ds2_nematoda_limpieza_de_datos

December 14, 2020

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 ="nematoda"
dataset = 2
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 nematoda dataset 2, con valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3      X4     X5     X6      X7     X8     X9  \
0    12.500  4.167  6.944  9.722   2.778  0.000  0.000  11.111  1.389  4.167
1     5.873  5.255  4.946  5.255   3.091  5.100  2.782   4.482  2.318  5.719
2     8.870  5.739  5.217  6.087   1.043  6.957  1.913   4.870  2.609  5.043
3     7.879  4.242  4.545  2.727  11.818  5.152  7.273   7.576  0.909  1.818
4     6.707  6.098  0.000  4.878   0.610  6.707  4.878   5.488  1.829  7.927
..      …      …      …      …       …      …      …       …      …      …
495   6.786  3.929  4.286  6.071  10.714  3.214  2.857   8.214  2.143  4.643
496   1.739  4.348  8.696  3.478  12.174  4.348  6.957   8.696  4.348  6.957
497   5.121  6.469  6.469  4.043   1.617  6.739  2.695   8.895  2.426  9.434
498   7.895  4.605  3.947  9.868   3.289  3.947  2.632   9.211  2.632  5.921
```

```
499   7.071  6.313  3.535  4.545   1.768  12.374  5.556   5.051  1.515  5.808

     …    X11    X12    X13    X14    X15    X16    X17    X18     X19  \
0    …  2.778  2.778  4.167  2.778  5.556  5.556  1.389  1.389  11.111
1    …  4.946  2.782  4.637  6.646  9.583  5.100  1.236  3.864   6.491
2    …  5.565  2.435  3.826  7.478  6.435  6.261  1.217  2.609   8.174
3    …  2.424  2.424  2.424  9.091  9.091  4.545  0.909  2.121   7.273
4    …  1.829  6.098  6.707  4.878  4.268  9.146  0.610  1.220   6.098
..   …    …    …    …    …    …    …    …    …     …
495  …  7.143  1.071  2.857  7.143  8.929  6.786  1.429  2.500   7.143
496  …  9.565  1.739  3.478  5.217  6.087  2.609  0.870  5.217   2.609
497  …  8.356  1.887  3.504  3.504  7.817  2.695  0.539  1.617   6.199
498  …  3.947  2.632  2.632  5.263  7.237  5.263  0.000  3.289   5.921
499  …  6.313  1.515  2.020  5.556  8.333  5.556  0.253  2.525   6.566

           X20
0     efectores
1     efectores
2     efectores
3     efectores
4     efectores
..          …
495   efectores
496   efectores
497   efectores
498   efectores
499   efectores

[500 rows x 21 columns]


Composición de aminoácidos (AAC) efectores nematoda dataset 2, con valores
atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     6.769708    5.977452    4.374810    5.139624    2.413246    6.187620
std      2.973360    2.608195    2.263652    2.290654    2.202631    3.051827
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.082250    4.319250    2.979500    3.694750    1.097750    4.190250
50%      6.457000    5.771500    4.150500    5.207000    1.821000    5.919500
75%      8.194750    7.171750    5.456750    6.405500    3.125000    7.692000
max     37.500000   18.182000   26.000000   13.158000   15.385000   23.958000

               X6          X7          X8          X9         X10         X11  \
count  500.00000  500.000000  500.000000  500.000000  500.000000  500.00000
mean     3.64686    5.919614    2.419426    5.682096    9.141018    5.86820
```

std	2.10183	4.491590	1.503801	2.304422	3.170897	2.89848
min	0.00000	0.000000	0.000000	0.000000	0.000000	0.00000
25%	2.36575	3.879000	1.451250	4.119500	7.430500	3.97025
50%	3.43900	5.242500	2.266000	5.613000	8.906500	5.56050
75%	4.40100	7.010500	3.100500	7.012000	10.915500	7.28075
max	26.00900	62.500000	11.628000	15.596000	20.000000	23.72900

	X12	X13	X14	X15	X16	X17 \
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	2.931122	4.473272	4.981440	7.743316	5.583704	1.152546
std	1.660197	2.284016	3.397498	3.003720	2.511056	1.093798
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.973500	2.972750	3.261000	5.837000	4.092500	0.382250
50%	2.632000	4.154000	4.401000	7.539500	5.263000	0.956000
75%	3.592500	5.674500	5.964000	9.414000	6.580250	1.665500
max	23.256000	17.045000	42.708000	22.449000	25.238000	10.448000

	X18	X19
count	500.000000	500.000000
mean	3.153048	6.441922
std	1.649533	2.467984
min	0.000000	0.000000
25%	2.005250	4.848000
50%	3.067500	6.484000
75%	4.115250	7.848500
max	10.185000	26.263000

no_efectores

Composición de aminoácidos (AAC) no_efectores nematoda dataset 2, con valores atípicos.

Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9 \
0	6.250	3.125	3.571	3.571	0.893	4.018	1.339	10.268	1.339	8.482
1	5.882	11.765	3.361	4.202	0.000	7.563	1.681	5.882	6.723	4.202
2	5.556	4.040	3.535	3.030	1.515	4.545	3.030	6.566	3.030	7.071
3	2.414	4.655	7.241	4.655	2.241	4.655	3.276	4.138	3.793	7.759
4	5.327	6.780	2.663	2.663	1.695	3.632	2.906	1.453	3.390	9.443
..	…	…	…	…	…	…	…	…	…	…
495	4.503	5.253	5.253	7.129	3.377	9.006	3.940	4.690	2.251	3.752
496	6.866	3.284	4.179	6.866	0.896	7.164	12.537	5.970	1.493	3.881
497	1.852	6.173	4.938	3.086	9.259	4.321	5.556	9.877	0.617	6.790
498	9.774	3.418	4.716	5.947	5.742	8.954	3.828	3.554	1.777	4.443
499	2.997	0.817	4.360	1.362	0.817	2.180	1.362	4.360	1.635	7.629

…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \

```
0    …  4.464  4.464   8.929  4.464  11.161  4.018  1.786  2.679   4.911
1    …  4.202  4.202   2.521  5.042  10.924  4.202  0.000  3.361   5.882
2    …  6.566  3.030   6.566  0.505   3.535  5.051  2.525  2.020  11.111
3    …  5.517  1.724   5.172  6.552   9.138  4.655  0.862  3.276   4.138
4    …  4.116  2.179   7.022  5.327   9.443  7.990  1.453  4.358   6.295
..   …    …    …     …    …     …     …    …     …      …
495  …  4.690  2.064   5.816  3.940   9.944  4.128  1.689  3.752   6.567
496  …  4.478  1.493   2.687  8.955   8.358  5.672  0.896  0.896   5.672
497  …  9.877  2.469   4.938  4.321   7.407  4.321  0.617  3.086   3.086
498  …  5.400  2.324   3.349  6.357   8.954  4.511  0.547  1.162   6.835
499  …  3.542  2.997  16.349  2.997   9.264  2.180  3.815  4.360   9.264

               X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
495   no_efectores
496   no_efectores
497   no_efectores
498   no_efectores
499   no_efectores

