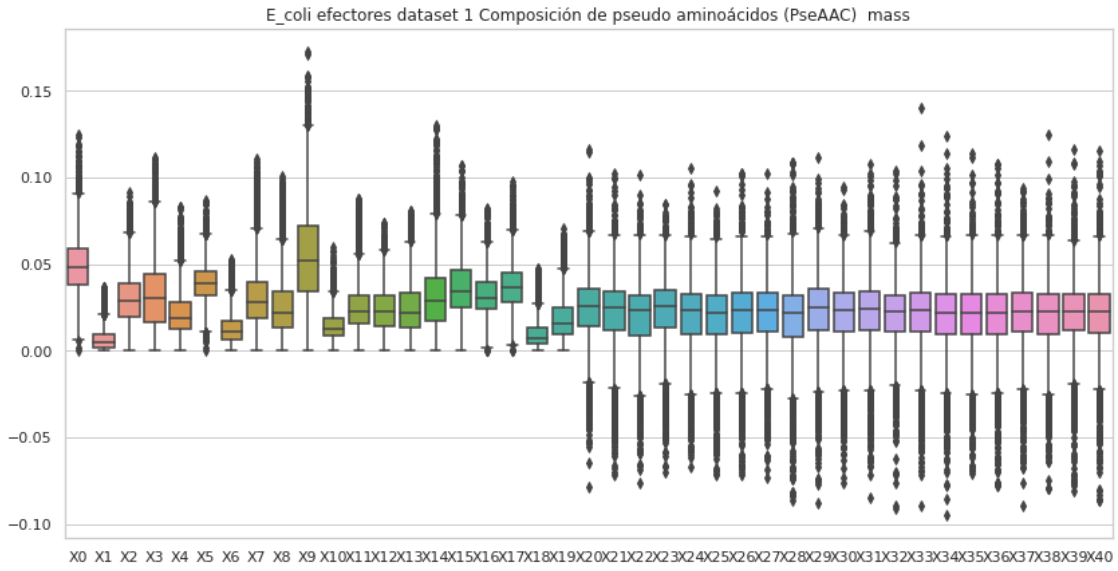
```
          X0        X1        X2        X3        X4        X5        X6  \
0   0.068881  0.006791  0.036866  0.041717  0.032015  0.044627  0.017463
1   0.058057  0.016328  0.036286  0.041728  0.014514  0.039914  0.018143
2   0.037989  0.003166  0.041155  0.072812  0.044320  0.063315  0.018994
3   0.036165  0.003288  0.042740  0.026302  0.016438  0.032877  0.013151
4   0.045906  0.005738  0.028691  0.051644  0.028691  0.034429  0.011476
…       …         …         …         …         …         …         …
```

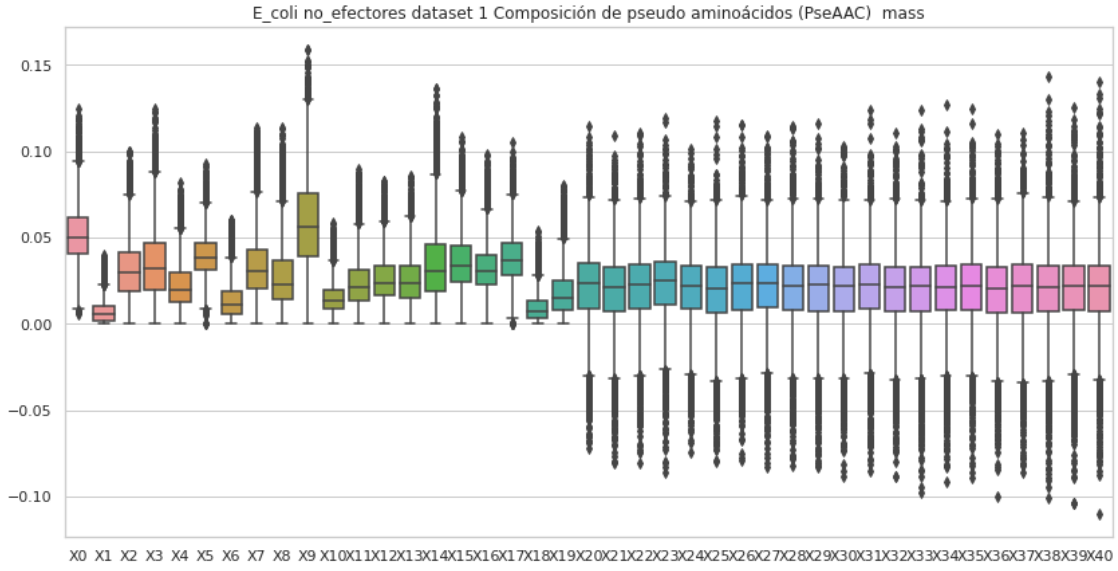
```
4995  0.047880  0.000000  0.031920  0.039900  0.015960  0.047880  0.023940
4996  0.061295  0.007355  0.029422  0.036777  0.012259  0.031873  0.012259
4997  0.046192  0.000000  0.029395  0.016797  0.029395  0.062989  0.012598
4998  0.050044  0.000000  0.008341  0.033363  0.033363  0.058384  0.008341
4999  0.037750  0.000000  0.006292  0.012583  0.006292  0.028313  0.003146

            X7        X8        X9  …       X32       X33       X34  \
0     0.040747  0.024254  0.065000  …  0.021796  0.003569  0.024928
1     0.030843  0.025400  0.058057  …  0.037758  0.016559  0.013925
2     0.044320  0.025326  0.075978  … -0.009578 -0.049752 -0.028086
3     0.032877  0.042740  0.062466  …  0.011103 -0.008341  0.008188
4     0.028691  0.034429  0.051644  …  0.000169  0.043277  0.073306
…        …         …        …  …        …         …         …
4995  0.055860  0.023940  0.055860  …  0.062893  0.016005 -0.007957
4996  0.034325  0.022066  0.031873  …  0.025929  0.007894  0.027633
4997  0.008399  0.029395  0.037794  …  0.038005 -0.020582  0.010328
4998  0.008341  0.025022  0.033363  …  0.054453  0.040053  0.019190
4999  0.037750  0.009438  0.025167  …  0.035924  0.047054  0.033626

           X35       X36       X37       X38       X39       X40           X41
0     0.002210  0.016530  0.027903  0.014497 -0.005863  0.012804  no_efectores
1     0.004807  0.025425  0.025757  0.033211  0.049007  0.014456  no_efectores
2     0.017733  0.028743  0.024973  0.039580  0.044360  0.055151  no_efectores
3     0.017738  0.003033  0.015507 -0.024430  0.043921  0.030731  no_efectores
4     0.063060  0.047587 -0.005800 -0.012489 -0.020664  0.002583  no_efectores
…        …         …         …         …         …         …            …
4995  0.008000  0.030342  0.044794 -0.008962  0.030155  0.026515  no_efectores
4996  0.021248  0.049276 -0.011532  0.023110 -0.017473  0.018668  no_efectores
4997  0.020624  0.012640  0.042657  0.032500  0.021818 -0.009689  no_efectores
4998  0.045823 -0.004871 -0.010252  0.006666  0.067750  0.044573  no_efectores
4999  0.025576  0.030986  0.017969  0.014803  0.018070  0.022063  no_efectores

[4213 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4213.000000  4213.000000  4213.000000  4213.000000  4213.000000
mean      0.051871     0.007230     0.031261     0.034887     0.022085
std       0.017269     0.006912     0.016716     0.021018     0.013684
min       0.005302     0.000000     0.000000     0.000000     0.000000
25%       0.040568     0.002133     0.019158     0.019571     0.012342
50%       0.050333     0.005738     0.029587     0.031756     0.019638
75%       0.061999     0.010246     0.041396     0.046925     0.029449
max       0.124352     0.039922     0.099978     0.124469     0.081615
```

```
                X5           X6           X7           X8           X9  …  \
count  4213.000000  4213.000000  4213.000000  4213.000000  4213.000000  …
mean      0.039917     0.013481     0.034114     0.027885     0.059027  …
std       0.012758     0.010228     0.019346     0.019642     0.026925  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.031528     0.005942     0.020668     0.013946     0.039082  …
50%       0.038642     0.011447     0.030209     0.022795     0.056268  …
75%       0.046975     0.019018     0.043008     0.036871     0.075567  …
max       0.092706     0.060395     0.113502     0.113477     0.158799  …

               X31          X32          X33          X34          X35  \
count  4213.000000  4213.000000  4213.000000  4213.000000  4213.000000
mean      0.020284     0.019278     0.018634     0.019488     0.020343
std       0.022809     0.022690     0.022975     0.022535     0.022833
min      -0.085231    -0.088846    -0.097832    -0.091836    -0.089411
25%       0.008864     0.006904     0.006824     0.007947     0.008206
50%       0.022651     0.021062     0.021674     0.021464     0.021636
75%       0.034129     0.033151     0.032769     0.033248     0.034071
max       0.123772     0.110537     0.124018     0.126645     0.124662

               X36          X37          X38          X39          X40
count  4213.000000  4213.000000  4213.000000  4213.000000  4213.000000
mean      0.019013     0.019634     0.019354     0.019725     0.019701
std       0.022956     0.023469     0.023829     0.023560     0.024003
min      -0.100498    -0.086427    -0.100876    -0.103810    -0.110259
25%       0.006116     0.006703     0.006835     0.008344     0.007231
50%       0.020675     0.022072     0.021559     0.021930     0.022170
75%       0.032677     0.034255     0.033435     0.033398     0.033689
max       0.109770     0.111023     0.143309     0.125136     0.140143

[8 rows x 41 columns]
```





5 Composición de pseudo aminoácidos (PseAAC) hidro

[9]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""
```

```
for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    #del df['X62']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  hidro efectores E_coli dataset 1,
con valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.083247  0.020812  0.076310  0.052030  0.027749  0.076310  0.020812
1     0.139307  0.028913  0.044683  0.068339  0.042055  0.086739  0.031541
2     0.127164  0.000000  0.033464  0.036811  0.020079  0.030118  0.036811
3     0.040486  0.080971  0.161943  0.121457  0.161943  0.121457  0.080971
4     0.038733  0.002734  0.006835  0.002734  0.011392  0.028252  0.003645
…          …         …         …         …         …         …         …
4995  0.098853  0.005815  0.046519  0.069778  0.021321  0.048457  0.021321
4996  0.059084  0.000000  0.038632  0.040905  0.029542  0.024997  0.004545
4997  0.088014  0.000000  0.060306  0.076605  0.044007  0.055416  0.017929
4998  0.112737  0.016105  0.083747  0.045095  0.064421  0.080526  0.032211
4999  0.096142  0.013735  0.051504  0.044637  0.027469  0.072106  0.034336

            X7        X8        X9  …       X53       X54       X55  \
0     0.076310  0.048561  0.062435  …  0.055120  0.015454 -0.006188
1     0.060454  0.028913  0.094624  …  0.028326 -0.029835 -0.028218
2     0.030118  0.036811  0.097046  … -0.042196  0.019540 -0.006719
```

```
3     0.040486  0.080971  0.283400  … -0.069996 -0.087585 -0.109054
4     0.020050  0.002278  0.056960  …  0.004353  0.016446  0.001876
…          …         …         …  …         …         …
4995  0.054272  0.032951  0.108544  …  0.034136  0.004832  0.019529
4996  0.045450  0.047722  0.031815  …  0.037563 -0.002930  0.012169
4997  0.039117  0.047267  0.097794  … -0.006245  0.005547  0.017216
4998  0.032211  0.041874  0.115958  …  0.049806 -0.060289 -0.038800
4999  0.027469  0.034336  0.113310  … -0.040850 -0.046078 -0.024391

           X56       X57       X58       X59       X60       X61         X62
0    -0.021828 -0.005979  0.033502  0.028689  0.051878  0.016908  efectores
1    -0.017720 -0.013807 -0.008766 -0.031883  0.016956 -0.001521  efectores
2     0.030002  0.025122  0.007263  0.002702 -0.021603 -0.041414  efectores
3    -0.269345 -0.114157 -0.110959 -0.155131  0.287642 -0.137646  efectores
4     0.017901  0.004884  0.030591  0.007205  0.020617  0.002836  efectores
…          …         …         …         …         …         …
4995  0.028460  0.046499  0.029774  0.005144  0.025838  0.022154  efectores
4996 -0.017497 -0.009147  0.057364  0.053892  0.006046  0.015457  efectores
4997  0.005844  0.020966 -0.010528  0.002822  0.000855 -0.009484  efectores
4998  0.003554  0.038901  0.000237 -0.025388  0.019307 -0.024867  efectores
4999  0.028413  0.016926  0.021272 -0.002848  0.004897 -0.004992  efectores

[5000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.078842     0.011670     0.048015     0.050680     0.033393
std       0.069333     0.025745     0.030385     0.060802     0.027282
min      -2.363606    -0.787869    -0.757291    -2.363606    -0.393934
25%       0.050784     0.002345     0.033558     0.031571     0.019665
50%       0.075849     0.007857     0.048511     0.048789     0.029434
75%       0.098539     0.015052     0.061826     0.065611     0.041650
max       2.620557     1.048223     0.524111     2.620557     0.544819

                X5           X6           X7           X8           X9  …  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000  …
mean      0.065849     0.021313     0.049866     0.042007     0.086775  …
std       0.047588     0.021373     0.059210     0.053433     0.087258  …
min      -0.787869    -0.151458    -1.969672    -2.363606    -2.757541  …
25%       0.039525     0.009730     0.030289     0.024428     0.054487  …
50%       0.061965     0.018684     0.044592     0.036693     0.078328  …
75%       0.083334     0.027322     0.060212     0.053927     0.105029  …
max       1.587555     0.596973     2.620557     1.572334     2.645925  …
```

```
                X52           X53           X54           X55           X56  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.006096     0.005460     0.001062     0.005772     0.003336
std       0.133289     0.073432     0.115422     0.093718     0.098428
min      -5.910111    -2.585537    -2.480351    -2.198688    -4.411328
25%      -0.013513    -0.008684    -0.017620    -0.007923    -0.015428
50%       0.007977     0.006896     0.004622     0.006237     0.007500
75%       0.027694     0.023356     0.022719     0.021226     0.026448
max       4.231394     2.611911     5.691478     4.631393     2.781439