[500 rows x 21 columns]
```

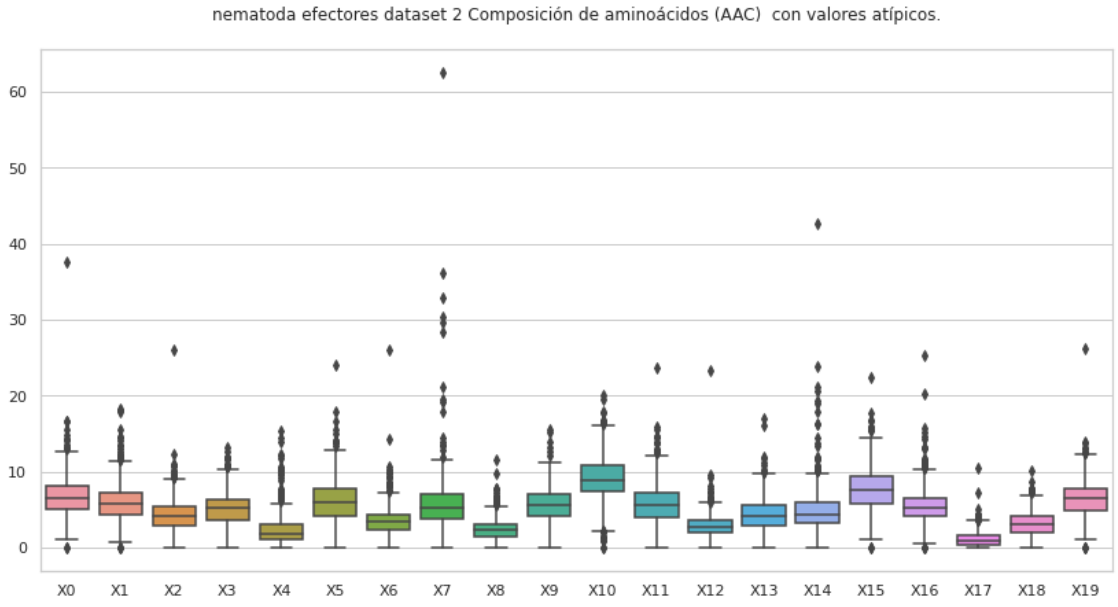
Composición de aminoácidos (AAC) no_efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

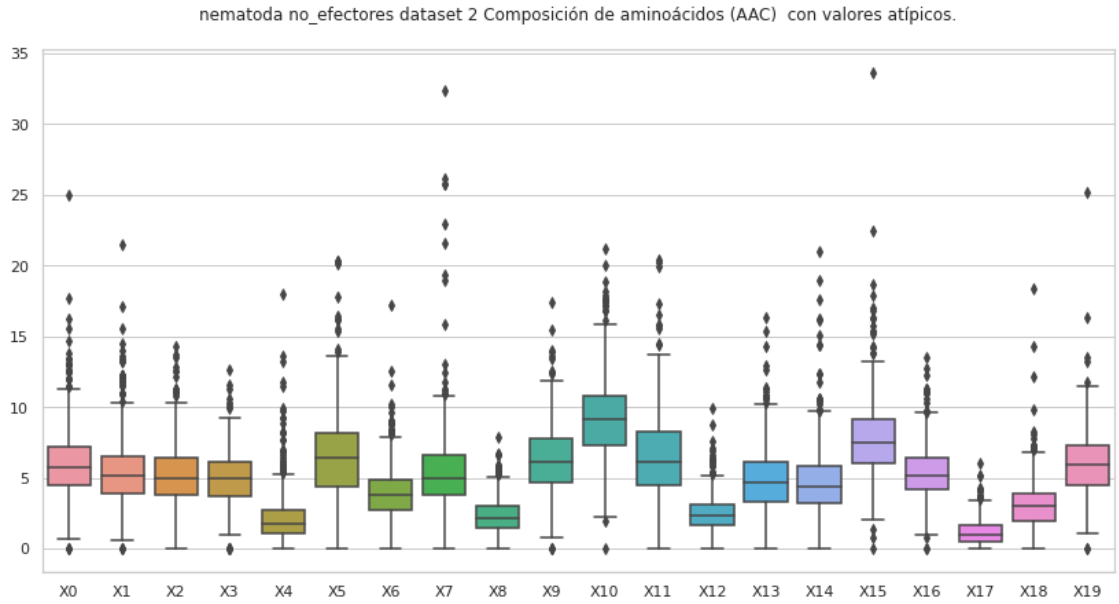
```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      6.078042    5.463254    5.252742    4.993012    2.232816    6.575662
std       2.693339    2.592927    2.227139    1.913910    2.048331    3.114682
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       4.478000    3.933250    3.794000    3.731250    1.064000    4.417750
50%       5.769000    5.197500    4.971000    4.991500    1.762000    6.419500
75%       7.240000    6.491000    6.440000    6.117750    2.740750    8.154000
max      25.000000   21.463000   14.286000   12.658000   17.937000   20.312000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      4.031770    5.562398    2.280386    6.223200    9.228780    6.463334
std       1.976829    3.280031    1.289290    2.403263    3.021454    3.085202
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       2.766500    3.780000    1.490000    4.701500    7.311500    4.444250
```

50%	3.792500	4.979000	2.129500	6.154000	9.091000	6.090500
75%	4.872000	6.573500	2.992500	7.747500	10.759500	8.217000
max	17.188000	32.328000	7.874000	17.442000	21.212000	20.438000

	X12	X13	X14	X15	X16	X17	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.502056	4.927292	4.752082	7.769228	5.310374	1.174472	
std	1.238051	2.423670	2.571427	2.963333	1.915720	0.972921	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	1.635000	3.305500	3.197750	6.067750	4.191750	0.484000	
50%	2.332500	4.678500	4.404000	7.493500	5.151000	0.990000	
75%	3.066250	6.085750	5.812500	9.091000	6.370000	1.690500	
max	9.910000	16.349000	21.002000	33.621000	13.504000	6.061000	

	X18	X19
count	500.000000	500.000000
mean	3.113798	6.065342
std	1.812588	2.383545
min	0.000000	0.000000
25%	1.971500	4.481250
50%	2.976500	5.882000
75%	3.923000	7.303000
max	18.331000	25.191000





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 nematoda dataset 2, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2     X3     X4      X5     X6      X7     X8     X9  \
0    12.500  4.167  6.944  9.722  2.778   0.000  0.000  11.111  1.389  4.167
1     5.873  5.255  4.946  5.255  3.091   5.100  2.782   4.482  2.318  5.719
2     8.870  5.739  5.217  6.087  1.043   6.957  1.913   4.870  2.609  5.043
4     6.707  6.098  0.000  4.878  0.610   6.707  4.878   5.488  1.829  7.927
5     6.757  8.108  4.054  8.108  0.676   8.784  2.703   3.378  6.081  4.054
..      …      …      …      …      …       …      …       …      …      …
492   7.864  9.189  2.235  5.215  1.407  11.093  4.719   3.146  2.401  4.305
493   9.890  7.692  4.396  2.198  1.099   7.692  2.198   2.198  4.396  1.099
497   5.121  6.469  6.469  4.043  1.617   6.739  2.695   8.895  2.426  9.434
498   7.895  4.605  3.947  9.868  3.289   3.947  2.632   9.211  2.632  5.921
499   7.071  6.313  3.535  4.545  1.768  12.374  5.556   5.051  1.515  5.808

     …     X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …   2.778  2.778  4.167  2.778   5.556  5.556  1.389  1.389  11.111
1    …   4.946  2.782  4.637  6.646   9.583  5.100  1.236  3.864   6.491
2    …   5.565  2.435  3.826  7.478   6.435  6.261  1.217  2.609   8.174
4    …   1.829  6.098  6.707  4.878   4.268  9.146  0.610  1.220   6.098
5    …   6.757  4.730  4.054  4.730   5.405  3.378  1.351  1.351   8.108
..   …     …      …      …      …       …      …      …      …       …
```

```
492  …   5.215  2.649  2.897  5.215  10.348  4.884  1.573  1.821   5.960
493  …  10.989  2.198  3.297  6.593   7.692  6.593  0.000  4.396   6.593
497  …   8.356  1.887  3.504  3.504   7.817  2.695  0.539  1.617   6.199
498  …   3.947  2.632  2.632  5.263   7.237  5.263  0.000  3.289   5.921
499  …   6.313  1.515  2.020  5.556   8.333  5.556  0.253  2.525   6.566

           X20
0    efectores
1    efectores
2    efectores
4    efectores
5    efectores
..         …
492  efectores
493  efectores
497  efectores
498  efectores
499  efectores

[422 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  422.000000  422.000000  422.000000  422.000000  422.000000  422.000000
mean     6.781140    5.974346    4.369540    5.331562    2.169744    6.376846
std      2.473537    2.175983    1.855312    2.084944    1.519152    2.657612
min      1.136000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.138500    4.505500    3.125000    3.950250    1.124000    4.738000
50%      6.485000    5.879500    4.202000    5.304000    1.790000    6.231000
75%      8.196250    7.186000    5.347000    6.439750    3.016250    7.764000
max     15.541000   13.462000   11.111000   11.940000    8.889000   14.980000

                X6          X7          X8          X9         X10         X11  \
count  422.000000  422.000000  422.000000  422.000000  422.000000  422.000000
mean     3.570611    5.416559    2.375002    5.823173    9.505178    5.841616
std      1.628760    2.203374    1.293722    1.953074    2.792333    2.480889
min      0.000000    0.000000    0.000000    0.787000    0.000000    0.000000
25%      2.446750    3.933750    1.507500    4.456000    7.858750    4.067250
50%      3.492500    5.211500    2.266000    5.807000    9.289500    5.614500
75%      4.384750    6.780000    3.064750    7.030000   11.111000    7.143000
max      9.890000   17.829000    6.818000   11.250000   17.822000   14.474000

               X12         X13         X14         X15         X16         X17  \
count  422.000000  422.000000  422.000000  422.000000  422.000000  422.000000
```

mean	2.868047	4.570737	4.642758	7.819412	5.491427	1.170682
std	1.221126	2.033308	2.116637	2.754518	1.903219	0.929110
min	0.495000	0.000000	0.000000	0.000000	1.075000	0.000000
25%	2.055250	3.131250	3.300750	5.959000	4.173250	0.509250
50%	2.663000	4.332500	4.384000	7.657000	5.263000	1.029500
75%	3.460750	5.715750	5.639250	9.455750	6.515750	1.686500
max	7.692000	11.236000	14.474000	16.667000	12.987000	4.369000

	X18	X19
count	422.000000	422.000000
mean	3.243690	6.657998
std	1.519236	2.136467
min	0.000000	1.274000
25%	2.151000	5.214000
50%	3.142000	6.663000
75%	4.181250	7.952000
max	7.742000	13.675000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	6.250	3.125	3.571	3.571	0.893	4.018	1.339	10.268	1.339	8.482	
2	5.556	4.040	3.535	3.030	1.515	4.545	3.030	6.566	3.030	7.071	
3	2.414	4.655	7.241	4.655	2.241	4.655	3.276	4.138	3.793	7.759	
4	5.327	6.780	2.663	2.663	1.695	3.632	2.906	1.453	3.390	9.443	
5	5.240	5.539	2.994	6.737	2.096	6.437	3.443	7.934	2.695	3.743	
..	…	…	…	…	…	…	…	…	…	…	
492	6.193	4.230	3.021	4.230	1.813	4.079	2.266	5.589	0.906	9.215	
493	7.080	4.425	4.646	4.646	2.655	8.186	3.097	7.743	2.655	6.195	
494	5.769	6.044	4.945	4.121	2.747	6.593	5.220	4.396	3.846	7.143	
495	4.503	5.253	5.253	7.129	3.377	9.006	3.940	4.690	2.251	3.752	
498	9.774	3.418	4.716	5.947	5.742	8.954	3.828	3.554	1.777	4.443	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	4.464	4.464	8.929	4.464	11.161	4.018	1.786	2.679	4.911	
2	…	6.566	3.030	6.566	0.505	3.535	5.051	2.525	2.020	11.111	
3	…	5.517	1.724	5.172	6.552	9.138	4.655	0.862	3.276	4.138	
4	…	4.116	2.179	7.022	5.327	9.443	7.990	1.453	4.358	6.295	
5	…	6.587	2.246	4.641	5.240	6.138	4.940	1.946	4.491	7.036	
..	…	…	…	…	…	…	…	…	…	…	
492	…	4.230	3.474	6.495	4.834	8.761	5.740	1.511	3.021	8.459	
493	…	5.973	1.991	4.204	4.425	6.637	5.088	0.885	4.867	7.301	
494	…	5.495	1.923	3.846	3.846	8.242	3.846	1.923	3.571	6.319	

```
495  …  4.690  2.064  5.816  3.940   9.944  4.128  1.689  3.752   6.567
498  …  5.400  2.324  3.349  6.357   8.954  4.511  0.547  1.162   6.835

              X20
0     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
5     no_efectores
..             …
492   no_efectores
493   no_efectores
494   no_efectores
495   no_efectores
498   no_efectores

[404 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

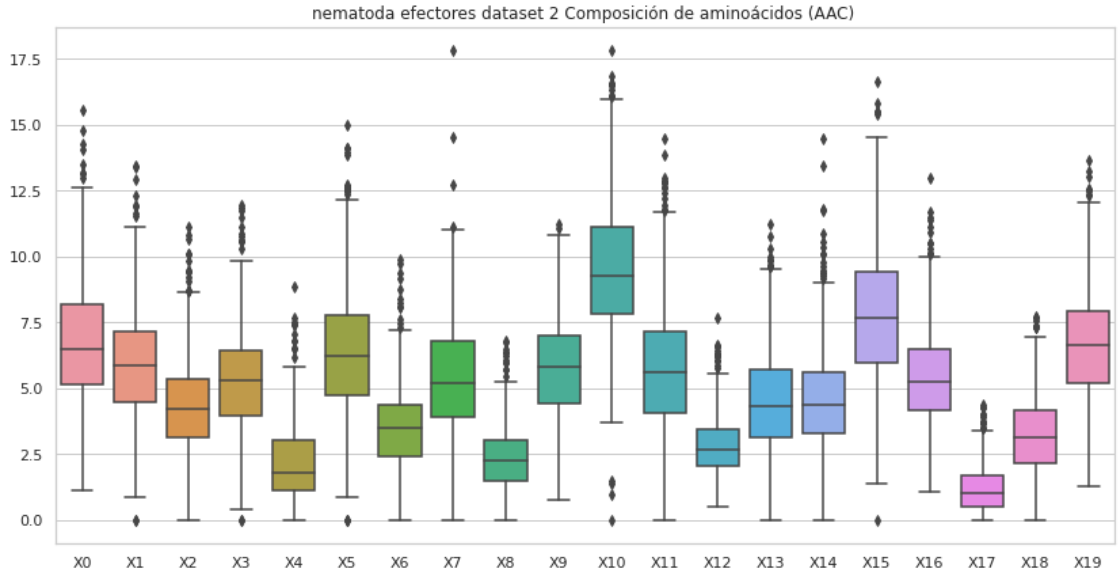
	X0	X1	X2	X3	X4	X5	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	6.072465	5.402129	5.193698	5.208042	2.113619	6.608928	
std	2.261114	2.153238	1.921941	1.707476	1.499090	2.706686	
min	0.847000	0.000000	0.800000	0.000000	0.000000	0.826000	
25%	4.670500	4.063250	3.874750	4.121000	1.149000	4.818000	
50%	5.846000	5.217500	4.939500	5.338500	1.778000	6.459500	
75%	7.195000	6.332750	6.289750	6.171000	2.716250	8.168250	
max	13.834000	13.187000	11.321000	10.596000	8.219000	15.517000	

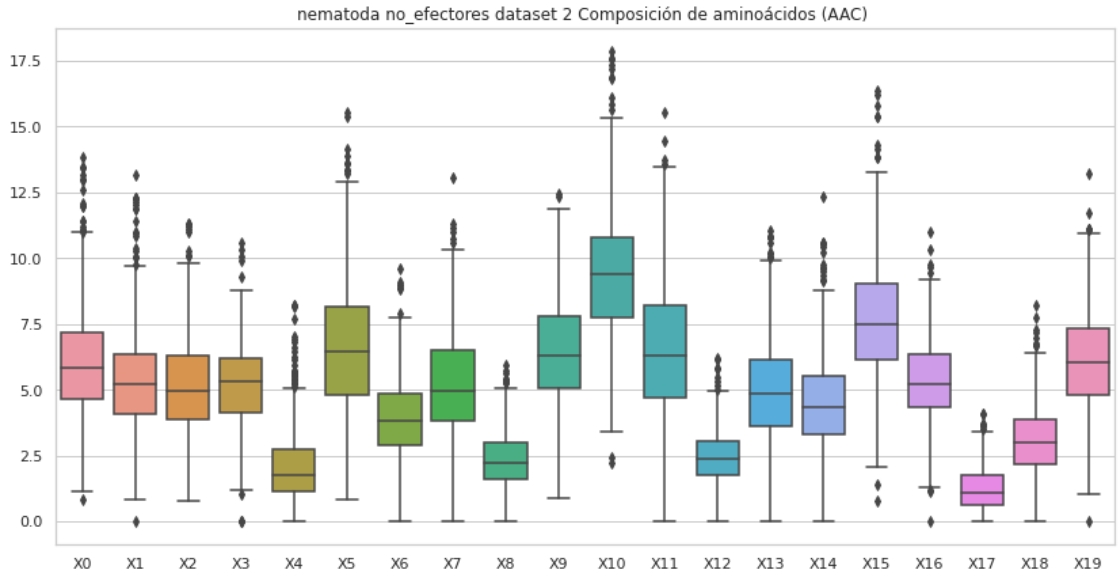
	X6	X7	X8	X9	X10	X11	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	3.968109	5.195535	2.307290	6.375631	9.445215	6.540616	
std	1.674127	2.018679	1.165503	2.049723	2.574510	2.582265	
min	0.000000	0.000000	0.000000	0.897000	2.242000	0.000000	
25%	2.870000	3.826500	1.578750	5.036250	7.749750	4.708750	
50%	3.810000	4.930000	2.228000	6.314000	9.378000	6.312500	
75%	4.828000	6.480750	2.992500	7.801000	10.813000	8.209500	
max	9.589000	13.043000	5.932000	12.452000	17.857000	15.517000	

	X12	X13	X14	X15	X16	X17	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	2.470196	5.047203	4.564797	7.679557	5.316453	1.223856	
std	1.010871	2.106332	1.908452	2.451105	1.597549	0.882013	
min	0.000000	0.000000	0.000000	0.787000	0.000000	0.000000	

25%	1.754000	3.603750	3.307500	6.163250	4.317000	0.629750
50%	2.368500	4.862500	4.348000	7.467000	5.201500	1.079000
75%	3.053750	6.143500	5.546250	9.012250	6.331250	1.762500
max	6.173000	11.047000	12.329000	16.352000	10.989000	4.068000

	X18	X19
count	404.000000	404.000000
mean	3.094735	6.171946
std	1.421695	2.091210
min	0.000000	0.000000
25%	2.145000	4.802250
50%	3.013000	6.050000
75%	3.879250	7.323750
max	8.213000	13.208000





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 nematoda
dataset 2, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.040396	0.008977	0.031419	0.000000	0.013465	0.035908	0.004488	
1	0.038366	0.020192	0.034327	0.033317	0.030289	0.029279	0.015144	
2	0.035370	0.004161	0.024274	0.027741	0.015258	0.019419	0.010403	
3	0.023973	0.035960	0.008298	0.015675	0.007376	0.023051	0.002766	
4	0.025031	0.002276	0.018205	0.025031	0.025031	0.020480	0.006827	
..	…	…	…	…	…	…	…	
495	0.023650	0.037342	0.021161	0.011203	0.009958	0.028629	0.007468	
496	0.012520	0.087640	0.025040	0.031300	0.025040	0.062600	0.031300	
497	0.027481	0.008678	0.021696	0.036160	0.018803	0.047731	0.013017	
498	0.024546	0.010228	0.030683	0.012273	0.008182	0.028637	0.008182	
499	0.024072	0.006018	0.015475	0.042126	0.006878	0.017194	0.005158	

	X7	X8	X9	…	X74	X75	X76	\
0	0.013465	0.008977	0.031419	…	-0.034659	-0.020911	0.008659	
1	0.037356	0.032308	0.064616	…	0.011141	0.017512	0.024784	
2	0.020113	0.022193	0.030516	…	-0.004863	0.005734	0.007552	
3	0.005532	0.007376	0.017519	…	0.001016	-0.002136	0.024227	
4	0.029583	0.006827	0.052338	…	-0.019335	-0.028585	-0.009569	
..	…	…	…	…	…	…	…	
495	0.016182	0.024895	0.007468	…	-0.000874	-0.001131	0.043695	
496	0.050080	0.068860	0.006260	…	-0.041295	-0.056304	0.003665	
497	0.050623	0.044838	0.053516	…	0.002126	0.037515	0.035376	
498	0.018410	0.012273	0.030683	…	0.024732	0.013592	0.027334	
499	0.019773	0.021493	0.026651	…	0.009287	0.028940	0.021130	

	X77	X78	X79	X80	X81	X82	X83
0	0.029195	0.022208	0.056263	-0.018757	-0.005883	0.026884	efectores
1	0.010862	-0.006575	0.000847	0.006473	0.009794	-0.005910	efectores
2	-0.001814	0.000832	0.012855	-0.005702	0.004960	0.010221	efectores
3	-0.005744	-0.005662	0.020793	-0.000685	0.007164	0.020006	efectores
4	0.033140	0.013122	0.021386	0.048924	0.026286	-0.029851	efectores
..	…	…	…	…	…	…	…
495	0.001489	-0.007676	0.010792	0.006028	0.008788	0.020217	efectores
496	0.034392	0.025112	-0.034088	-0.003367	-0.007701	0.070198	efectores