                X57           X58           X59           X60           X61
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.005480    -0.003062     0.002465     0.001414     0.003776
std       0.087266     0.150189     0.081121     0.078025     0.071752
min      -4.228185    -8.210816    -3.939156    -1.735380    -1.562272
25%      -0.009858    -0.018449    -0.010823    -0.019525    -0.012473
50%       0.006236     0.003614     0.006330     0.004776     0.004472
75%       0.022345     0.022021     0.020884     0.023634     0.020242
max       2.952840     3.842909     1.770755     1.583027     2.253361

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores E_coli dataset 1,
con valores atípicos.
Valores del documento csv.

```
            X0        X1        X2        X3        X4        X5        X6  \
0     0.142090  0.014009  0.076048  0.086054  0.066042  0.092058  0.036023
1     0.093605  0.026326  0.058503  0.067279  0.023401  0.064353  0.029252
2     0.042972  0.003581  0.046553  0.082363  0.050134  0.071620  0.021486
3     0.065717  0.005974  0.077666  0.047794  0.029872  0.059743  0.023897
4     0.070448  0.008806  0.044030  0.079254  0.044030  0.052836  0.017612
…          …         …         …         …         …         …         …
4995  0.085060  0.000000  0.056707  0.070883  0.028353  0.085060  0.042530
4996  0.124181  0.014902  0.059607  0.074509  0.024836  0.064574  0.024836
4997  0.059909  0.000000  0.038124  0.021785  0.038124  0.081694  0.016339
4998  0.095051  0.000000  0.015842  0.063367  0.063367  0.110893  0.015842
4999  0.054942  0.000000  0.009157  0.018314  0.009157  0.041207  0.004578

            X7        X8        X9  …       X53       X54       X55  \
0     0.084053  0.050032  0.134084  …  0.011077 -0.041925 -0.032368
1     0.049728  0.040952  0.093605  …  0.004749  0.018852 -0.002525
2     0.050134  0.028648  0.085944  …  0.043586  0.006943 -0.011443
3     0.059743  0.077666  0.113512  …  0.031968 -0.039131 -0.002064
```

```
4     0.044030  0.052836  0.079254  …  0.050869  0.001328  0.100770
…        …        …        …  …        …        …        …
4995  0.099236  0.042530  0.099236  …  0.019631 -0.024360 -0.062091
4996  0.069541  0.044705  0.064574  …  0.022797  0.023980  0.040672
4997  0.010893  0.038124  0.049017  …  0.016232 -0.012615 -0.040594
4998  0.015842  0.047526  0.063367  … -0.090359  0.052033  0.163502
4999  0.054942  0.013736  0.036628  … -0.005934 -0.012813  0.016109

           X56       X57       X58       X59       X60       X61           X62
0     0.042088 -0.020983  0.018397  0.011122  0.011102  0.000935  no_efectores
1     0.001378  0.040447  0.031530  0.010392  0.040637  0.021704  no_efectores
2     0.002358 -0.040644  0.000836  0.029708  0.050715  0.037081  no_efectores
3    -0.005225  0.049235 -0.046823 -0.006001  0.012118  0.043877  no_efectores
4    -0.116895 -0.057017  0.007675 -0.029580  0.014582  0.046067  no_efectores
…        …        …        …        …        …        …           …
4995 -0.058400 -0.055478  0.018733 -0.003350  0.028190  0.136742  no_efectores
4996  0.036855  0.009762 -0.017297 -0.047150  0.026248 -0.004613  no_efectores
4997  0.044051  0.005018 -0.016430  0.036683  0.004468  0.017749  no_efectores
4998  0.024171  0.037073 -0.072360 -0.089832 -0.000067  0.056857  no_efectores
4999 -0.022300 -0.020808  0.007750 -0.012368 -0.002923  0.003399  no_efectores

[5000 rows x 63 columns]
```

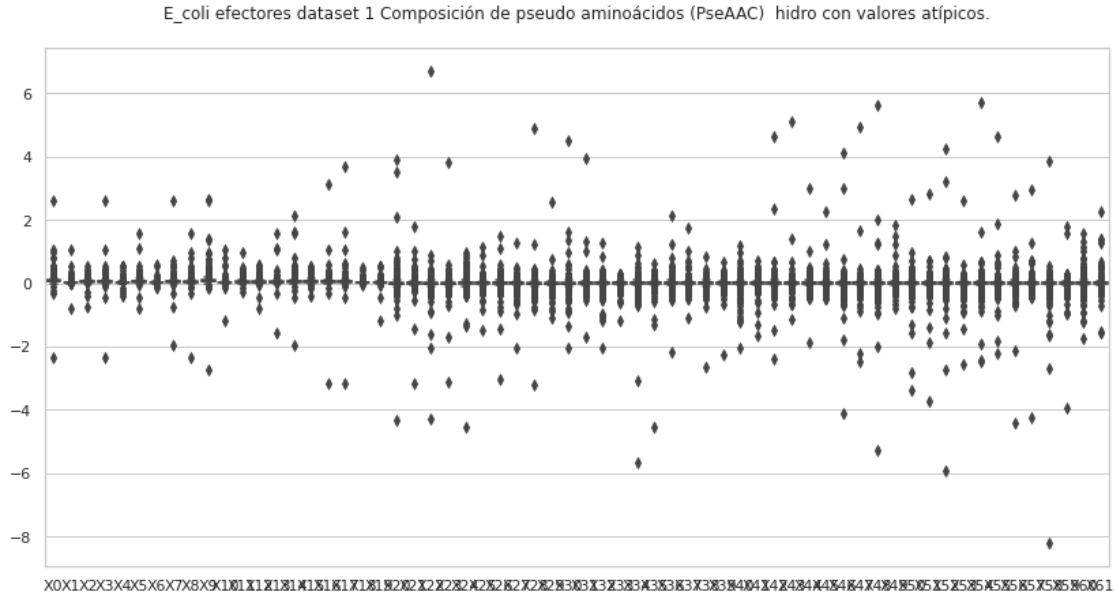
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

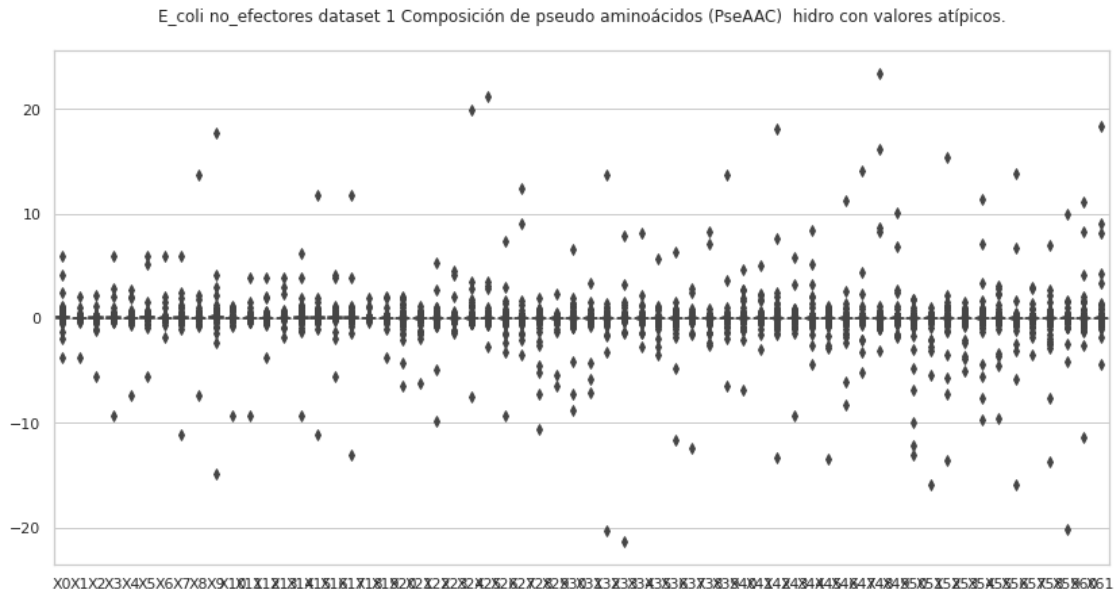
```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.082071     0.011786     0.047133     0.053355     0.033045
std       0.136187     0.065200     0.094530     0.167898     0.123168
min      -3.721382    -3.721382    -5.582073    -9.303455    -7.442764
25%       0.047654     0.002154     0.029316     0.032922     0.018413
50%       0.075811     0.008192     0.047390     0.050598     0.029072
75%       0.102341     0.016198     0.063566     0.069012     0.043129
max       5.896170     2.071803     2.139556     5.896170     2.674445

                X5           X6           X7           X8           X9  …  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000  …
mean      0.064483     0.023183     0.050589     0.044034     0.089404  …
std       0.145299     0.098902     0.190287     0.227915     0.343243  …
min      -5.582073    -1.860691   -11.164146    -7.442764   -14.885528  …
25%       0.036486     0.008270     0.029036     0.022188     0.054727  …
50%       0.057758     0.017453     0.044693     0.037321     0.080830  …
75%       0.081343     0.029632     0.064187     0.055401     0.110249  …
max       5.896170     5.896170     5.896170    13.757729    17.688509  …
```

	X52	X53	X54	X55	X56	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.000841	0.003963	0.002552	0.003347	-0.001272	
std	0.331965	0.127467	0.290705	0.190125	0.335659	
min	-13.601385	-4.998501	-9.699616	-9.595966	-15.935615	
25%	-0.018566	-0.011851	-0.019367	-0.012130	-0.022128	
50%	0.007859	0.007388	0.006930	0.006512	0.005187	
75%	0.028278	0.027417	0.028273	0.025541	0.026647	
max	15.326231	1.603853	11.422487	3.084844	13.824440	

	X57	X58	X59	X60	X61
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000
mean	0.003868	-0.002847	0.001030	0.004583	0.013454
std	0.103572	0.269431	0.331590	0.276255	0.334545
min	-3.517489	-13.687412	-20.135201	-11.345174	-4.370196
25%	-0.014732	-0.019748	-0.013325	-0.019451	-0.012076
50%	0.005716	0.006275	0.006033	0.007140	0.006439
75%	0.025486	0.027824	0.024237	0.027787	0.026053
max	3.016312	6.952892	10.011803	11.141032	18.376311

[8 rows x 62 columns]





5.1 Composición de pseudo aminoácidos (PseAAC) hidro, sin valores atípicos

[10]:
```
#hidro
transf = "Composición de pseudo aminoácidos (PseAAC) "
transf2 = "PseAAC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()


for etiq in "efectores", "no_efectores":
    titulo = (str(transf) +" "+ str(etiq) + " " + str(nombre2) + ", " +␣
 ↪str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=PseAAC_hidro_efec

    if etiq == "no_efectores":
        df=PseAAC_hidro_no_efec

    del df['X62']
```

```
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X62'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)


    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
efectores
Composición de pseudo aminoácidos (PseAAC)  efectores E_coli dataset 1, sin
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.083247  0.020812  0.076310  0.052030  0.027749  0.076310  0.020812
1     0.139307  0.028913  0.044683  0.068339  0.042055  0.086739  0.031541
2     0.127164  0.000000  0.033464  0.036811  0.020079  0.030118  0.036811
4     0.038733  0.002734  0.006835  0.002734  0.011392  0.028252  0.003645
5     0.118599  0.010542  0.060617  0.055346  0.034262  0.084337  0.039533
…          …         …         …         …         …         …         …
4995  0.098853  0.005815  0.046519  0.069778  0.021321  0.048457  0.021321
4996  0.059084  0.000000  0.038632  0.040905  0.029542  0.024997  0.004545
4997  0.088014  0.000000  0.060306  0.076605  0.044007  0.055416  0.017929
4998  0.112737  0.016105  0.083747  0.045095  0.064421  0.080526  0.032211
4999  0.096142  0.013735  0.051504  0.044637  0.027469  0.072106  0.034336

            X7        X8        X9  …       X53       X54       X55  \
0     0.076310  0.048561  0.062435  …  0.055120  0.015454 -0.006188
1     0.060454  0.028913  0.094624  …  0.028326 -0.029835 -0.028218
2     0.030118  0.036811  0.097046  … -0.042196  0.019540 -0.006719
4     0.020050  0.002278  0.056960  …  0.004353  0.016446  0.001876
5     0.050075  0.034262  0.134412  … -0.025273  0.016144 -0.006764
…          …         …         …  …         …         …         …
```

```
4995  0.054272  0.032951  0.108544  …  0.034136  0.004832  0.019529
4996  0.045450  0.047722  0.031815  …  0.037563 -0.002930  0.012169
4997  0.039117  0.047267  0.097794  … -0.006245  0.005547  0.017216
4998  0.032211  0.041874  0.115958  …  0.049806 -0.060289 -0.038800
4999  0.027469  0.034336  0.113310  … -0.040850 -0.046078 -0.024391

          X56       X57       X58       X59       X60       X61        X62
0    -0.021828 -0.005979  0.033502  0.028689  0.051878  0.016908  efectores
1    -0.017720 -0.013807 -0.008766 -0.031883  0.016956 -0.001521  efectores
2     0.030002  0.025122  0.007263  0.002702 -0.021603 -0.041414  efectores
4     0.017901  0.004884  0.030591  0.007205  0.020617  0.002836  efectores
5     0.017123 -0.024103 -0.025490 -0.027059 -0.032442 -0.025210  efectores
…        …         …         …         …         …         …         …
4995  0.028460  0.046499  0.029774  0.005144  0.025838  0.022154  efectores
4996 -0.017497 -0.009147  0.057364  0.053892  0.006046  0.015457  efectores
4997  0.005844  0.020966 -0.010528  0.002822  0.000855 -0.009484  efectores
4998  0.003554  0.038901  0.000237 -0.025388  0.019307 -0.024867  efectores
4999  0.028413  0.016926  0.021272 -0.002848  0.004897 -0.004992  efectores

[4705 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4705.000000  4705.000000  4705.000000  4705.000000  4705.000000
mean      0.075507     0.009908     0.046711     0.048267     0.030854
std       0.033824     0.010075     0.022126     0.024801     0.016068
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.050339     0.002452     0.033452     0.031025     0.019368
50%       0.074916     0.007593     0.048031     0.048047     0.028644
75%       0.096653     0.014176     0.060873     0.064260     0.039815
max       0.239950     0.083693     0.136847     0.231065     0.110504

                X5           X6           X7           X8           X9  …  \
count  4705.000000  4705.000000  4705.000000  4705.000000  4705.000000  …
mean      0.063343     0.019534     0.046241     0.038818     0.080179  …
std       0.032782     0.013406     0.022978     0.021417     0.035695  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.038842     0.009569     0.029954     0.024034     0.053515  …
50%       0.060873     0.018137     0.043576     0.035964     0.076224  …
75%       0.081770     0.026269     0.057859     0.051210     0.100289  …
max       0.194898     0.085376     0.178988     0.145689     0.260029  …

               X52          X53          X54          X55          X56  \
count  4705.000000  4705.000000  4705.000000  4705.000000  4705.000000
```

mean	0.008489	0.007654	0.003090	0.006425	0.005678
std	0.037230	0.028764	0.036470	0.028487	0.038613
min	-0.215698	-0.131821	-0.236773	-0.181901	-0.223338
25%	-0.011688	-0.007448	-0.015713	-0.006793	-0.013518
50%	0.008459	0.007045	0.004941	0.006418	0.007840
75%	0.027159	0.022678	0.022273	0.020411	0.025731
max	0.262080	0.168555	0.253163	0.198642	0.268368