```
497 -0.005860  0.023509  0.008350 -0.015469 -0.000498  0.017781  efectores
498  0.017109  0.005775  0.042681  0.005137  0.028349  0.021954  efectores
499  0.015082  0.008012  0.021785  0.002439  0.015847  0.029849  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.035026	0.014032	0.028596	0.033470	0.026172	0.029261	
std	0.021542	0.018398	0.021901	0.024156	0.026558	0.018827	
min	-0.024404	0.000000	-0.073211	-0.048807	-0.073211	-0.024404	
25%	0.021886	0.004501	0.014081	0.017208	0.011633	0.016592	
50%	0.032163	0.008958	0.025187	0.030189	0.020885	0.025540	
75%	0.043528	0.017665	0.036452	0.043464	0.031150	0.038096	
max	0.197480	0.215735	0.167589	0.192282	0.329627	0.154846	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.014026	0.032218	0.031811	0.052203	…	0.009782	
std	0.015219	0.027202	0.025367	0.039913	…	0.028918	
min	-0.073211	-0.097615	-0.073211	-0.097615	…	-0.197318	
25%	0.005354	0.016321	0.015953	0.027234	…	-0.000719	
50%	0.010833	0.027710	0.027115	0.045620	…	0.009736	
75%	0.018039	0.041290	0.041614	0.065291	…	0.021993	
max	0.161801	0.302158	0.302158	0.357096	…	0.176885	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.001161	0.005630	0.011731	0.003880	0.006711	0.008927	
std	0.038225	0.030339	0.028263	0.042581	0.029312	0.027027	
min	-0.189542	-0.257527	-0.159129	-0.391272	-0.251625	-0.183585	
25%	-0.011242	-0.005279	-0.000037	-0.008533	-0.003861	-0.001061	
50%	0.002662	0.008055	0.011391	0.004974	0.007791	0.010473	
75%	0.017468	0.019983	0.023724	0.017184	0.019401	0.021327	
max	0.264887	0.155891	0.219974	0.344425	0.148443	0.134011	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.001647	0.004630	0.011447
std	0.040353	0.033523	0.026788
min	-0.339401	-0.301141	-0.128760
25%	-0.010808	-0.005921	0.000508
50%	0.003472	0.006659	0.010912

```
75%      0.016631    0.020516    0.022570
max      0.164607    0.124685    0.235183

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 2, con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.016992  0.002427  0.009710  0.010924  0.024275  0.027916  0.003641
1    0.023510  0.000000  0.016793  0.030228  0.010076  0.023510  0.026869
2    0.027050  0.007377  0.014755  0.022132  0.031969  0.031969  0.014755
3    0.019480  0.018089  0.037569  0.037569  0.041744  0.033395  0.030612
4    0.029942  0.009527  0.014971  0.020415  0.039469  0.008166  0.019054
..        …         …         …         …         …         …         …
495  0.024710  0.018533  0.039124  0.049420  0.031917  0.025740  0.012355
496  0.020769  0.002709  0.020769  0.021672  0.008127  0.018060  0.004515
497  0.010368  0.051840  0.017280  0.024192  0.027648  0.055296  0.003456
498  0.034289  0.020142  0.020861  0.031412  0.011749  0.012469  0.006234
499  0.004844  0.001321  0.002202  0.003523  0.026424  0.007047  0.002642

           X7        X8        X9  …       X74       X75       X76  \
0    0.023061  0.012137  0.027916  …  0.004525  0.003384 -0.002456
1    0.016793  0.016793  0.033586  …  0.016827 -0.002067  0.021211
2    0.034428  0.031969  0.083610  …  0.002137 -0.010547  0.007776
3    0.062616  0.044527  0.114100  …  0.013697  0.009938  0.001464
4    0.053079  0.023137  0.066688  … -0.026315 -0.010014 -0.003372
..        …         …         …  …         …         …         …
495  0.020592  0.025740  0.045302  …  0.000453  0.020947  0.005681
496  0.011739  0.013545  0.023478  … -0.009425  0.002633  0.015512
497  0.038016  0.055296  0.041472  … -0.009636 -0.008957 -0.007762
498  0.015586  0.018943  0.029493  … -0.004047  0.006407  0.012670
499  0.012331  0.005725  0.028626  …  0.022697  0.010784 -0.004667

          X77       X78       X79       X80       X81       X82           X83
0    0.004378  0.006896  0.021716  0.022090  0.008710  0.008143  no_efectores
1   -0.006079  0.015438  0.002684 -0.026162 -0.025301 -0.002550  no_efectores
2   -0.008629  0.005019 -0.020604  0.017886  0.007805  0.002220  no_efectores
3   -0.030484 -0.024147  0.014043  0.014695 -0.000611  0.004804  no_efectores
4    0.007643 -0.006692  0.026342  0.016641  0.009007 -0.002693  no_efectores
..        …         …         …         …         …         …           …
495  0.008172  0.018081  0.007305 -0.000700  0.007038  0.018774  no_efectores
496  0.001434  0.018016  0.012346  0.012178  0.006855  0.013556  no_efectores
497 -0.025315 -0.007538 -0.002534  0.028767  0.000512 -0.015347  no_efectores
```

```
498  0.001032  0.012185  0.022461 -0.001108  0.004431  0.015035  no_efectores
499  0.028337  0.015488 -0.000646  0.024374  0.013074 -0.002364  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 2, con valores atípicos.
Estadísticas.

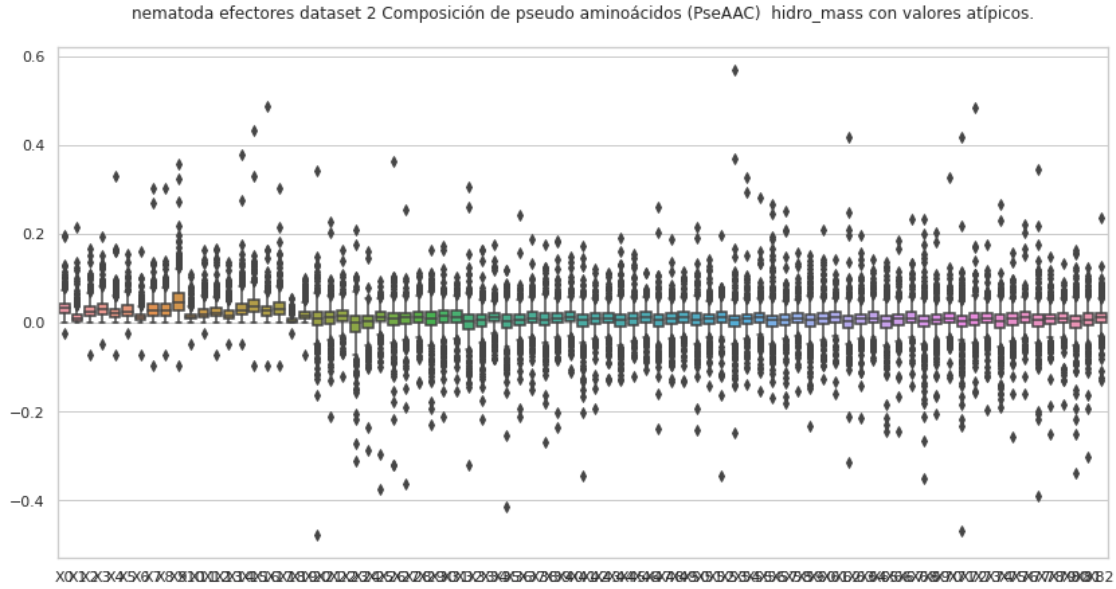
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.029504	0.010836	0.026086	0.033537	0.026022	0.025439	
std	0.023301	0.032034	0.024966	0.024317	0.026099	0.032914	
min	-0.313772	-0.627545	-0.313772	-0.000000	-0.313772	-0.627545	
25%	0.018779	0.004408	0.013590	0.018585	0.012560	0.015566	
50%	0.027383	0.008271	0.023256	0.030762	0.022570	0.023891	
75%	0.036750	0.014923	0.035489	0.042197	0.033581	0.035209	
max	0.107286	0.166843	0.250264	0.333685	0.208553	0.106380	

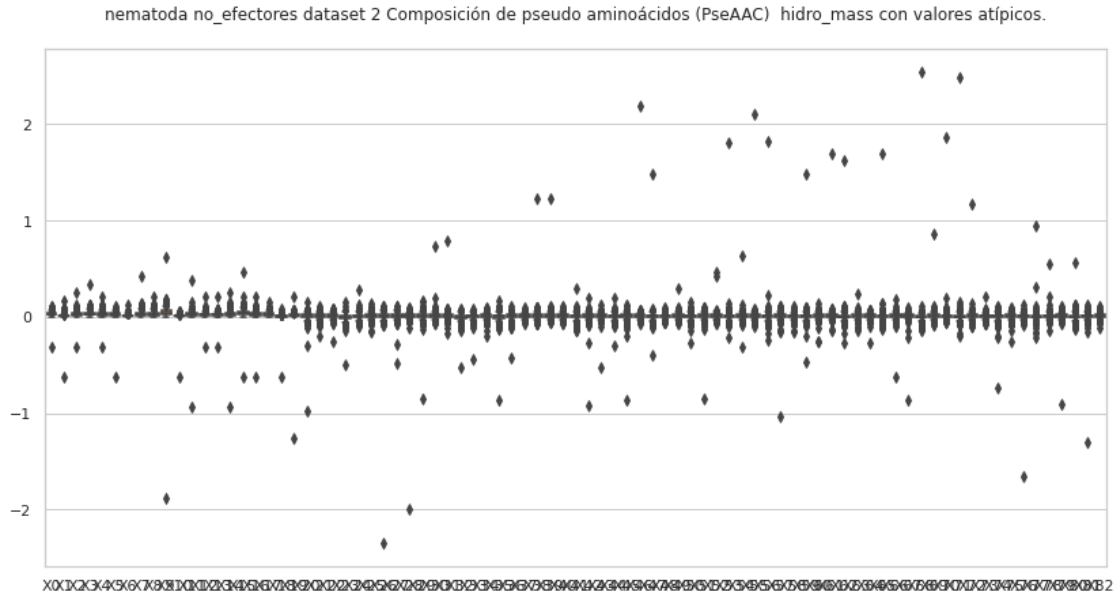
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.012464	0.034317	0.033492	0.045341	…	0.009351	
std	0.011016	0.028799	0.023676	0.094708	…	0.019757	
min	-0.000000	-0.000000	-0.000000	-1.882634	…	-0.069143	
25%	0.004894	0.016695	0.017896	0.027866	…	0.000101	
50%	0.010110	0.028434	0.028886	0.041745	…	0.008675	
75%	0.016976	0.043964	0.042599	0.062736	…	0.018280	
max	0.125132	0.417106	0.208553	0.625660	…	0.214273	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.000937	0.007613	0.005734	0.003694	0.008619	0.007841	
std	0.043992	0.026339	0.076622	0.051958	0.033678	0.045012	
min	-0.736537	-0.263752	-1.650320	-0.221613	-0.140266	-0.909004	
25%	-0.007120	-0.000692	0.000102	-0.007237	-0.002681	0.000077	
50%	0.005014	0.008295	0.008559	0.003694	0.008802	0.008821	
75%	0.015517	0.019554	0.018231	0.013168	0.018031	0.018977	
max	0.111025	0.090784	0.118630	0.944926	0.542137	0.103922	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.003265	0.004248	0.009099
std	0.036082	0.062760	0.018297
min	-0.135559	-1.300567	-0.110793
25%	-0.008593	-0.002965	0.000381
50%	0.002885	0.006927	0.008330
75%	0.014783	0.018282	0.018956