	X57	X58	X59	X60	X61
count	4705.000000	4705.000000	4705.000000	4705.000000	4705.000000
mean	0.005998	0.001655	0.005086	0.002714	0.004149
std	0.030423	0.036819	0.030646	0.036634	0.030296
min	-0.160271	-0.263728	-0.230029	-0.190016	-0.148394
25%	-0.008654	-0.016602	-0.009342	-0.017243	-0.011099
50%	0.006421	0.004127	0.006541	0.005234	0.004681
75%	0.021732	0.021558	0.020415	0.023137	0.019788
max	0.217120	0.295391	0.244817	0.221736	0.174176

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.142090	0.014009	0.076048	0.086054	0.066042	0.092058	0.036023	
1	0.093605	0.026326	0.058503	0.067279	0.023401	0.064353	0.029252	
2	0.042972	0.003581	0.046553	0.082363	0.050134	0.071620	0.021486	
3	0.065717	0.005974	0.077666	0.047794	0.029872	0.059743	0.023897	
4	0.070448	0.008806	0.044030	0.079254	0.044030	0.052836	0.017612	
…	…	…	…	…	…	…	…	
4995	0.085060	0.000000	0.056707	0.070883	0.028353	0.085060	0.042530	
4996	0.124181	0.014902	0.059607	0.074509	0.024836	0.064574	0.024836	
4997	0.059909	0.000000	0.038124	0.021785	0.038124	0.081694	0.016339	
4998	0.095051	0.000000	0.015842	0.063367	0.063367	0.110893	0.015842	
4999	0.054942	0.000000	0.009157	0.018314	0.009157	0.041207	0.004578	

	X7	X8	X9	…	X53	X54	X55	\
0	0.084053	0.050032	0.134084	…	0.011077	-0.041925	-0.032368	
1	0.049728	0.040952	0.093605	…	0.004749	0.018852	-0.002525	
2	0.050134	0.028648	0.085944	…	0.043586	0.006943	-0.011443	
3	0.059743	0.077666	0.113512	…	0.031968	-0.039131	-0.002064	
4	0.044030	0.052836	0.079254	…	0.050869	0.001328	0.100770	
…	…	…	…	…	…	…		
4995	0.099236	0.042530	0.099236	…	0.019631	-0.024360	-0.062091	

```
4996  0.069541  0.044705  0.064574  …  0.022797  0.023980  0.040672
4997  0.010893  0.038124  0.049017  …  0.016232 -0.012615 -0.040594
4998  0.015842  0.047526  0.063367  … -0.090359  0.052033  0.163502
4999  0.054942  0.013736  0.036628  … -0.005934 -0.012813  0.016109

          X56       X57       X58       X59       X60       X61           X62
0     0.042088 -0.020983  0.018397  0.011122  0.011102  0.000935  no_efectores
1     0.001378  0.040447  0.031530  0.010392  0.040637  0.021704  no_efectores
2     0.002358 -0.040644  0.000836  0.029708  0.050715  0.037081  no_efectores
3    -0.005225  0.049235 -0.046823 -0.006001  0.012118  0.043877  no_efectores
4    -0.116895 -0.057017  0.007675 -0.029580  0.014582  0.046067  no_efectores
…        …         …         …         …         …         …           …
4995 -0.058400 -0.055478  0.018733 -0.003350  0.028190  0.136742  no_efectores
4996  0.036855  0.009762 -0.017297 -0.047150  0.026248 -0.004613  no_efectores
4997  0.044051  0.005018 -0.016430  0.036683  0.004468  0.017749  no_efectores
4998  0.024171  0.037073 -0.072360 -0.089832 -0.000067  0.056857  no_efectores
4999 -0.022300 -0.020808  0.007750 -0.012368 -0.002923  0.003399  no_efectores

[4953 rows x 63 columns]
```

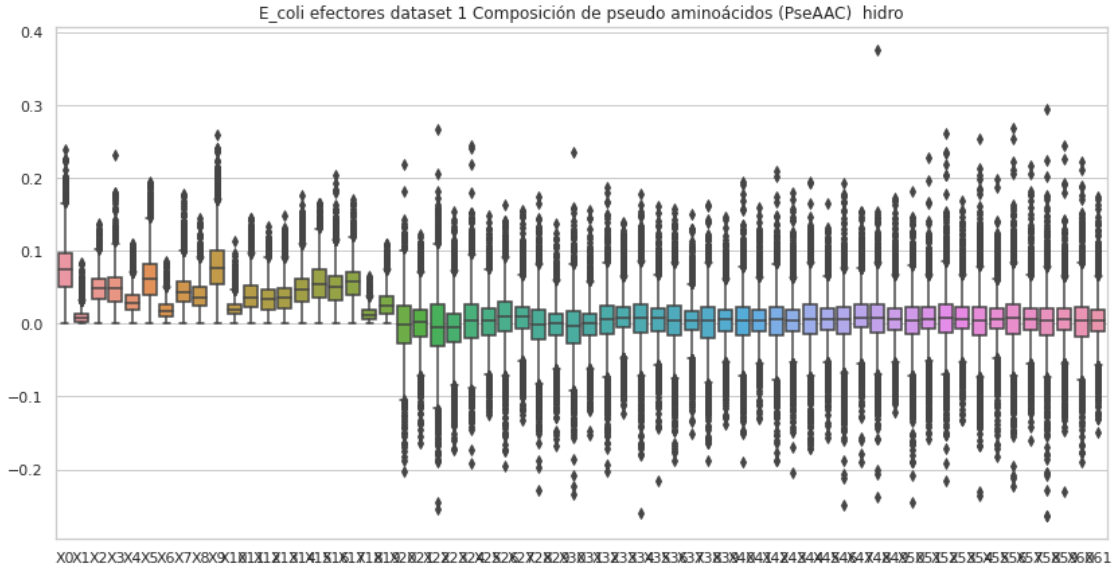
Composición de pseudo aminoácidos (PseAAC) no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

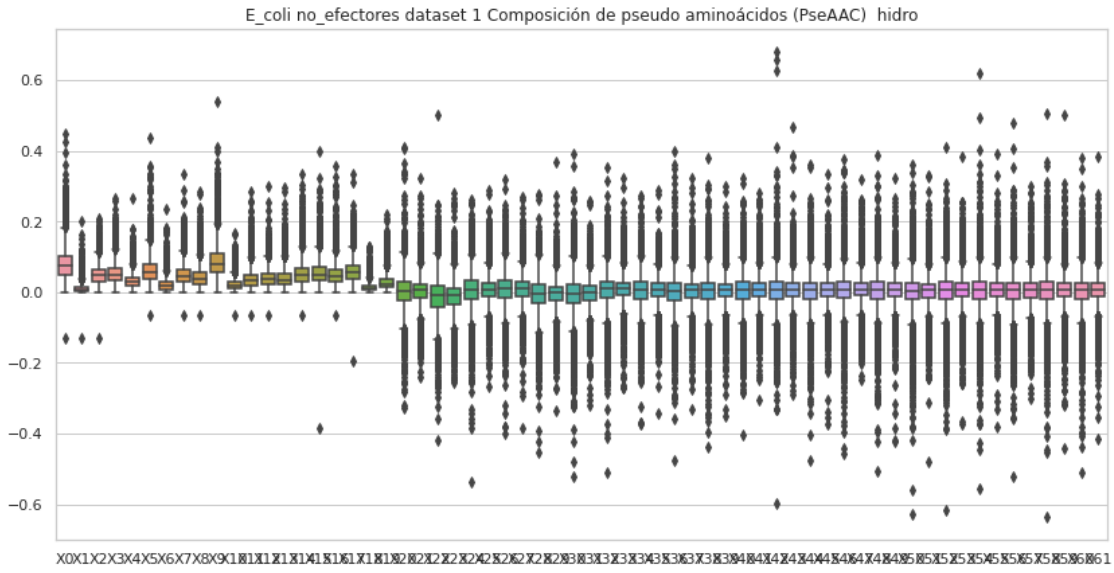
```
                X0           X1           X2           X3           X4  \
count  4953.000000  4953.000000  4953.000000  4953.000000  4953.000000
mean      0.079533     0.011613     0.047934     0.052300     0.032742
std       0.043897     0.014171     0.027619     0.030604     0.021571
min      -0.128595    -0.128595    -0.128595     0.000000     0.000000
25%       0.047639     0.002182     0.029388     0.032895     0.018453
50%       0.075655     0.008175     0.047286     0.050465     0.028979
75%       0.101942     0.016092     0.063347     0.068645     0.042767
max       0.449115     0.199631     0.209206     0.265582     0.266058

                X5           X6           X7           X8           X9  …  \
count  4953.000000  4953.000000  4953.000000  4953.000000  4953.000000  …
mean      0.062542     0.021340     0.049929     0.041606     0.086716  …
std       0.037021     0.019108     0.030266     0.028339     0.045333  …
min      -0.064297     0.000000    -0.064297    -0.064297     0.000000  …
25%       0.036522     0.008349     0.029036     0.022219     0.054583  …
50%       0.057603     0.017422     0.044538     0.037273     0.080554  …
75%       0.080893     0.029458     0.063810     0.055028     0.109715  …
max       0.436581     0.234520     0.331978     0.284506     0.540956  …

               X52          X53          X54          X55          X56  \
count  4953.000000  4953.000000  4953.000000  4953.000000  4953.000000
mean      0.002985     0.006677     0.003473     0.005807     0.001478
```

std	0.055281	0.045661	0.055573	0.044040	0.054829
min	-0.618137	-0.373969	-0.554091	-0.379347	-0.522599
25%	-0.018210	-0.011615	-0.019019	-0.011770	-0.021594
50%	0.007886	0.007423	0.006943	0.006512	0.005470
75%	0.028112	0.027238	0.028095	0.025325	0.026678
max	0.408408	0.381713	0.619191	0.389376	0.478866

	X57	X58	X59	X60	X61
count	4953.000000	4953.000000	4953.000000	4953.000000	4953.000000
mean	0.004754	0.002585	0.004832	0.002357	0.005904
std	0.045332	0.055434	0.045197	0.054042	0.044002
min	-0.303039	-0.635821	-0.443099	-0.508998	-0.415490
25%	-0.014506	-0.019339	-0.012714	-0.019121	-0.011846
50%	0.005806	0.006346	0.006152	0.007137	0.006461
75%	0.025374	0.027778	0.024216	0.027559	0.025865
max	0.300839	0.505929	0.499523	0.378580	0.381652

[8 rows x 62 columns]





6 Covarianza de auto cruzamiento (ACC) hidro_mass

[11]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "hidro_mass"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

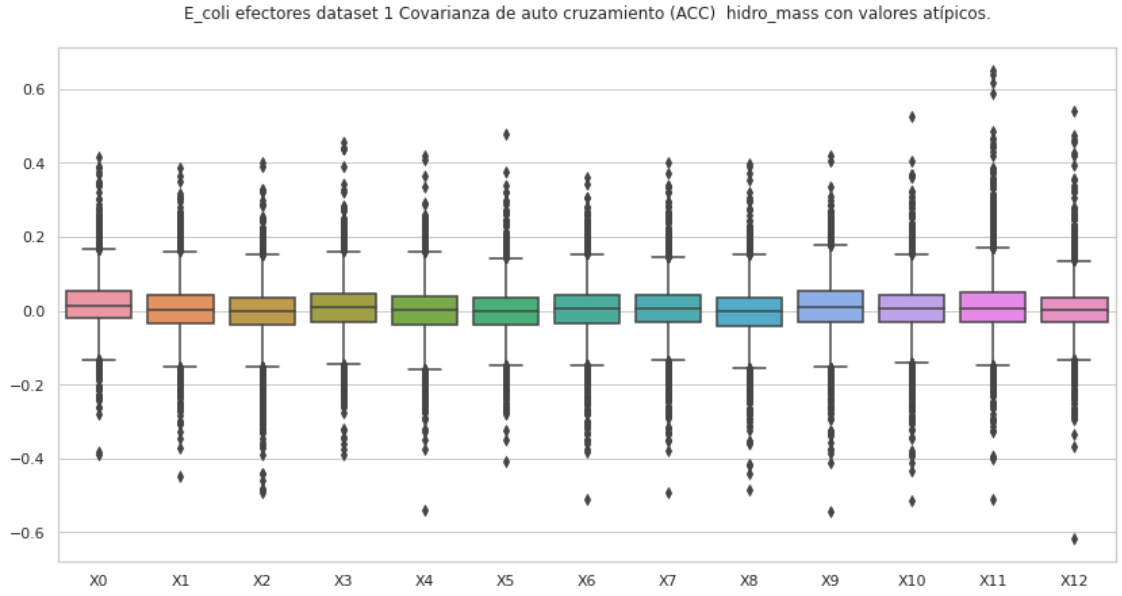
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass efectores E_coli dataset 1, con
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.077376 -0.088196  0.075833  0.091523 -0.099552 -0.098420 -0.073963
1     0.028381 -0.016699 -0.002134  0.037529 -0.036442 -0.004850  0.083475
2     0.025856  0.026529 -0.063750 -0.002298  0.019052  0.029324 -0.023916
3    -0.280045  0.164138 -0.186895 -0.042533 -0.038458 -0.014183  0.138796
4     0.038466  0.075383 -0.025156  0.100673 -0.031296  0.017963  0.015848
…          …         …         …         …         …         …         …
4995  0.002929  0.001960 -0.022317  0.011868  0.027216  0.008513 -0.009864
4996  0.016178 -0.028796 -0.054783 -0.001316  0.053138  0.067539  0.005426
4997  0.000080 -0.018600  0.010681  0.031095  0.002605 -0.048100  0.039584
4998 -0.015484 -0.052498  0.024306  0.036469  0.014764  0.094604 -0.016827
4999  0.100350 -0.006800 -0.009294 -0.025997 -0.113469  0.037093 -0.021071

            X7        X8        X9       X10       X11       X12         X13
0    -0.017186  0.016062 -0.086360 -0.023622  0.103487  0.020307  efectores
1     0.073620 -0.002637  0.075957 -0.055765 -0.045228 -0.003370  efectores
2    -0.004301  0.087321 -0.091086 -0.030830 -0.015681 -0.016724  efectores
3    -0.145100  0.202127 -0.120026  0.100044 -0.053028 -0.045605  efectores
4    -0.009901 -0.046369  0.022224 -0.053325 -0.021419 -0.077382  efectores
…          …         …         …         …         …         …          …
4995  0.002943 -0.049892 -0.058970  0.036229 -0.009857  0.093824  efectores
4996 -0.018677 -0.080764  0.008985  0.023915 -0.024210 -0.035600  efectores
4997 -0.039774 -0.010608 -0.064678  0.038887  0.027728  0.046387  efectores
4998  0.023680  0.063598 -0.067560  0.009771 -0.081678 -0.010186  efectores
4999  0.005203 -0.076145 -0.035380 -0.112944  0.099287  0.018103  efectores

[5000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores E_coli dataset 1, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.017376     0.004223    -0.004548     0.008934    -0.000542
```

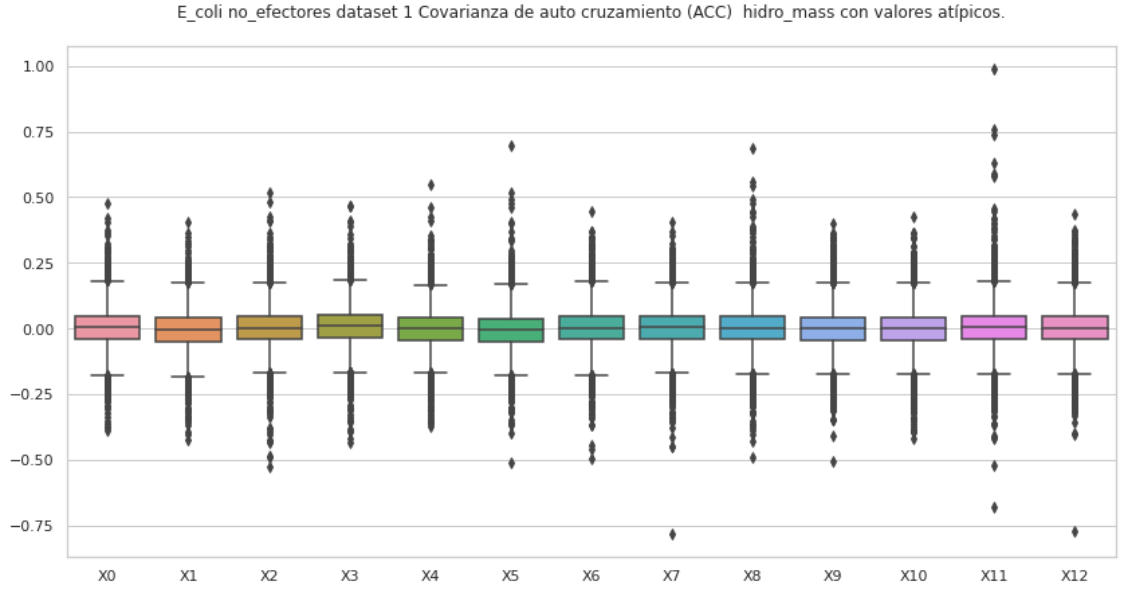
```
std       0.070101     0.070272     0.073687     0.068221     0.071053
min      -0.389204    -0.449937    -0.491468    -0.388473    -0.541536
25%      -0.021487    -0.034310    -0.039345    -0.029753    -0.039609
50%       0.012158     0.003326    -0.000209     0.008689     0.000987
75%       0.053456     0.043968     0.036748     0.046939     0.040058
max       0.417228     0.387012     0.400099     0.457085     0.418008

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.002470     0.000357     0.004138    -0.002841     0.013150
std       0.066380     0.070633     0.069762     0.071075     0.078919
min      -0.408338    -0.510631    -0.491194    -0.486654    -0.545628
25%      -0.039036    -0.034827    -0.029744    -0.040227    -0.029620
50%      -0.000069     0.004968     0.006095     0.000113     0.010301
75%       0.033983     0.040654     0.040666     0.036461     0.053445
max       0.480281     0.359964     0.401713     0.396334     0.418365

               X10          X11          X12
count  5000.000000  5000.000000  5000.000000
mean      0.005179     0.012791     0.001849
std       0.071990     0.080750     0.068541
min      -0.513401    -0.512738    -0.617534
25%      -0.031233    -0.029792    -0.032513
50%       0.006346     0.006511     0.001443
75%       0.042409     0.049666     0.034906
max       0.525613     0.650461     0.539462



no_efectores
Covarianza de auto cruzamiento (ACC)  hidro_mass no_efectores E_coli dataset 1,
con valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0    -0.030243 -0.019672 -0.045151  0.042283 -0.009105  0.006268 -0.033475
1    -0.080824 -0.040629  0.118567 -0.052110 -0.063758  0.016124  0.013795
2     0.106439  0.151949 -0.012083 -0.035980 -0.012053 -0.102279 -0.025317
3     0.087044  0.125425  0.067460  0.035318  0.016650  0.000040  0.028263
4     0.194816 -0.029300 -0.078677 -0.044346  0.061402 -0.041087  0.076805
…        …         …         …         …         …         …         …
4995  0.010320 -0.065405 -0.048549 -0.127174  0.117198  0.013278  0.050625
4996  0.043010  0.068418 -0.044684 -0.010425  0.041706  0.142078  0.064334
4997  0.011844  0.015145 -0.135385  0.050560  0.003012 -0.025702  0.104379
4998 -0.017986 -0.195728 -0.060715  0.011054  0.023115 -0.174729  0.043133
4999 -0.101247  0.050261  0.143514  0.178042 -0.205156  0.134546  0.091264

            X7        X8        X9       X10       X11       X12           X13
```

```
0     0.045322  0.081635  0.025543  0.019580 -0.039865  0.022301  no_efectores
1     0.000198 -0.020210 -0.046555  0.021890 -0.093262  0.053022  no_efectores
2     0.101325  0.016975  0.028377  0.015806 -0.013050 -0.050760  no_efectores
3     0.029972  0.070662  0.003929  0.035565 -0.012748 -0.012343  no_efectores
4     0.188631  0.064457 -0.066790 -0.070802 -0.002704 -0.035132  no_efectores
…          …         …         …         …         …         …           …
4995  0.122009  0.017848 -0.140898 -0.016816  0.171124  0.173876  no_efectores
4996  0.041866  0.213042  0.083344 -0.005997 -0.101175  0.018472  no_efectores
4997  0.017186  0.155909 -0.078553  0.082439  0.119343  0.076685  no_efectores
4998  0.061635  0.052842 -0.164454 -0.164007  0.158024  0.108366  no_efectores
4999  0.145237 -0.116825  0.168332  0.026004  0.003203  0.052573  no_efectores

[5000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean      0.003737    -0.003258     0.002292     0.009168    -0.001117
std       0.080548     0.079047     0.081600     0.079467     0.078526
min      -0.385849    -0.421979    -0.525876    -0.435517    -0.374406
25%      -0.042499    -0.048508    -0.040329    -0.034350    -0.042616
50%       0.004392    -0.002478     0.001331     0.010015     0.000995
75%       0.047902     0.041731     0.046162     0.053461     0.041490
max       0.477074     0.404719     0.518108     0.467046     0.546652

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.006316     0.001815     0.001767     0.002933     0.000536
std       0.079907     0.080122     0.080333     0.083339     0.081269
min      -0.511494    -0.496049    -0.781817    -0.492111    -0.504838
25%      -0.048213    -0.042461    -0.040539    -0.041223    -0.042767
50%      -0.005374     0.003091     0.004224     0.002081     0.000462
75%       0.038484     0.046657     0.045888     0.045479     0.043492
max       0.695117     0.445295     0.405503     0.688808     0.398259

               X10          X11          X12
count  5000.000000  5000.000000  5000.000000
mean     -0.000803     0.005494     0.000638
std       0.081009     0.087162     0.082518
min      -0.416273    -0.677384    -0.770224
25%      -0.043165    -0.039097    -0.042460
50%       0.001212     0.004188     0.000624
75%       0.043629     0.048655     0.044581
max       0.424455     0.987969     0.438151
```


E_coli efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


E_coli no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

6.1 Covarianza de auto cruzamiento (ACC) hidro_mass, sin valores atípicos

[12]:
```
#hidro_mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro_mass"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_mass_efec

    if etiq == "no_efectores":
        df=ACC_hidro_mass_no_efec

    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

efectores
Covarianza de auto cruzamiento (ACC) hidro_mass efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.077376	-0.088196	0.075833	0.091523	-0.099552	-0.098420	-0.073963	
1	0.028381	-0.016699	-0.002134	0.037529	-0.036442	-0.004850	0.083475	
2	0.025856	0.026529	-0.063750	-0.002298	0.019052	0.029324	-0.023916	
4	0.038466	0.075383	-0.025156	0.100673	-0.031296	0.017963	0.015848	
5	-0.016277	-0.000190	0.077194	0.031105	0.082260	-0.034954	-0.116922	
…	…	…	…	…	…	…	…	
4995	0.002929	0.001960	-0.022317	0.011868	0.027216	0.008513	-0.009864	
4996	0.016178	-0.028796	-0.054783	-0.001316	0.053138	0.067539	0.005426	
4997	0.000080	-0.018600	0.010681	0.031095	0.002605	-0.048100	0.039584	
4998	-0.015484	-0.052498	0.024306	0.036469	0.014764	0.094604	-0.016827	
4999	0.100350	-0.006800	-0.009294	-0.025997	-0.113469	0.037093	-0.021071	

	X7	X8	X9	X10	X11	X12	X13
0	-0.017186	0.016062	-0.086360	-0.023622	0.103487	0.020307	efectores
1	0.073620	-0.002637	0.075957	-0.055765	-0.045228	-0.003370	efectores
2	-0.004301	0.087321	-0.091086	-0.030830	-0.015681	-0.016724	efectores
4	-0.009901	-0.046369	0.022224	-0.053325	-0.021419	-0.077382	efectores
5	0.057125	-0.041936	0.119070	0.103691	0.068622	-0.022121	efectores
…	…	…	…	…	…	…	…
4995	0.002943	-0.049892	-0.058970	0.036229	-0.009857	0.093824	efectores
4996	-0.018677	-0.080764	0.008985	0.023915	-0.024210	-0.035600	efectores
4997	-0.039774	-0.010608	-0.064678	0.038887	0.027728	0.046387	efectores
4998	0.023680	0.063598	-0.067560	0.009771	-0.081678	-0.010186	efectores
4999	0.005203	-0.076145	-0.035380	-0.112944	0.099287	0.018103	efectores

[4514 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4514.000000	4514.000000	4514.000000	4514.000000	4514.000000	
mean	0.015282	0.004708	-0.001311	0.010091	0.000510	
std	0.059844	0.059888	0.059723	0.057606	0.059897	
min	-0.187214	-0.205644	-0.222450	-0.193394	-0.212271	
25%	-0.020693	-0.031918	-0.035373	-0.026852	-0.036940	
50%	0.011209	0.003575	0.000771	0.009435	0.001528	
75%	0.050029	0.042168	0.035755	0.045635	0.038537	
max	0.226703	0.214795	0.208688	0.210972	0.207420	

```
                X5           X6           X7           X8           X9  \
count  4514.000000  4514.000000  4514.000000  4514.000000  4514.000000
mean     -0.001637     0.003224     0.004765    -0.001234     0.013188
std       0.054713     0.058798     0.056193     0.058521     0.067301
min      -0.196102    -0.208205    -0.204273    -0.215975    -0.221482
25%      -0.035300    -0.030728    -0.026846    -0.036327    -0.027035
50%       0.000509     0.005983     0.006288     0.000744     0.009848
75%       0.031935     0.039835     0.037951     0.034591     0.048786
max       0.193310     0.210487     0.212980     0.198690     0.246190

               X10          X11          X12
count  4514.000000  4514.000000  4514.000000
mean      0.006404     0.008789     0.001218
std       0.057555     0.061914     0.053991
min      -0.210468    -0.225026    -0.197507
25%      -0.027976    -0.028200    -0.030481
50%       0.006766     0.005468     0.000644
75%       0.040715     0.044005     0.031912
max       0.220853     0.254103     0.199337
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

```
            X0        X1        X2        X3        X4        X5        X6  \
0    -0.030243 -0.019672 -0.045151  0.042283 -0.009105  0.006268 -0.033475
1    -0.080824 -0.040629  0.118567 -0.052110 -0.063758  0.016124  0.013795
2     0.106439  0.151949 -0.012083 -0.035980 -0.012053 -0.102279 -0.025317
3     0.087044  0.125425  0.067460  0.035318  0.016650  0.000040  0.028263
4     0.194816 -0.029300 -0.078677 -0.044346  0.061402 -0.041087  0.076805
…        …         …         …         …         …         …
4995  0.010320 -0.065405 -0.048549 -0.127174  0.117198  0.013278  0.050625
4996  0.043010  0.068418 -0.044684 -0.010425  0.041706  0.142078  0.064334
4997  0.011844  0.015145 -0.135385  0.050560  0.003012 -0.025702  0.104379
4998 -0.017986 -0.195728 -0.060715  0.011054  0.023115 -0.174729  0.043133
4999 -0.101247  0.050261  0.143514  0.178042 -0.205156  0.134546  0.091264