```
max       0.560350     0.118929     0.104230

[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 nematoda
dataset 2, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.040396	0.008977	0.031419	0.000000	0.013465	0.035908	0.004488	
1	0.038366	0.020192	0.034327	0.033317	0.030289	0.029279	0.015144	
2	0.035370	0.004161	0.024274	0.027741	0.015258	0.019419	0.010403	
3	0.023973	0.035960	0.008298	0.015675	0.007376	0.023051	0.002766	
4	0.025031	0.002276	0.018205	0.025031	0.025031	0.020480	0.006827	
..	…	…	…	…	…	…	…	
494	0.004751	0.033254	0.019002	0.042755	0.009501	0.000000	0.009501	
495	0.023650	0.037342	0.021161	0.011203	0.009958	0.028629	0.007468	
497	0.027481	0.008678	0.021696	0.036160	0.018803	0.047731	0.013017	
498	0.024546	0.010228	0.030683	0.012273	0.008182	0.028637	0.008182	
499	0.024072	0.006018	0.015475	0.042126	0.006878	0.017194	0.005158	

	X7	X8	X9	…	X74	X75	X76	\
0	0.013465	0.008977	0.031419	…	-0.034659	-0.020911	0.008659	
1	0.037356	0.032308	0.064616	…	0.011141	0.017512	0.024784	
2	0.020113	0.022193	0.030516	…	-0.004863	0.005734	0.007552	
3	0.005532	0.007376	0.017519	…	0.001016	-0.002136	0.024227	
4	0.029583	0.006827	0.052338	…	-0.019335	-0.028585	-0.009569	
..	…	…	…	…	…	…	…	
494	0.014252	0.019002	0.023753	…	-0.045924	0.009083	-0.021728	
495	0.016182	0.024895	0.007468	…	-0.000874	-0.001131	0.043695	
497	0.050623	0.044838	0.053516	…	0.002126	0.037515	0.035376	
498	0.018410	0.012273	0.030683	…	0.024732	0.013592	0.027334	
499	0.019773	0.021493	0.026651	…	0.009287	0.028940	0.021130	

	X77	X78	X79	X80	X81	X82	X83
0	0.029195	0.022208	0.056263	-0.018757	-0.005883	0.026884	efectores
1	0.010862	-0.006575	0.000847	0.006473	0.009794	-0.005910	efectores
2	-0.001814	0.000832	0.012855	-0.005702	0.004960	0.010221	efectores
3	-0.005744	-0.005662	0.020793	-0.000685	0.007164	0.020006	efectores
4	0.033140	0.013122	0.021386	0.048924	0.026286	-0.029851	efectores
..	…	…	…	…	…	…	
494	-0.056796	-0.047375	-0.012600	-0.031787	-0.017206	0.008843	efectores
495	0.001489	-0.007676	0.010792	0.006028	0.008788	0.020217	efectores
497	-0.005860	0.023509	0.008350	-0.015469	-0.000498	0.017781	efectores
498	0.017109	0.005775	0.042681	0.005137	0.028349	0.021954	efectores
499	0.015082	0.008012	0.021785	0.002439	0.015847	0.029849	efectores

[412 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	412.000000	412.000000	412.000000	412.000000	412.000000	412.000000	
mean	0.030922	0.010709	0.024278	0.028944	0.020109	0.025301	
std	0.014456	0.009688	0.013972	0.016294	0.012514	0.012807	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.020389	0.004137	0.013703	0.016306	0.010435	0.015781	
50%	0.030056	0.007908	0.023537	0.027939	0.018784	0.023036	
75%	0.039062	0.014402	0.032932	0.039216	0.028084	0.033073	
max	0.087821	0.063696	0.083772	0.096798	0.065796	0.066024	

	X6	X7	X8	X9	…	X73	\
count	412.000000	412.000000	412.000000	412.000000	…	412.000000	
mean	0.011435	0.027352	0.026643	0.043129	…	0.011993	
std	0.008668	0.016566	0.015442	0.023686	…	0.017764	
min	0.000000	0.000000	0.000000	0.000000	…	-0.047539	
25%	0.004959	0.015058	0.014481	0.025925	…	0.001888	
50%	0.009780	0.025519	0.023829	0.040573	…	0.010201	
75%	0.015402	0.035939	0.036074	0.056370	…	0.021688	
max	0.051931	0.092031	0.079434	0.150186	…	0.090289	