            X7        X8        X9       X10       X11       X12           X13
0     0.045322  0.081635  0.025543  0.019580 -0.039865  0.022301  no_efectores
1     0.000198 -0.020210 -0.046555  0.021890 -0.093262  0.053022  no_efectores
2     0.101325  0.016975  0.028377  0.015806 -0.013050 -0.050760  no_efectores
3     0.029972  0.070662  0.003929  0.035565 -0.012748 -0.012343  no_efectores
4     0.188631  0.064457 -0.066790 -0.070802 -0.002704 -0.035132  no_efectores
…        …         …         …         …         …         …
4995  0.122009  0.017848 -0.140898 -0.016816  0.171124  0.173876  no_efectores
```

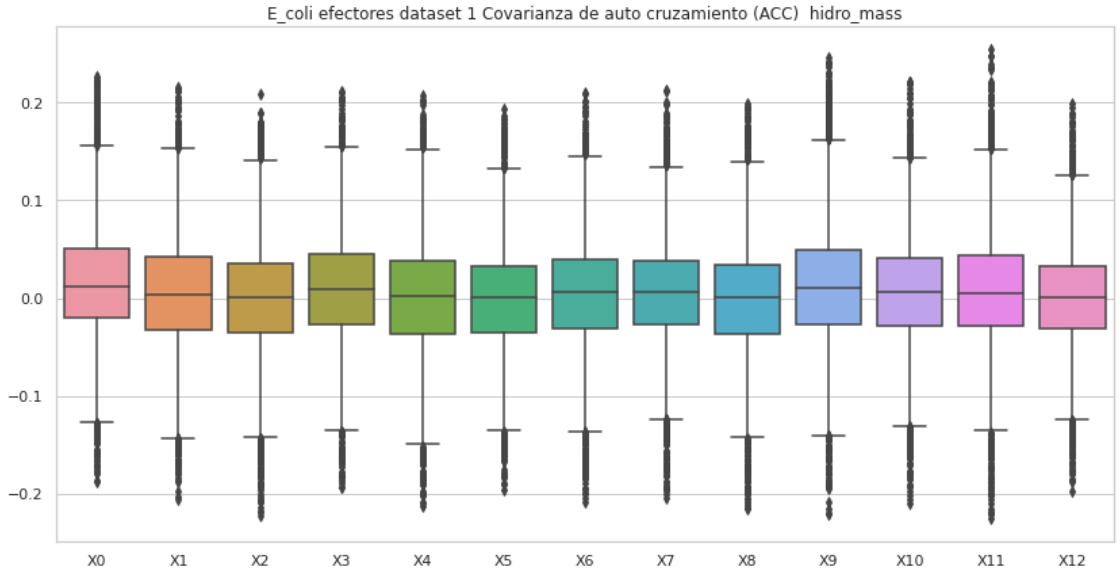
```
4996  0.041866  0.213042  0.083344 -0.005997 -0.101175  0.018472  no_efectores
4997  0.017186  0.155909 -0.078553  0.082439  0.119343  0.076685  no_efectores
4998  0.061635  0.052842 -0.164454 -0.164007  0.158024  0.108366  no_efectores
4999  0.145237 -0.116825  0.168332  0.026004  0.003203  0.052573  no_efectores

[4510 rows x 14 columns]
```

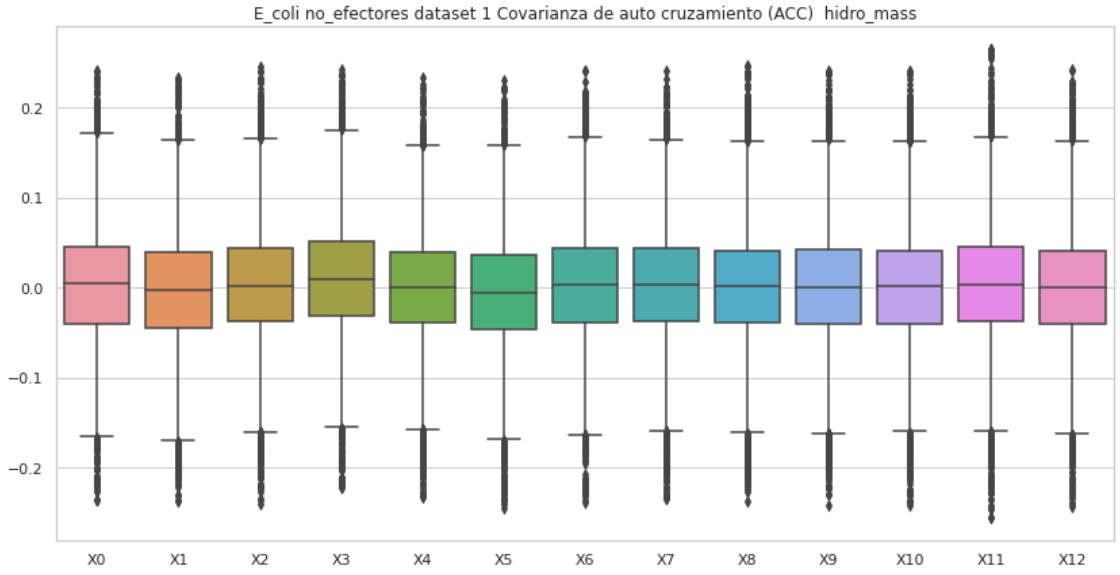
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4510.000000	4510.000000	4510.000000	4510.000000	4510.000000	
mean	0.003612	-0.002579	0.001971	0.010228	-0.001069	
std	0.070171	0.067938	0.068251	0.067955	0.066650	
min	-0.235970	-0.236584	-0.239624	-0.221514	-0.232679	
25%	-0.039429	-0.044401	-0.037586	-0.031100	-0.039067	
50%	0.004509	-0.001939	0.001439	0.010397	0.001043	
75%	0.045482	0.039274	0.043817	0.051525	0.039716	
max	0.241954	0.233257	0.245462	0.243121	0.233533	

	X5	X6	X7	X8	X9	\
count	4510.000000	4510.000000	4510.000000	4510.000000	4510.000000	
mean	-0.006208	0.002149	0.003043	0.002022	0.001104	
std	0.068111	0.068640	0.068200	0.068938	0.069791	
min	-0.243782	-0.238171	-0.233614	-0.236330	-0.241096	
25%	-0.045267	-0.038821	-0.037458	-0.038852	-0.039656	
50%	-0.004890	0.003095	0.004344	0.001622	0.000650	
75%	0.036566	0.044053	0.043532	0.041746	0.042025	
max	0.230045	0.241528	0.241550	0.247061	0.240597	

	X10	X11	X12
count	4510.000000	4510.000000	4510.000000
mean	0.000888	0.003875	0.000661
std	0.068066	0.070401	0.069715
min	-0.241115	-0.255038	-0.242382
25%	-0.039315	-0.036614	-0.040147
50%	0.001605	0.004034	-0.000089
75%	0.041503	0.045414	0.041452
max	0.240643	0.265228	0.242459





7 Covarianza de auto cruzamiento (ACC) mass

[13]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "mass"
df=""
```

```
for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
 ↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores E_coli dataset 1, con
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.077376 -0.088196  0.075833  0.091523 -0.099552 -0.098420 -0.073963
1     0.028381 -0.016699 -0.002134  0.037529 -0.036442 -0.004850  0.083475
2     0.025856  0.026529 -0.063750 -0.002298  0.019052  0.029324 -0.023916
3    -0.280045  0.164138 -0.186895 -0.042533 -0.038458 -0.014183  0.138796
4     0.038466  0.075383 -0.025156  0.100673 -0.031296  0.017963  0.015848
…        …         …         …         …         …         …         …
4995  0.002929  0.001960 -0.022317  0.011868  0.027216  0.008513 -0.009864
4996  0.016178 -0.028796 -0.054783 -0.001316  0.053138  0.067539  0.005426
4997  0.000080 -0.018600  0.010681  0.031095  0.002605 -0.048100  0.039584
4998 -0.015484 -0.052498  0.024306  0.036469  0.014764  0.094604 -0.016827
4999  0.100350 -0.006800 -0.009294 -0.025997 -0.113469  0.037093 -0.021071

            X7        X8        X9       X10       X11       X12        X13
0    -0.017186  0.016062 -0.086360 -0.023622  0.103487  0.020307  efectores
1     0.073620 -0.002637  0.075957 -0.055765 -0.045228 -0.003370  efectores
2    -0.004301  0.087321 -0.091086 -0.030830 -0.015681 -0.016724  efectores
```

```
3    -0.145100  0.202127 -0.120026  0.100044 -0.053028 -0.045605  efectores
4    -0.009901 -0.046369  0.022224 -0.053325 -0.021419 -0.077382  efectores
…         …        …        …        …        …        …         …
4995  0.002943 -0.049892 -0.058970  0.036229 -0.009857  0.093824  efectores
4996 -0.018677 -0.080764  0.008985  0.023915 -0.024210 -0.035600  efectores
4997 -0.039774 -0.010608 -0.064678  0.038887  0.027728  0.046387  efectores
4998  0.023680  0.063598 -0.067560  0.009771 -0.081678 -0.010186  efectores
4999  0.005203 -0.076145 -0.035380 -0.112944  0.099287  0.018103  efectores

[5000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.017376	0.004223	-0.004548	0.008934	-0.000542	
std	0.070101	0.070272	0.073687	0.068221	0.071053	
min	-0.389204	-0.449937	-0.491468	-0.388473	-0.541536	
25%	-0.021487	-0.034310	-0.039345	-0.029753	-0.039609	
50%	0.012158	0.003326	-0.000209	0.008689	0.000987	
75%	0.053456	0.043968	0.036748	0.046939	0.040058	
max	0.417228	0.387012	0.400099	0.457085	0.418008	

	X5	X6	X7	X8	X9	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	-0.002470	0.000357	0.004138	-0.002841	0.013150	
std	0.066380	0.070633	0.069762	0.071075	0.078919	
min	-0.408338	-0.510631	-0.491194	-0.486654	-0.545628	
25%	-0.039036	-0.034827	-0.029744	-0.040227	-0.029620	
50%	-0.000069	0.004968	0.006095	0.000113	0.010301	
75%	0.033983	0.040654	0.040666	0.036461	0.053445	
max	0.480281	0.359964	0.401713	0.396334	0.418365	

	X10	X11	X12
count	5000.000000	5000.000000	5000.000000
mean	0.005179	0.012791	0.001849
std	0.071990	0.080750	0.068541
min	-0.513401	-0.512738	-0.617534
25%	-0.031233	-0.029792	-0.032513
50%	0.006346	0.006511	0.001443
75%	0.042409	0.049666	0.034906
max	0.525613	0.650461	0.539462

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores E_coli dataset 1, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.030243	-0.019672	-0.045151	0.042283	-0.009105	0.006268	-0.033475	
1	-0.080824	-0.040629	0.118567	-0.052110	-0.063758	0.016124	0.013795	
2	0.106439	0.151949	-0.012083	-0.035980	-0.012053	-0.102279	-0.025317	
3	0.087044	0.125425	0.067460	0.035318	0.016650	0.000040	0.028263	
4	0.194816	-0.029300	-0.078677	-0.044346	0.061402	-0.041087	0.076805	
…	…	…	…	…	…	…	…	
4995	0.010320	-0.065405	-0.048549	-0.127174	0.117198	0.013278	0.050625	
4996	0.043010	0.068418	-0.044684	-0.010425	0.041706	0.142078	0.064334	
4997	0.011844	0.015145	-0.135385	0.050560	0.003012	-0.025702	0.104379	
4998	-0.017986	-0.195728	-0.060715	0.011054	0.023115	-0.174729	0.043133	
4999	-0.101247	0.050261	0.143514	0.178042	-0.205156	0.134546	0.091264	

	X7	X8	X9	X10	X11	X12	X13
0	0.045322	0.081635	0.025543	0.019580	-0.039865	0.022301	no_efectores
1	0.000198	-0.020210	-0.046555	0.021890	-0.093262	0.053022	no_efectores
2	0.101325	0.016975	0.028377	0.015806	-0.013050	-0.050760	no_efectores
3	0.029972	0.070662	0.003929	0.035565	-0.012748	-0.012343	no_efectores
4	0.188631	0.064457	-0.066790	-0.070802	-0.002704	-0.035132	no_efectores
…	…	…	…	…	…	…	…
4995	0.122009	0.017848	-0.140898	-0.016816	0.171124	0.173876	no_efectores
4996	0.041866	0.213042	0.083344	-0.005997	-0.101175	0.018472	no_efectores
4997	0.017186	0.155909	-0.078553	0.082439	0.119343	0.076685	no_efectores
4998	0.061635	0.052842	-0.164454	-0.164007	0.158024	0.108366	no_efectores
4999	0.145237	-0.116825	0.168332	0.026004	0.003203	0.052573	no_efectores

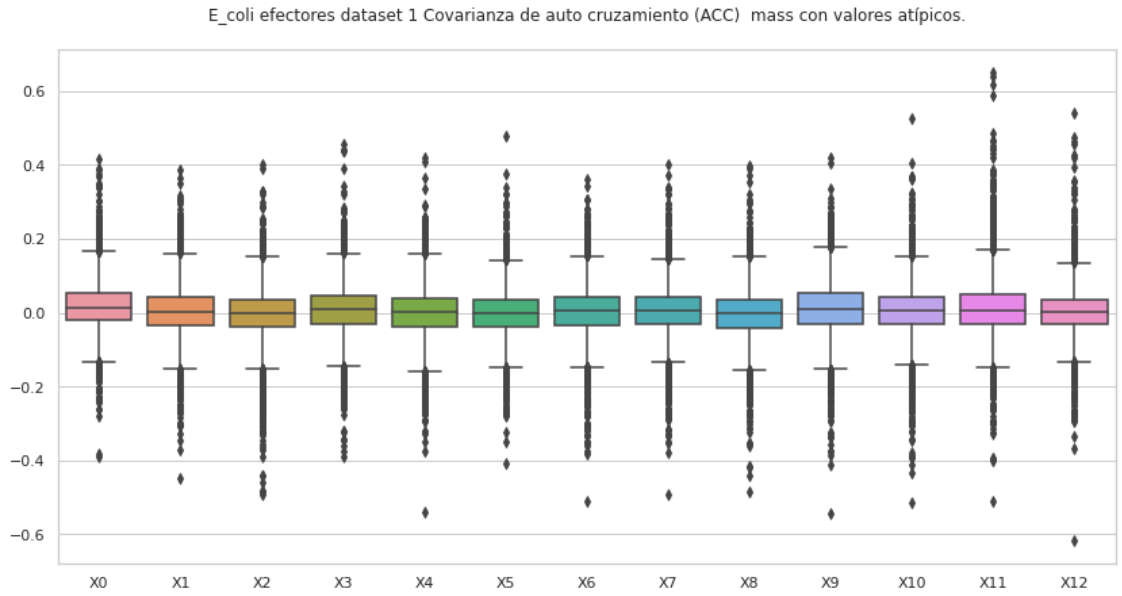
[5000 rows x 14 columns]

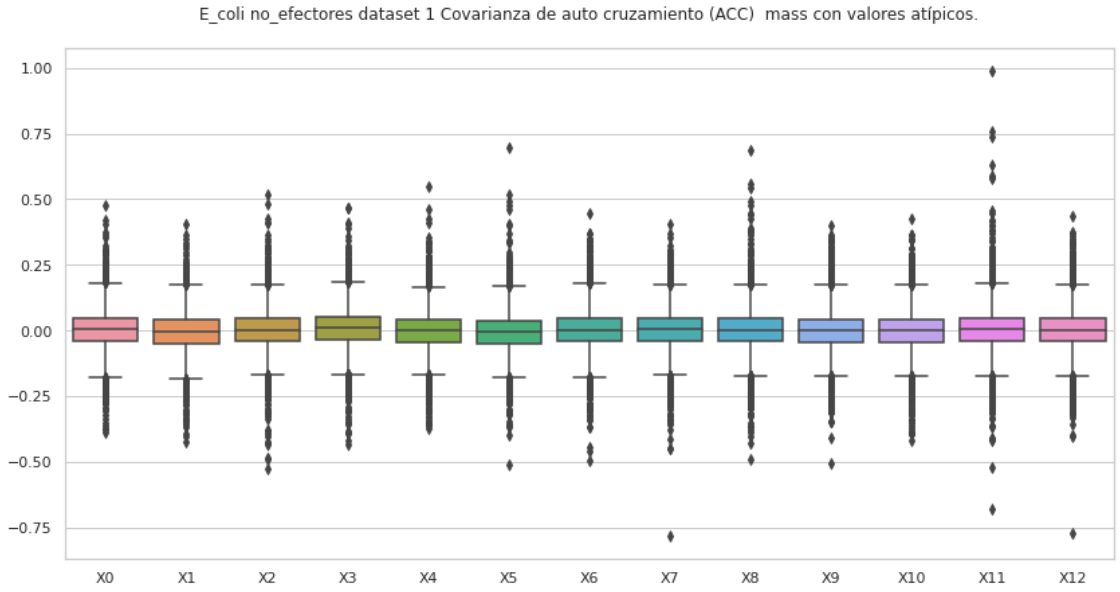
Covarianza de auto cruzamiento (ACC) mass no_efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	0.003737	-0.003258	0.002292	0.009168	-0.001117	
std	0.080548	0.079047	0.081600	0.079467	0.078526	
min	-0.385849	-0.421979	-0.525876	-0.435517	-0.374406	
25%	-0.042499	-0.048508	-0.040329	-0.034350	-0.042616	
50%	0.004392	-0.002478	0.001331	0.010015	0.000995	
75%	0.047902	0.041731	0.046162	0.053461	0.041490	
max	0.477074	0.404719	0.518108	0.467046	0.546652	

	X5	X6	X7	X8	X9	\
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	
mean	-0.006316	0.001815	0.001767	0.002933	0.000536	
std	0.079907	0.080122	0.080333	0.083339	0.081269	
min	-0.511494	-0.496049	-0.781817	-0.492111	-0.504838	
25%	-0.048213	-0.042461	-0.040539	-0.041223	-0.042767	
50%	-0.005374	0.003091	0.004224	0.002081	0.000462	
75%	0.038484	0.046657	0.045888	0.045479	0.043492	
max	0.695117	0.445295	0.405503	0.688808	0.398259	

	X10	X11	X12
count	5000.000000	5000.000000	5000.000000
mean	-0.000803	0.005494	0.000638
std	0.081009	0.087162	0.082518
min	-0.416273	-0.677384	-0.770224
25%	-0.043165	-0.039097	-0.042460
50%	0.001212	0.004188	0.000624
75%	0.043629	0.048655	0.044581
max	0.424455	0.987969	0.438151





7.1 Covarianza de auto cruzamiento (ACC) mass, sin valores atípicos

[14]:
```
#mass
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "mass"
df=""

#Se eliminan todas las filas que tengan valores atípicos en al menos una de sus␣
 ↪columnas.
out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
 ↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df=""
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))

    if etiq == "efectores":
        df=ACC_mass_efec

    if etiq == "no_efectores":
        df=ACC_mass_no_efec
```

```
    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores E_coli dataset 1, sin
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.077376 -0.088196  0.075833  0.091523 -0.099552 -0.098420 -0.073963
1     0.028381 -0.016699 -0.002134  0.037529 -0.036442 -0.004850  0.083475
2     0.025856  0.026529 -0.063750 -0.002298  0.019052  0.029324 -0.023916
4     0.038466  0.075383 -0.025156  0.100673 -0.031296  0.017963  0.015848
5    -0.016277 -0.000190  0.077194  0.031105  0.082260 -0.034954 -0.116922
…         …         …         …         …         …         …         …
4995  0.002929  0.001960 -0.022317  0.011868  0.027216  0.008513 -0.009864
4996  0.016178 -0.028796 -0.054783 -0.001316  0.053138  0.067539  0.005426
4997  0.000080 -0.018600  0.010681  0.031095  0.002605 -0.048100  0.039584
4998 -0.015484 -0.052498  0.024306  0.036469  0.014764  0.094604 -0.016827
4999  0.100350 -0.006800 -0.009294 -0.025997 -0.113469  0.037093 -0.021071

            X7        X8        X9       X10       X11       X12        X13
0    -0.017186  0.016062 -0.086360 -0.023622  0.103487  0.020307  efectores
1     0.073620 -0.002637  0.075957 -0.055765 -0.045228 -0.003370  efectores
2    -0.004301  0.087321 -0.091086 -0.030830 -0.015681 -0.016724  efectores
4    -0.009901 -0.046369  0.022224 -0.053325 -0.021419 -0.077382  efectores
5     0.057125 -0.041936  0.119070  0.103691  0.068622 -0.022121  efectores
…         …         …         …         …         …         …         …
4995  0.002943 -0.049892 -0.058970  0.036229 -0.009857  0.093824  efectores
```

```
4996 -0.018677 -0.080764  0.008985  0.023915 -0.024210 -0.035600  efectores
4997 -0.039774 -0.010608 -0.064678  0.038887  0.027728  0.046387  efectores
4998  0.023680  0.063598 -0.067560  0.009771 -0.081678 -0.010186  efectores
4999  0.005203 -0.076145 -0.035380 -0.112944  0.099287  0.018103  efectores

[4514 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4514.000000	4514.000000	4514.000000	4514.000000	4514.000000	
mean	0.015282	0.004708	-0.001311	0.010091	0.000510	
std	0.059844	0.059888	0.059723	0.057606	0.059897	
min	-0.187214	-0.205644	-0.222450	-0.193394	-0.212271	
25%	-0.020693	-0.031918	-0.035373	-0.026852	-0.036940	
50%	0.011209	0.003575	0.000771	0.009435	0.001528	
75%	0.050029	0.042168	0.035755	0.045635	0.038537	
max	0.226703	0.214795	0.208688	0.210972	0.207420	

	X5	X6	X7	X8	X9	\
count	4514.000000	4514.000000	4514.000000	4514.000000	4514.000000	
mean	-0.001637	0.003224	0.004765	-0.001234	0.013188	
std	0.054713	0.058798	0.056193	0.058521	0.067301	
min	-0.196102	-0.208205	-0.204273	-0.215975	-0.221482	
25%	-0.035300	-0.030728	-0.026846	-0.036327	-0.027035	
50%	0.000509	0.005983	0.006288	0.000744	0.009848	
75%	0.031935	0.039835	0.037951	0.034591	0.048786	
max	0.193310	0.210487	0.212980	0.198690	0.246190	

	X10	X11	X12
count	4514.000000	4514.000000	4514.000000
mean	0.006404	0.008789	0.001218
std	0.057555	0.061914	0.053991
min	-0.210468	-0.225026	-0.197507
25%	-0.027976	-0.028200	-0.030481
50%	0.006766	0.005468	0.000644
75%	0.040715	0.044005	0.031912
max	0.220853	0.254103	0.199337

Covarianza de auto cruzamiento (ACC) mass no_efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

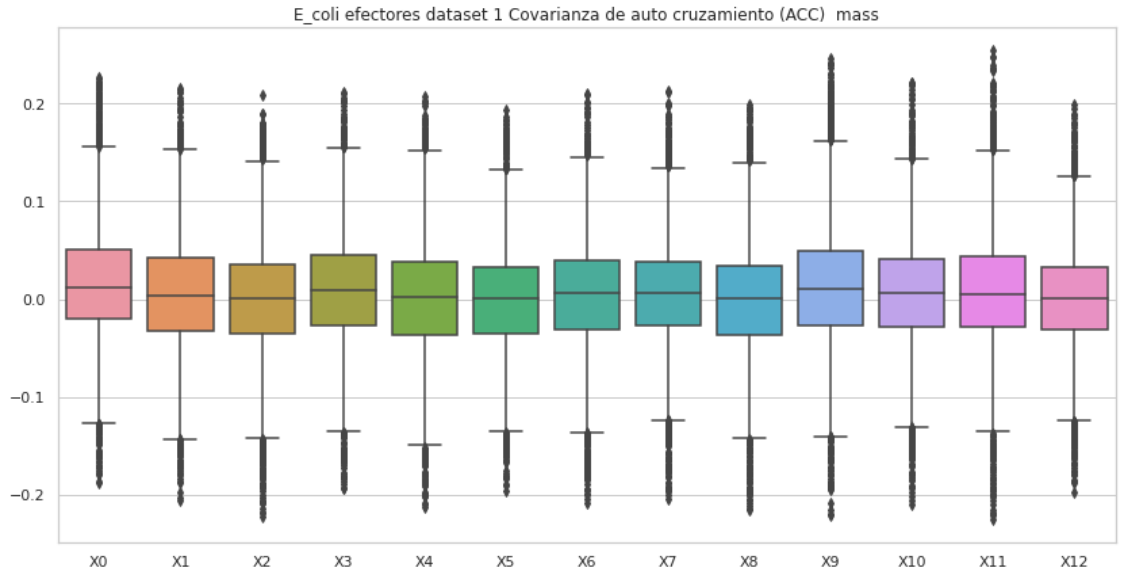
```
            X0        X1        X2        X3        X4        X5        X6  \
0    -0.030243 -0.019672 -0.045151  0.042283 -0.009105  0.006268 -0.033475
1    -0.080824 -0.040629  0.118567 -0.052110 -0.063758  0.016124  0.013795
2     0.106439  0.151949 -0.012083 -0.035980 -0.012053 -0.102279 -0.025317
3     0.087044  0.125425  0.067460  0.035318  0.016650  0.000040  0.028263
4     0.194816 -0.029300 -0.078677 -0.044346  0.061402 -0.041087  0.076805
…          …         …         …         …         …         …         …
4995  0.010320 -0.065405 -0.048549 -0.127174  0.117198  0.013278  0.050625
4996  0.043010  0.068418 -0.044684 -0.010425  0.041706  0.142078  0.064334
4997  0.011844  0.015145 -0.135385  0.050560  0.003012 -0.025702  0.104379
4998 -0.017986 -0.195728 -0.060715  0.011054  0.023115 -0.174729  0.043133
4999 -0.101247  0.050261  0.143514  0.178042 -0.205156  0.134546  0.091264

            X7        X8        X9       X10       X11       X12           X13
0     0.045322  0.081635  0.025543  0.019580 -0.039865  0.022301  no_efectores
1     0.000198 -0.020210 -0.046555  0.021890 -0.093262  0.053022  no_efectores
2     0.101325  0.016975  0.028377  0.015806 -0.013050 -0.050760  no_efectores
3     0.029972  0.070662  0.003929  0.035565 -0.012748 -0.012343  no_efectores
4     0.188631  0.064457 -0.066790 -0.070802 -0.002704 -0.035132  no_efectores
…          …         …         …         …         …         …           …
4995  0.122009  0.017848 -0.140898 -0.016816  0.171124  0.173876  no_efectores
4996  0.041866  0.213042  0.083344 -0.005997 -0.101175  0.018472  no_efectores
4997  0.017186  0.155909 -0.078553  0.082439  0.119343  0.076685  no_efectores
4998  0.061635  0.052842 -0.164454 -0.164007  0.158024  0.108366  no_efectores
4999  0.145237 -0.116825  0.168332  0.026004  0.003203  0.052573  no_efectores

[4510 rows x 14 columns]
```

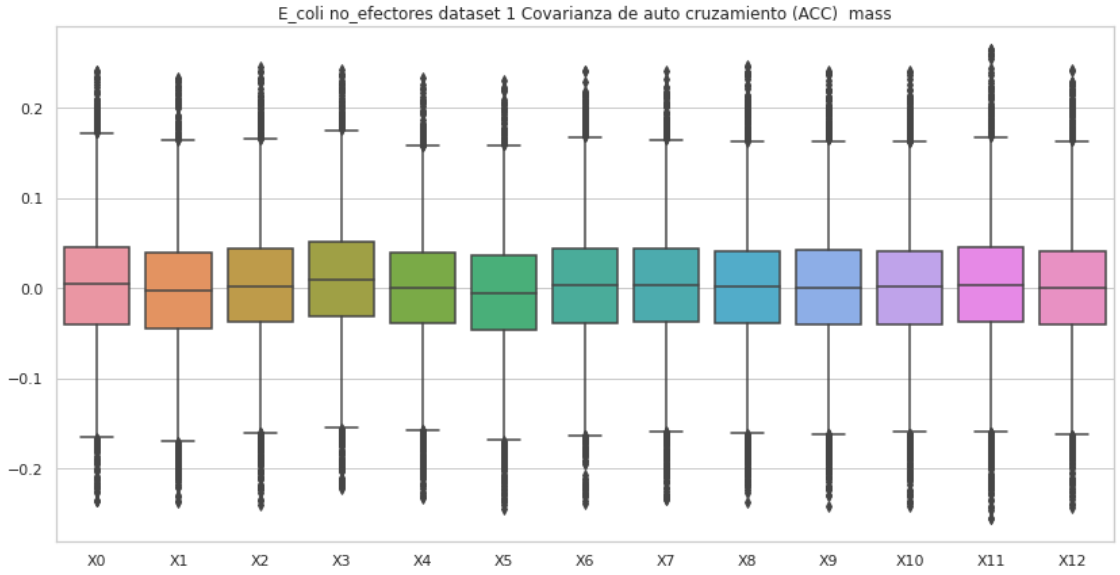
Covarianza de auto cruzamiento (ACC) mass no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  4510.000000  4510.000000  4510.000000  4510.000000  4510.000000
mean      0.003612    -0.002579     0.001971     0.010228    -0.001069
std       0.070171     0.067938     0.068251     0.067955     0.066650
min      -0.235970    -0.236584    -0.239624    -0.221514    -0.232679
25%      -0.039429    -0.044401    -0.037586    -0.031100    -0.039067
50%       0.004509    -0.001939     0.001439     0.010397     0.001043
75%       0.045482     0.039274     0.043817     0.051525     0.039716
max       0.241954     0.233257     0.245462     0.243121     0.233533