	X74	X75	X76	X77	X78	X79	\
count	412.000000	412.000000	412.000000	412.000000	412.000000	412.000000	
mean	0.004454	0.008434	0.012192	0.004719	0.007803	0.011259	
std	0.021988	0.017881	0.016521	0.020832	0.017838	0.017096	
min	-0.078879	-0.073147	-0.038545	-0.089484	-0.079017	-0.051342	
25%	-0.006557	-0.001529	0.001575	-0.005823	-0.002440	0.001046	
50%	0.003625	0.009314	0.011573	0.005014	0.007961	0.010757	
75%	0.016349	0.019439	0.023112	0.016469	0.018706	0.020820	
max	0.080212	0.081819	0.074978	0.108522	0.070940	0.074097	

	X80	X81	X82
count	412.000000	412.000000	412.000000
mean	0.003172	0.007484	0.012847
std	0.020919	0.019563	0.016791
min	-0.064022	-0.084056	-0.055769
25%	-0.007857	-0.003416	0.002272
50%	0.004146	0.007197	0.011181
75%	0.015224	0.019855	0.022013
max	0.072873	0.074439	0.086609

[8 rows x 83 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 2, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.016992	0.002427	0.009710	0.010924	0.024275	0.027916	0.003641	
1	0.023510	0.000000	0.016793	0.030228	0.010076	0.023510	0.026869	
2	0.027050	0.007377	0.014755	0.022132	0.031969	0.031969	0.014755	
3	0.019480	0.018089	0.037569	0.037569	0.041744	0.033395	0.030612	
4	0.029942	0.009527	0.014971	0.020415	0.039469	0.008166	0.019054	
..	…	…	…	…	…	…	…	
495	0.024710	0.018533	0.039124	0.049420	0.031917	0.025740	0.012355	
496	0.020769	0.002709	0.020769	0.021672	0.008127	0.018060	0.004515	
497	0.010368	0.051840	0.017280	0.024192	0.027648	0.055296	0.003456	
498	0.034289	0.020142	0.020861	0.031412	0.011749	0.012469	0.006234	
499	0.004844	0.001321	0.002202	0.003523	0.026424	0.007047	0.002642	

	X7	X8	X9	…	X74	X75	X76	\
0	0.023061	0.012137	0.027916	…	0.004525	0.003384	-0.002456	
1	0.016793	0.016793	0.033586	…	0.016827	-0.002067	0.021211	
2	0.034428	0.031969	0.083610	…	0.002137	-0.010547	0.007776	
3	0.062616	0.044527	0.114100	…	0.013697	0.009938	0.001464	
4	0.053079	0.023137	0.066688	…	-0.026315	-0.010014	-0.003372	
..	…	…	…	…	…	…	…	
495	0.020592	0.025740	0.045302	…	0.000453	0.020947	0.005681	
496	0.011739	0.013545	0.023478	…	-0.009425	0.002633	0.015512	
497	0.038016	0.055296	0.041472	…	-0.009636	-0.008957	-0.007762	
498	0.015586	0.018943	0.029493	…	-0.004047	0.006407	0.012670	
499	0.012331	0.005725	0.028626	…	0.022697	0.010784	-0.004667	

	X77	X78	X79	X80	X81	X82	X83
0	0.004378	0.006896	0.021716	0.022090	0.008710	0.008143	no_efectores
1	-0.006079	0.015438	0.002684	-0.026162	-0.025301	-0.002550	no_efectores
2	-0.008629	0.005019	-0.020604	0.017886	0.007805	0.002220	no_efectores
3	-0.030484	-0.024147	0.014043	0.014695	-0.000611	0.004804	no_efectores
4	0.007643	-0.006692	0.026342	0.016641	0.009007	-0.002693	no_efectores
..	…	…	…	…	…	…	…
495	0.008172	0.018081	0.007305	-0.000700	0.007038	0.018774	no_efectores
496	0.001434	0.018016	0.012346	0.012178	0.006855	0.013556	no_efectores
497	-0.025315	-0.007538	-0.002534	0.028767	0.000512	-0.015347	no_efectores
498	0.001032	0.012185	0.022461	-0.001108	0.004431	0.015035	no_efectores
499	0.028337	0.015488	-0.000646	0.024374	0.013074	-0.002364	no_efectores

[442 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 2, sin valores atípicos.
Estadísticas.

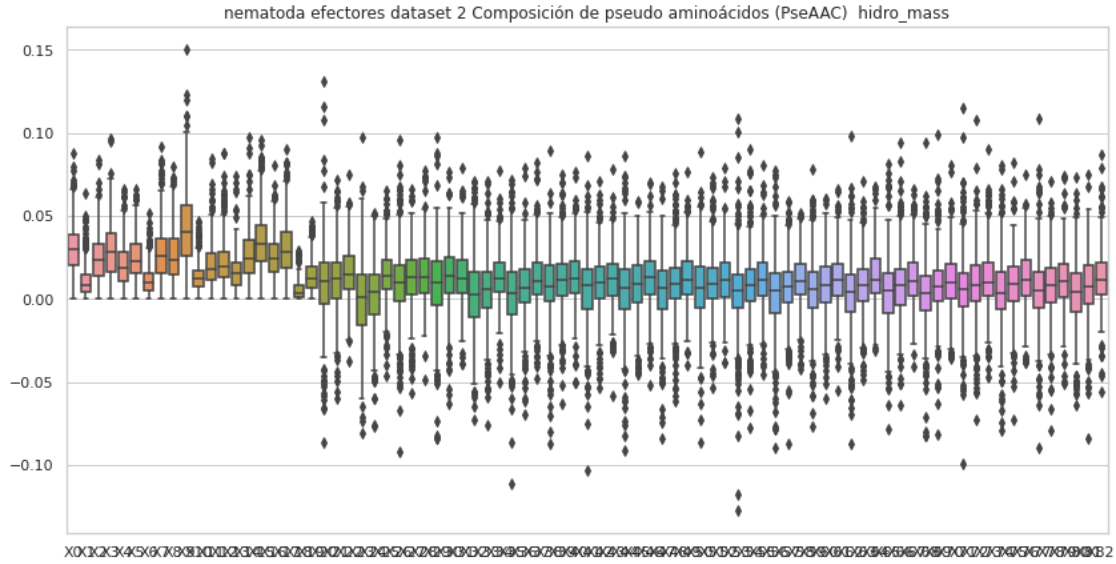
	X0	X1	X2	X3	X4	X5	\
count	442.000000	442.000000	442.000000	442.000000	442.000000	442.000000	
mean	0.027371	0.010597	0.023774	0.029996	0.022913	0.024753	
std	0.013492	0.010986	0.013715	0.016384	0.014401	0.012686	
min	0.000000	0.000000	0.000000	0.000684	0.000000	0.000000	
25%	0.018293	0.004275	0.012861	0.017336	0.012075	0.014841	
50%	0.025990	0.007706	0.021597	0.028441	0.021137	0.023066	
75%	0.034406	0.013812	0.033358	0.039365	0.030609	0.032715	
max	0.072505	0.084879	0.075322	0.094677	0.081102	0.063128	

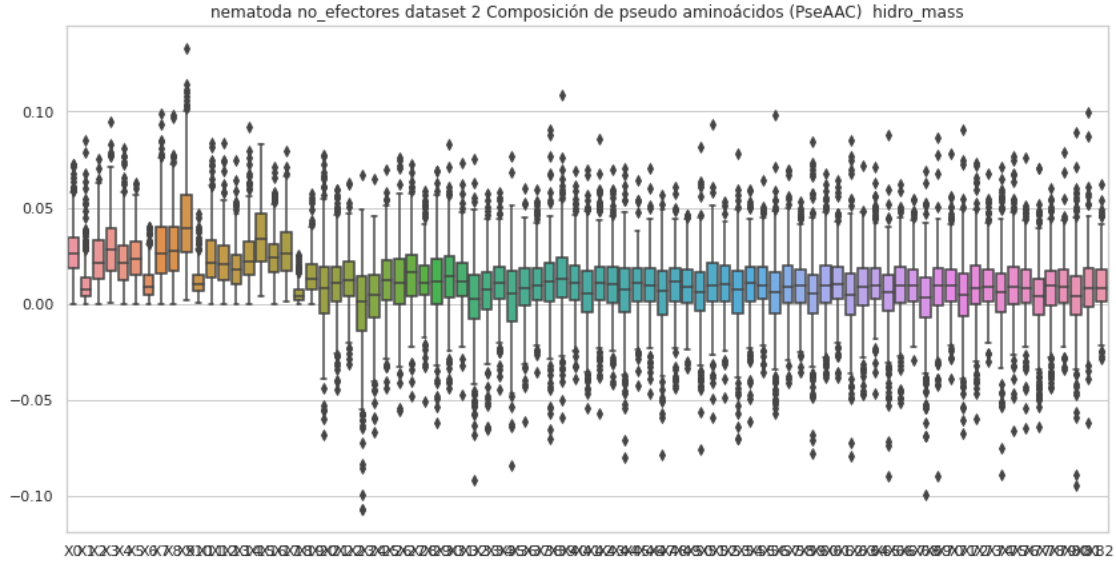
	X6	X7	X8	X9	…	X73	\
count	442.000000	442.000000	442.000000	442.000000	…	442.000000	
mean	0.010724	0.029603	0.029639	0.043186	…	0.010007	
std	0.007808	0.017905	0.017032	0.022711	…	0.013969	
min	0.000000	0.000000	0.000000	0.001845	…	-0.030125	
25%	0.004662	0.015740	0.017222	0.027061	…	0.001387	
50%	0.008880	0.026380	0.027405	0.039662	…	0.008813	
75%	0.015201	0.039978	0.039872	0.056690	…	0.017722	
max	0.040020	0.099058	0.098432	0.132875	…	0.067684	