                X5           X6           X7           X8           X9  \
count  4510.000000  4510.000000  4510.000000  4510.000000  4510.000000
mean     -0.006208     0.002149     0.003043     0.002022     0.001104
std       0.068111     0.068640     0.068200     0.068938     0.069791
min      -0.243782    -0.238171    -0.233614    -0.236330    -0.241096
```

25%	-0.045267	-0.038821	-0.037458	-0.038852	-0.039656
50%	-0.004890	0.003095	0.004344	0.001622	0.000650
75%	0.036566	0.044053	0.043532	0.041746	0.042025
max	0.230045	0.241528	0.241550	0.247061	0.240597

	X10	X11	X12
count	4510.000000	4510.000000	4510.000000
mean	0.000888	0.003875	0.000661
std	0.068066	0.070401	0.069715
min	-0.241115	-0.255038	-0.242382
25%	-0.039315	-0.036614	-0.040147
50%	0.001605	0.004034	-0.000089
75%	0.041503	0.045414	0.041452
max	0.240643	0.265228	0.242459





8 Covarianza de auto cruzamiento (ACC) hidro

[15]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "con valores atípicos.\n"
comp = "hidro"
df=""

for etiq in "efectores", "no_efectores":
    titulo = (str(transf)+" "+ str(comp)+" "+ str(etiq) + " "+ str(nombre2) +",␣
 ↪" + str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    #del df['X13']
    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")
```

```
    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp)+" "+str(estado))
```

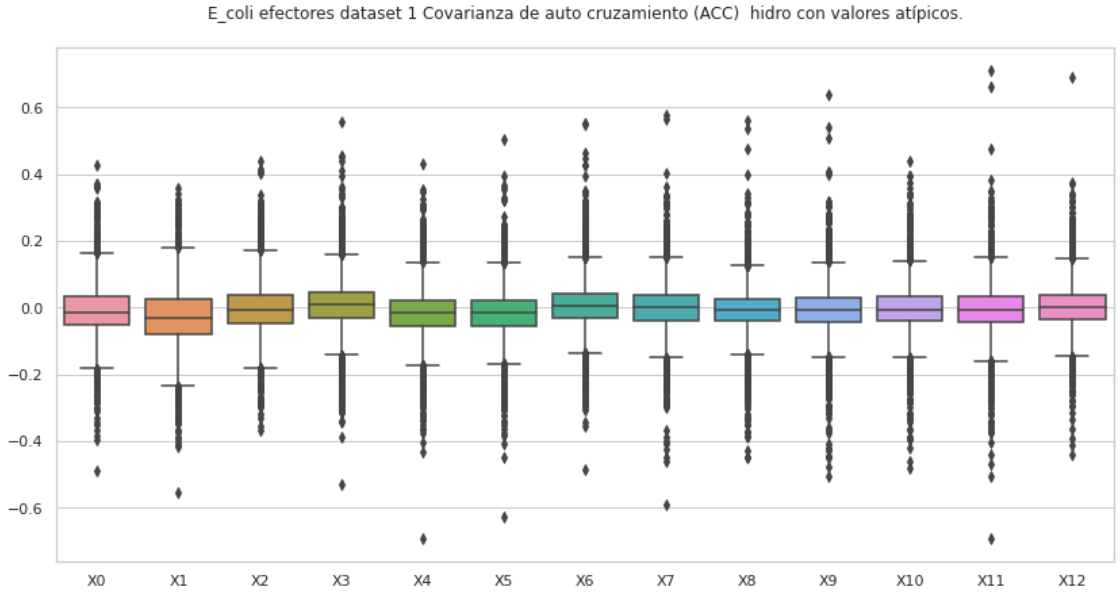
```
efectores
Covarianza de auto cruzamiento (ACC)  hidro efectores E_coli dataset 1, con
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0    -0.085097 -0.178797  0.059962  0.015801  0.004252 -0.036535 -0.132634
1     0.006439 -0.095642 -0.026894  0.033728 -0.060276  0.020702  0.015548
2     0.014060  0.017170  0.039138  0.057428 -0.066596  0.031059 -0.063172
3    -0.042148 -0.203709 -0.177273  0.218670  0.033664  0.083698 -0.023079
4     0.059792 -0.029199 -0.004267  0.002365  0.051751 -0.027156 -0.040303
…          …         …         …         …         …         …         …
4995  0.005648 -0.021203 -0.015085 -0.056617 -0.006074 -0.092652  0.011622
4996 -0.035779 -0.086199  0.009579  0.000224  0.049863  0.040137  0.007543
4997 -0.058582 -0.082639 -0.048476 -0.065264 -0.079524 -0.067068  0.020121
4998 -0.074828 -0.035255  0.017399  0.008488 -0.083313 -0.072412 -0.021019
4999  0.028067 -0.068206  0.041069 -0.066946 -0.055858 -0.005209  0.050478

            X7        X8        X9       X10       X11       X12         X13
0    -0.065495 -0.021032 -0.040802 -0.007870  0.038143 -0.016109  efectores
1     0.018032 -0.015680 -0.048546  0.057091 -0.018787 -0.037275  efectores
2     0.030360  0.082217  0.027592  0.003096  0.005174 -0.015149  efectores
3    -0.048300  0.017779  0.121604 -0.027496 -0.121096 -0.042552  efectores
4    -0.000420 -0.026349  0.013526  0.012078 -0.025232 -0.011417  efectores
…          …         …         …         …         …         …          …
4995  0.032056  0.021602 -0.044768 -0.096074  0.045958 -0.081244  efectores
4996 -0.005477 -0.065490  0.087498 -0.001420 -0.055983  0.035982  efectores
4997  0.125068 -0.016432  0.009432  0.021310 -0.026549  0.011957  efectores
4998 -0.071509 -0.030906 -0.037714 -0.008780 -0.016260 -0.001249  efectores
4999  0.121382 -0.003885  0.036279  0.040518 -0.101848 -0.019123  efectores

[5000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores E_coli dataset 1, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.010537    -0.028183    -0.003081     0.008472    -0.018236
```

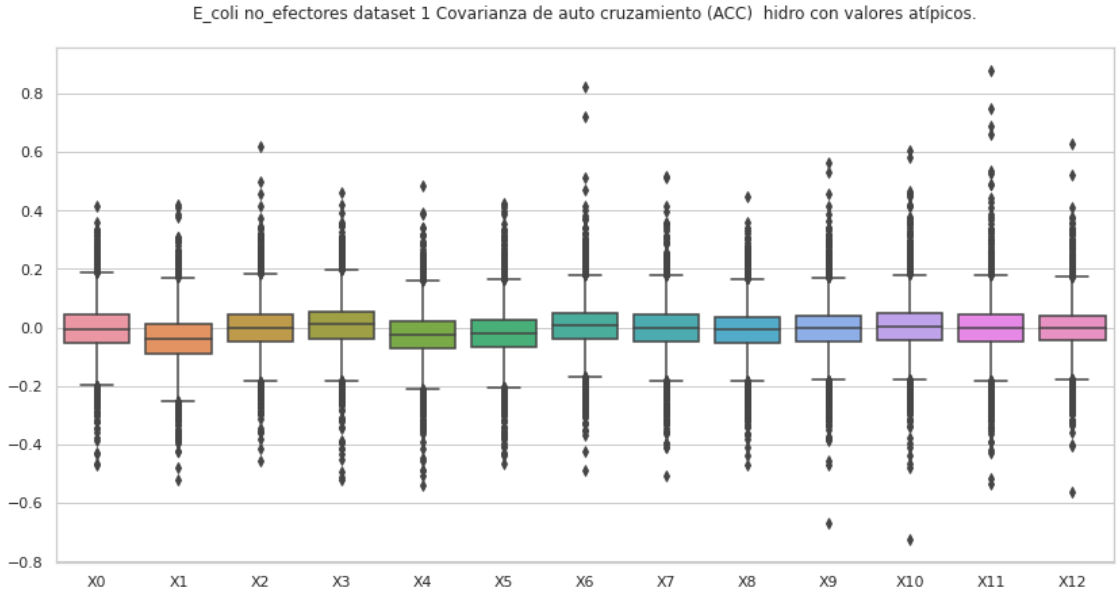
```
std       0.076850     0.086570     0.076016     0.074388     0.072470
min      -0.491061    -0.555820    -0.367204    -0.528854    -0.692255
25%      -0.053326    -0.080118    -0.049472    -0.029605    -0.057123
50%      -0.013872    -0.031064    -0.005478     0.008430    -0.015100
75%       0.033071     0.023786     0.039245     0.045872     0.020519
max       0.428395     0.355807     0.439100     0.557568     0.432137

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.018515     0.006250     0.000317    -0.007192    -0.006416
std       0.072402     0.072372     0.071832     0.070873     0.073657
min      -0.625392    -0.485139    -0.592353    -0.450514    -0.507036
25%      -0.055211    -0.029591    -0.037926    -0.039915    -0.041890
50%      -0.016265     0.003250    -0.000237    -0.006368    -0.005173
75%       0.020337     0.042119     0.037334     0.026633     0.029641
max       0.504897     0.554135     0.578443     0.559193     0.638483

               X10          X11          X12
count  5000.000000  5000.000000  5000.000000
mean     -0.003925    -0.005985    -0.000252
std       0.073226     0.074995     0.072047
min      -0.480310    -0.690370    -0.439851
25%      -0.041317    -0.044350    -0.036780
50%      -0.005959    -0.005919    -0.000698
75%       0.031608     0.033507     0.036210
max       0.437091     0.710177     0.688220



no_efectores
Covarianza de auto cruzamiento (ACC)  hidro no_efectores E_coli dataset 1, con
valores atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.001739 -0.078496  0.038463  0.032532 -0.021198 -0.043329 -0.015968
1    -0.039241 -0.056567  0.038432 -0.071892  0.015493 -0.090237  0.013220
2     0.027974 -0.111850 -0.061178 -0.101606  0.057302 -0.026872 -0.130451
3     0.009808 -0.096546 -0.143432 -0.086488  0.024776  0.024199 -0.095189
4     0.006976 -0.269785  0.092910  0.105968 -0.165162 -0.044997  0.198744
…          …         …         …         …         …         …         …
4995 -0.006795 -0.035274 -0.294082 -0.093986  0.061549  0.112034 -0.069085
4996 -0.028256 -0.151939  0.128824  0.022512 -0.135511  0.008281 -0.003447
4997  0.009588 -0.015444 -0.023656 -0.191519 -0.018206 -0.035709 -0.059783
4998 -0.146105 -0.110212  0.058503  0.103403 -0.120239 -0.083412  0.074698
4999  0.045403  0.034867  0.121220  0.194081  0.005273  0.013215  0.066636

            X7        X8        X9       X10       X11       X12           X13
```

```
0    -0.038174  0.007484 -0.008325  0.023791 -0.051827 -0.017317  no_efectores
1    -0.052817  0.031388 -0.043547 -0.030782 -0.006512 -0.059326  no_efectores
2     0.052936  0.022370 -0.057908  0.018082  0.043212 -0.057748  no_efectores
3     0.010149  0.028660 -0.086532  0.044837 -0.041295  0.067074  no_efectores
4    -0.035102 -0.101665 -0.016415  0.101661 -0.002755  0.153120  no_efectores
…         …        …        …        …        …         …            …
4995 -0.097833 -0.103088  0.021346  0.030427  0.037445  0.129196  no_efectores
4996 -0.006933 -0.100158  0.017152 -0.099123 -0.003293  0.047938  no_efectores
4997  0.165935 -0.095479 -0.015736 -0.019552  0.078014 -0.071760  no_efectores
4998 -0.002443 -0.030510 -0.071303  0.094809  0.080384 -0.008898  no_efectores
4999  0.155927 -0.078532  0.076280  0.078659  0.031242 -0.075832  no_efectores

[5000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores E_coli dataset 1, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.005893    -0.040463     0.001906     0.008657    -0.026891
std       0.083746     0.087833     0.082822     0.086976     0.084152
min      -0.469100    -0.520733    -0.454879    -0.519341    -0.539813
25%      -0.052963    -0.092092    -0.046400    -0.039202    -0.071779
50%      -0.006693    -0.041072     0.000236     0.010160    -0.025269
75%       0.044052     0.013289     0.046206     0.055813     0.021400
max       0.414064     0.421449     0.617335     0.460106     0.484819

                X5           X6           X7           X8           X9  \
count  5000.000000  5000.000000  5000.000000  5000.000000  5000.000000
mean     -0.020535     0.007026    -0.003550    -0.008689    -0.004454
std       0.084034     0.084892     0.082958     0.083058     0.084727
min      -0.466053    -0.490005    -0.507022    -0.470000    -0.670131
25%      -0.065985    -0.037514    -0.047872    -0.051847    -0.047935
50%      -0.019670     0.005174    -0.001942    -0.007002    -0.003428
75%       0.026309     0.049811     0.042621     0.035553     0.039997
max       0.424762     0.822456     0.515623     0.445890     0.562969

               X10          X11          X12
count  5000.000000  5000.000000  5000.000000
mean      0.002078    -0.001802    -0.002524
std       0.087312     0.088332     0.084168
min      -0.724813    -0.534222    -0.564545
25%      -0.043540    -0.046792    -0.045882
50%       0.002571    -0.002160    -0.002799
75%       0.046674     0.043904     0.041670
max       0.602789     0.877803     0.629447
```


E_coli efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


E_coli no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

8.1 Covarianza de auto cruzamiento (ACC) hidro, sin valores atípicos

[16]:
```
#hidro
transf = "Covarianza de auto cruzamiento (ACC) "
transf2 = "ACC"
estado = "sin valores atípicos.\n"
comp = "hidro"
df=""

out = (str(r3) + '/ds' + str(dataset) + '_' + str(transf2) + '_' + str(comp) +␣
↪'_' + str(organismo) + '.csv')
os.makedirs(str(r3), exist_ok=True)
df_out = pd.DataFrame()

for etiq in "efectores", "no_efectores":
    titulo = (str(transf) +" "+ str(etiq) + " " + str(nombre2) + ", " +␣
↪str(estado))
    print (str(etiq))

    if etiq == "efectores":
        df=ACC_hidro_efec

    if etiq == "no_efectores":
        df=ACC_hidro_no_efec

    del df['X13']
    #Se eliminan todas las filas que tengan valores atípicos en al menos una de␣
↪sus columnas.
    df = (df[(np.abs(stats.zscore(df)) < 3).all(axis=1)])
    df['X13'] = etiq
    df_out = pd.concat([df_out,df])

    #Guarda la lista csv sin valores atípicos.
    df_out.to_csv(str(out), index=False, header=False)

    print (str(titulo) + "Valores del documento csv.\n")
    print (df)
    print ("\n\n" + str(titulo) + "Estadísticas.\n")
    print(df.describe())
    print ("\n\n")

    #Gráfica de caja y bigotes
    sns.set(style="whitegrid")
    fig , ax = plt.subplots(figsize=(14,7))
    ax = sns.boxplot(data=df)
    ax.set_title(organismo +' '+str(etiq)+" dataset "+str(dataset)+"␣
↪"+str(transf)+" "+str(comp))
```

efectores
Covarianza de auto cruzamiento (ACC) efectores E_coli dataset 1, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.085097	-0.178797	0.059962	0.015801	0.004252	-0.036535	-0.132634	
1	0.006439	-0.095642	-0.026894	0.033728	-0.060276	0.020702	0.015548	
2	0.014060	0.017170	0.039138	0.057428	-0.066596	0.031059	-0.063172	
3	-0.042148	-0.203709	-0.177273	0.218670	0.033664	0.083698	-0.023079	
4	0.059792	-0.029199	-0.004267	0.002365	0.051751	-0.027156	-0.040303	
…	…	…	…	…	…	…	…	
4995	0.005648	-0.021203	-0.015085	-0.056617	-0.006074	-0.092652	0.011622	
4996	-0.035779	-0.086199	0.009579	0.000224	0.049863	0.040137	0.007543	
4997	-0.058582	-0.082639	-0.048476	-0.065264	-0.079524	-0.067068	0.020121	
4998	-0.074828	-0.035255	0.017399	0.008488	-0.083313	-0.072412	-0.021019	
4999	0.028067	-0.068206	0.041069	-0.066946	-0.055858	-0.005209	0.050478	

	X7	X8	X9	X10	X11	X12	X13
0	-0.065495	-0.021032	-0.040802	-0.007870	0.038143	-0.016109	efectores
1	0.018032	-0.015680	-0.048546	0.057091	-0.018787	-0.037275	efectores
2	0.030360	0.082217	0.027592	0.003096	0.005174	-0.015149	efectores
3	-0.048300	0.017779	0.121604	-0.027496	-0.121096	-0.042552	efectores
4	-0.000420	-0.026349	0.013526	0.012078	-0.025232	-0.011417	efectores
…	…	…	…	…	…	…	
4995	0.032056	0.021602	-0.044768	-0.096074	0.045958	-0.081244	efectores
4996	-0.005477	-0.065490	0.087498	-0.001420	-0.055983	0.035982	efectores
4997	0.125068	-0.016432	0.009432	0.021310	-0.026549	0.011957	efectores
4998	-0.071509	-0.030906	-0.037714	-0.008780	-0.016260	-0.001249	efectores
4999	0.121382	-0.003885	0.036279	0.040518	-0.101848	-0.019123	efectores

[4516 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4516.000000	4516.000000	4516.000000	4516.000000	4516.000000	
mean	-0.011391	-0.027374	-0.005024	0.007128	-0.017125	
std	0.065449	0.074413	0.064731	0.059941	0.058093	
min	-0.236679	-0.286178	-0.230575	-0.211722	-0.226830	
25%	-0.051369	-0.076376	-0.047880	-0.027566	-0.053169	
50%	-0.014240	-0.029852	-0.006152	0.007991	-0.014255	
75%	0.029245	0.021908	0.035015	0.042385	0.018069	
max	0.218055	0.220816	0.220489	0.229459	0.192167	

```
                X5           X6           X7           X8           X9  \
count  4516.000000  4516.000000  4516.000000  4516.000000  4516.000000
mean     -0.016795     0.006012    -0.000289    -0.006624    -0.006730
std       0.058124     0.058143     0.057634     0.056342     0.057443
min      -0.234713    -0.207040    -0.213914    -0.214800    -0.226049
25%      -0.051517    -0.026660    -0.035637    -0.037931    -0.039468
50%      -0.015503     0.003248    -0.000524    -0.006542    -0.005190
75%       0.019206     0.038811     0.033928     0.024081     0.026785
max       0.195060     0.215558     0.209931     0.201795     0.209574

               X10          X11          X12
count  4516.000000  4516.000000  4516.000000
mean     -0.004134    -0.005540     0.000361
std       0.057625     0.058182     0.059542
min      -0.223312    -0.229773    -0.211944
25%      -0.038254    -0.041906    -0.032816
50%      -0.005945    -0.005628    -0.000015
75%       0.028884     0.030904     0.034336
max       0.215291     0.213081     0.213499



no_efectores
Covarianza de auto cruzamiento (ACC)  no_efectores E_coli dataset 1, sin valores
atípicos.
Valores del documento csv.

            X0        X1        X2        X3        X4        X5        X6  \
0     0.001739 -0.078496  0.038463  0.032532 -0.021198 -0.043329 -0.015968
1    -0.039241 -0.056567  0.038432 -0.071892  0.015493 -0.090237  0.013220
2     0.027974 -0.111850 -0.061178 -0.101606  0.057302 -0.026872 -0.130451
3     0.009808 -0.096546 -0.143432 -0.086488  0.024776  0.024199 -0.095189
4     0.006976 -0.269785  0.092910  0.105968 -0.165162 -0.044997  0.198744
…        …        …        …        …        …        …        …
4993  0.065994 -0.110054  0.039000 -0.072459 -0.087210  0.096618  0.175336
4996 -0.028256 -0.151939  0.128824  0.022512 -0.135511  0.008281 -0.003447
4997  0.009588 -0.015444 -0.023656 -0.191519 -0.018206 -0.035709 -0.059783
4998 -0.146105 -0.110212  0.058503  0.103403 -0.120239 -0.083412  0.074698
4999  0.045403  0.034867  0.121220  0.194081  0.005273  0.013215  0.066636

            X7        X8        X9       X10       X11       X12           X13
0    -0.038174  0.007484 -0.008325  0.023791 -0.051827 -0.017317  no_efectores
1    -0.052817  0.031388 -0.043547 -0.030782 -0.006512 -0.059326  no_efectores
2     0.052936  0.022370 -0.057908  0.018082  0.043212 -0.057748  no_efectores
3     0.010149  0.028660 -0.086532  0.044837 -0.041295  0.067074  no_efectores
4    -0.035102 -0.101665 -0.016415  0.101661 -0.002755  0.153120  no_efectores
…        …        …        …        …        …        …         …
4993 -0.045333  0.127399  0.193264 -0.177705 -0.053640  0.103411  no_efectores
```

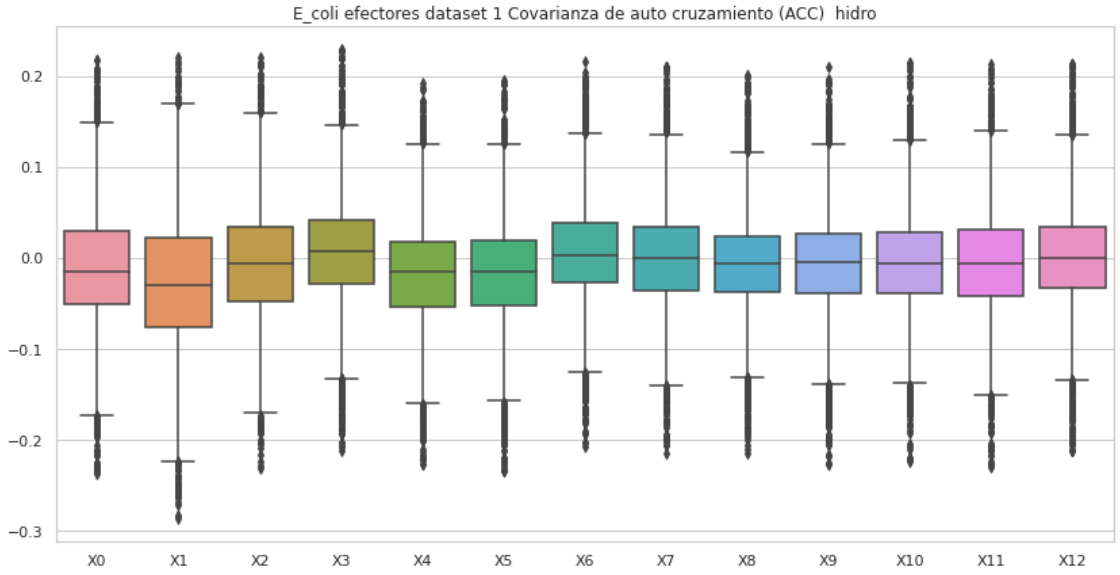
```
4996 -0.006933 -0.100158  0.017152 -0.099123 -0.003293  0.047938  no_efectores
4997  0.165935 -0.095479 -0.015736 -0.019552  0.078014 -0.071760  no_efectores
4998 -0.002443 -0.030510 -0.071303  0.094809  0.080384 -0.008898  no_efectores
4999  0.155927 -0.078532  0.076280  0.078659  0.031242 -0.075832  no_efectores

[4523 rows x 14 columns]
```

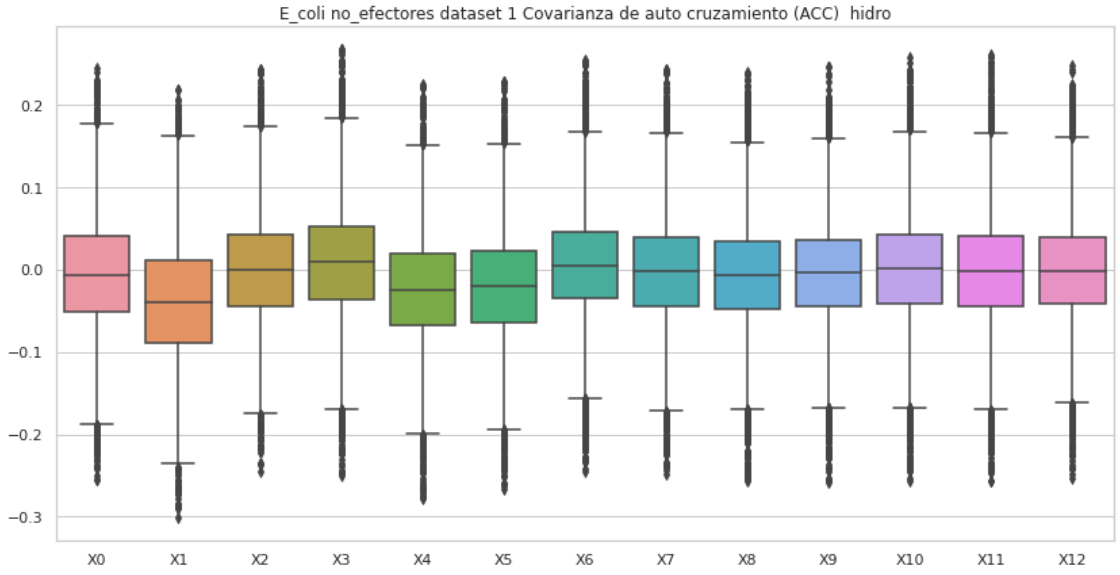
Covarianza de auto cruzamiento (ACC) no_efectores E_coli dataset 1, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	4523.000000	4523.000000	4523.000000	4523.000000	4523.000000	
mean	-0.005723	-0.038294	-0.000236	0.007861	-0.025497	
std	0.073474	0.076292	0.071457	0.075335	0.072731	
min	-0.254540	-0.301103	-0.245846	-0.250785	-0.278159	
25%	-0.050820	-0.088407	-0.044575	-0.036903	-0.068346	
50%	-0.006889	-0.039751	-0.000912	0.009664	-0.024640	
75%	0.040547	0.012027	0.042894	0.051719	0.019826	
max	0.244900	0.218009	0.243021	0.267778	0.224496	

	X5	X6	X7	X8	X9	\
count	4523.000000	4523.000000	4523.000000	4523.000000	4523.000000	
mean	-0.020800	0.005841	-0.002866	-0.007607	-0.004907	
std	0.071773	0.071902	0.071436	0.072217	0.070852	
min	-0.266258	-0.244951	-0.247685	-0.256013	-0.258240	
25%	-0.063560	-0.034921	-0.044473	-0.047780	-0.045308	
50%	-0.019731	0.004739	-0.001489	-0.006419	-0.003542	
75%	0.023468	0.046190	0.040130	0.033813	0.036704	
max	0.228015	0.254189	0.243100	0.240424	0.247061	

	X10	X11	X12
count	4523.000000	4523.000000	4523.000000
mean	0.000831	-0.001906	-0.001673
std	0.072440	0.071332	0.070416
min	-0.257311	-0.256966	-0.253958
25%	-0.040955	-0.043750	-0.041142
50%	0.002301	-0.001971	-0.002369
75%	0.043308	0.040641	0.039538
max	0.257906	0.260870	0.247754


E_coli efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


E_coli no_efectores dataset 1 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12