	X74	X75	X76	X77	X78	X79	\
count	442.000000	442.000000	442.000000	442.000000	442.000000	442.000000	
mean	0.005023	0.009940	0.008712	0.002615	0.007617	0.009774	
std	0.019413	0.016811	0.015160	0.018662	0.017015	0.014099	
min	-0.089230	-0.066025	-0.064341	-0.064130	-0.044214	-0.033992	
25%	-0.004537	0.000108	0.000439	-0.006037	-0.000875	0.000521	
50%	0.005701	0.008883	0.008266	0.003867	0.009342	0.008856	
75%	0.015479	0.019216	0.017832	0.012639	0.017239	0.017880	
max	0.070983	0.077010	0.076031	0.070782	0.060344	0.078539	

	X80	X81	X82
count	442.000000	442.000000	442.000000
mean	0.003781	0.008473	0.009576
std	0.019241	0.016889	0.013666
min	-0.094392	-0.062195	-0.028525
25%	-0.005744	-0.001807	0.001527
50%	0.003900	0.007812	0.008271
75%	0.014693	0.018369	0.017909
max	0.089380	0.099853	0.062360

[8 rows x 83 columns]


nematoda efectores dataset 2 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.15
0.10
0.05
0.00
−0.05
−0.10


nematoda no_efectores dataset 2 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.10
0.05
0.00
−0.05
−0.10

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 nematoda dataset 2,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.043307  0.009624  0.033683  0.000000  0.014436  0.038495  0.004812
1    0.042765  0.022508  0.038263  0.037138  0.033761  0.032636  0.016881
2    0.045888  0.005399  0.031492  0.035990  0.019795  0.025193  0.013496
3    0.025753  0.038630  0.008915  0.016839  0.007924  0.024763  0.002972
4    0.048083  0.004371  0.034969  0.048083  0.048083  0.039341  0.013114
..        …         …         …         …         …         …         …
495  0.025570  0.040373  0.022878  0.012112  0.010766  0.030953  0.008075
```

```
496  0.014258  0.099808  0.028516  0.035646  0.028516  0.071291  0.035646
497  0.027640  0.008728  0.021821  0.036368  0.018911  0.048006  0.013092
498  0.028525  0.011886  0.035657  0.014263  0.009508  0.033280  0.009508
499  0.033791  0.008448  0.021723  0.059135  0.009655  0.024137  0.007241

           X7        X8        X9  …       X32       X33       X34  \
0    0.014436  0.009624  0.033683  …  0.029468  0.006771 -0.016933
1    0.041639  0.036012  0.072024  …  0.000682  0.011295  0.017977
2    0.026093  0.028792  0.039589  …  0.034140  0.028684  0.018508
3    0.005943  0.007924  0.018820  …  0.029950  0.027727  0.037623
4    0.056825  0.013114  0.100537  …  0.010711  0.074836  0.007183
..        …         …         …  …         …         …         …
495  0.017495  0.026916  0.008075  …  0.024157  0.027050  0.042969
496  0.057033  0.078420  0.007129  … -0.023733  0.032797  0.064658
497  0.050915  0.045096  0.053825  …  0.018611  0.035137 -0.000164
498  0.021394  0.014263  0.035657  …  0.040637  0.013316  0.031709
499  0.027757  0.030171  0.037412  …  0.007688  0.026500  0.019044

           X35       X36       X37       X38       X39       X40         X41
0    0.026668  0.038044  0.004171  0.009283  0.060316  0.028821  efectores
1    0.002736  0.003929  0.022341  0.027626  0.000945 -0.006588  efectores
2    0.027734  0.027339  0.023605  0.009798  0.016678  0.013260  efectores
3    0.031189  0.029797  0.030261  0.026025  0.022337  0.021491  efectores
4    0.016562  0.009495  0.025954 -0.018381  0.041080 -0.057341  efectores
..        …         …         …         …         …         …         …
495  0.060019  0.030931  0.019357  0.047242  0.011669  0.021858  efectores
496  0.001128 -0.004069  0.066157  0.004174 -0.038820  0.079944  efectores
497  0.034526  0.037499  0.018686  0.035580  0.008398  0.017883  efectores
498  0.057614  0.032870  0.005631  0.031765  0.049600  0.025513  efectores
499  0.016579  0.027770  0.020932  0.029662  0.030582  0.041902  efectores

[500 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.044595    0.016581    0.036183    0.044347    0.032495    0.036508
std      0.017992    0.015110    0.020183    0.027555    0.021807    0.015448
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.033290    0.006843    0.023149    0.025140    0.017445    0.026232
50%      0.043642    0.012463    0.033980    0.040252    0.027791    0.034531
75%      0.054408    0.022517    0.045700    0.056793    0.043222    0.044406
max      0.131000    0.109637    0.119309    0.162298    0.152973    0.117738
```

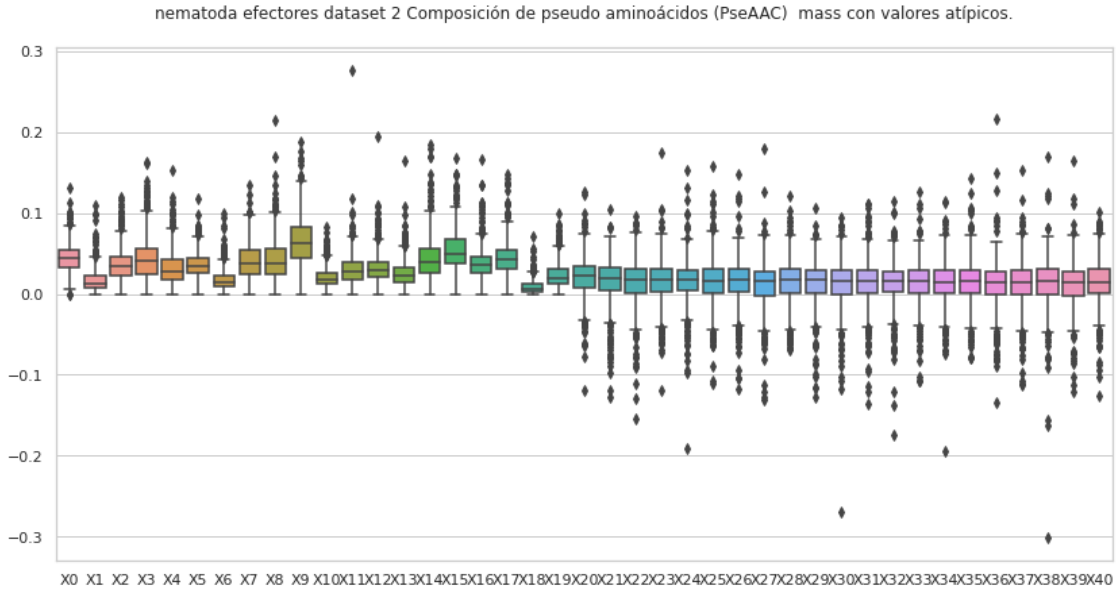
```
              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.017401    0.040605    0.042276    0.065012  …    0.013156
std      0.013298    0.021572    0.026848    0.031051  …    0.028990
min      0.000000    0.000000    0.000000    0.000000  …   -0.135811
25%      0.008942    0.024872    0.024078    0.044234  …    0.000657
50%      0.014749    0.038301    0.037463    0.062296  …    0.015286
75%      0.023220    0.054680    0.055384    0.082944  …    0.028583
max      0.100354    0.133806    0.214168    0.188097  …    0.111610

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.013087    0.014374    0.014312    0.014688    0.012393    0.012618
std      0.027532    0.027068    0.028276    0.027510    0.029944    0.029588
min     -0.174600   -0.107489   -0.194879   -0.079616   -0.134111   -0.113452
25%      0.001980    0.000916    0.000882    0.000627   -0.000687   -0.001014
50%      0.015588    0.016401    0.014855    0.015513    0.013560    0.014842
75%      0.028357    0.029241    0.030057    0.029802    0.028091    0.029181
max      0.114534    0.126861    0.114446    0.142614    0.217046    0.152680

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.013054    0.011979    0.014240
std      0.032661    0.028471    0.027686
min     -0.301530   -0.121783   -0.125489
25%     -0.000360   -0.001937    0.000792
50%      0.015343    0.014455    0.015073
75%      0.030926    0.028246    0.030712
max      0.168996    0.164865    0.100988

[8 rows x 41 columns]
```

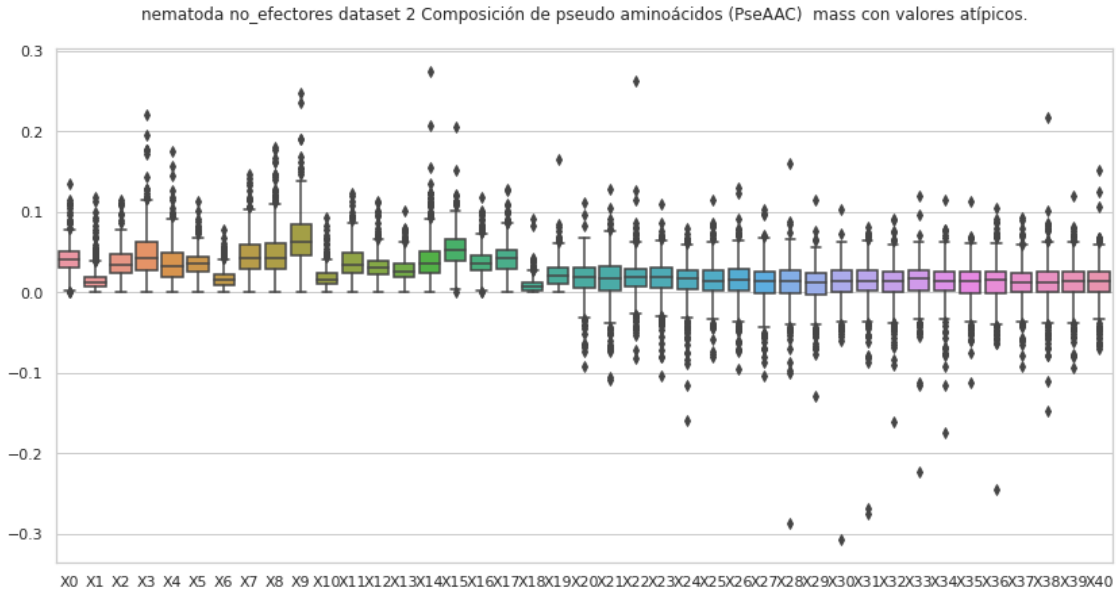
no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.034627  0.004947  0.019787  0.022260  0.049468  0.056888  0.007420
1    0.048338  0.000000  0.034527  0.062149  0.020716  0.048338  0.055244
2    0.049676  0.013548  0.027096  0.040644  0.058708  0.058708  0.027096
3    0.019340  0.017958  0.037298  0.037298  0.041442  0.033154  0.030391
4    0.045678  0.014534  0.022839  0.031144  0.060212  0.012458  0.029068
..        …         …         …         …         …         …         …
495  0.034014  0.025510  0.053855  0.068028  0.043935  0.035431  0.017007
496  0.033325  0.004347  0.033325  0.034774  0.013040  0.028978  0.007245
```

```
497  0.014013  0.070067  0.023356  0.032698  0.037369  0.074738  0.004671
498  0.044574  0.026183  0.027118  0.040833  0.015274  0.016209  0.008104
499  0.032116  0.008759  0.014598  0.023357  0.175176  0.046714  0.017518

           X7        X8        X9  …       X32       X33       X34  \
0    0.046994  0.024734  0.056888  …  0.028875 -0.020245 -0.008724
1    0.034527  0.034527  0.069055  … -0.028712 -0.055522 -0.033364
2    0.063224  0.058708  0.153544  … -0.029635  0.021992  0.025927
3    0.062163  0.044205  0.113275  … -0.007813 -0.004888  0.031515
4    0.080974  0.035296  0.101737  …  0.003687  0.015463 -0.016523
..        …         …         …  …       …         …         …
495  0.028345  0.035431  0.062359  … -0.010329  0.020742  0.013356
496  0.018836  0.021734  0.037672  …  0.017081  0.019761  0.029700
497  0.051382  0.074738  0.056053  …  0.023603  0.026805  0.038473
498  0.020261  0.024625  0.038340  …  0.024787  0.023389  0.017417
499  0.081749  0.037955  0.189774  …  0.014798  0.031644 -0.001021

          X35       X36       X37       X38       X39       X40           X41
0    0.068993  0.021776  0.019051 -0.005004  0.044253  0.016594  no_efectores
1    0.021601 -0.020367 -0.035039  0.043610  0.005519 -0.005243  no_efectores
2   -0.000578 -0.026345  0.016201  0.014279 -0.037838  0.004077  no_efectores
3    0.020548  0.010312  0.023627  0.001454  0.013942  0.004769  no_efectores
4    0.005485 -0.030424 -0.026392 -0.005144  0.040186 -0.004108  no_efectores
..        …         …         …         …         …         …           …
495  0.028819  0.003969 -0.002085  0.007820  0.010056  0.025843  no_efectores
496  0.023251  0.028827  0.022656  0.024889  0.019810  0.021751  no_efectores
497  0.045391 -0.017533  0.008231 -0.010491 -0.003425 -0.020742  no_efectores
498  0.026812  0.027117  0.031449  0.016471  0.029198  0.019545  no_efectores
499 -0.013136 -0.063045  0.015972 -0.030941 -0.004283 -0.015674  no_efectores

[500 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.041770    0.016149    0.036506    0.048643    0.037199    0.036465
std      0.017998    0.015704    0.018450    0.030260    0.024082    0.014605
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.031053    0.006717    0.023741    0.027577    0.019438    0.026363
50%      0.040376    0.012319    0.035012    0.043211    0.032870    0.035657
75%      0.050481    0.019768    0.047367    0.062432    0.049104    0.044713
max      0.135424    0.118052    0.114382    0.220064    0.175176    0.113267

                X6          X7          X8          X9  …         X31  \
```

```
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016901    0.045803    0.048053    0.067454  …    0.011876
std      0.011473    0.024177    0.029655    0.031686  …    0.029550
min      0.000000    0.000000    0.000000    0.000000  …   -0.275799
25%      0.009290    0.028648    0.029040    0.046588  …    0.001615
50%      0.015159    0.042108    0.042469    0.063028  …    0.013566
75%      0.022306    0.059305    0.061995    0.083983  …    0.027250
max      0.078461    0.146497    0.179863    0.247366  …    0.080943

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011273    0.013552    0.012629    0.011364    0.012635    0.012029
std      0.024530    0.025622    0.024986    0.023032    0.026321    0.022380
min     -0.161238   -0.222784   -0.173412   -0.111999   -0.245101   -0.091224
25%     -0.000013    0.002394    0.001808   -0.000553   -0.000262    0.000205
50%      0.014019    0.016712    0.013687    0.013724    0.015701    0.013204
75%      0.025161    0.027822    0.026127    0.025544    0.026455    0.024219
max      0.091261    0.120458    0.114528    0.113755    0.104806    0.092766

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.012000    0.012635    0.012685
std      0.027226    0.024008    0.023508
min     -0.146899   -0.094331   -0.070375
25%      0.000295    0.000216    0.001124
50%      0.012455    0.014167    0.014153
75%      0.025669    0.025589    0.025136
max      0.216842    0.119438    0.152184

[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 nematoda dataset 2,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.043307  0.009624  0.033683  0.000000  0.014436  0.038495  0.004812
1    0.042765  0.022508  0.038263  0.037138  0.033761  0.032636  0.016881
2    0.045888  0.005399  0.031492  0.035990  0.019795  0.025193  0.013496
3    0.025753  0.038630  0.008915  0.016839  0.007924  0.024763  0.002972
```

```
4    0.048083  0.004371  0.034969  0.048083  0.048083  0.039341  0.013114
..        …         …         …         …         …         …
492  0.058345  0.010441  0.038692  0.082296  0.021495  0.023338  0.017810
495  0.025570  0.040373  0.022878  0.012112  0.010766  0.030953  0.008075
497  0.027640  0.008728  0.021821  0.036368  0.018911  0.048006  0.013092
498  0.028525  0.011886  0.035657  0.014263  0.009508  0.033280  0.009508
499  0.033791  0.008448  0.021723  0.059135  0.009655  0.024137  0.007241

           X7        X8        X9  …       X32       X33       X34  \
0    0.014436  0.009624  0.033683  …  0.029468  0.006771 -0.016933
1    0.041639  0.036012  0.072024  …  0.000682  0.011295  0.017977
2    0.026093  0.028792  0.039589  …  0.034140  0.028684  0.018508
3    0.005943  0.007924  0.018820  …  0.029950  0.027727  0.037623
4    0.056825  0.013114  0.100537  …  0.010711  0.074836  0.007183
..        …         …         …  …         …         …         …
492  0.031936  0.038692  0.058345  …  0.008673  0.014651 -0.004425
495  0.017495  0.026916  0.008075  …  0.024157  0.027050  0.042969
497  0.050915  0.045096  0.053825  …  0.018611  0.035137 -0.000164
498  0.021394  0.014263  0.035657  …  0.040637  0.013316  0.031709
499  0.027757  0.030171  0.037412  …  0.007688  0.026500  0.019044

           X35       X36       X37       X38       X39       X40         X41
0    0.026668  0.038044  0.004171  0.009283  0.060316  0.028821  efectores
1    0.002736  0.003929  0.022341  0.027626  0.000945 -0.006588  efectores
2    0.027734  0.027339  0.023605  0.009798  0.016678  0.013260  efectores
3    0.031189  0.029797  0.030261  0.026025  0.022337  0.021491  efectores
4    0.016562  0.009495  0.025954 -0.018381  0.041080 -0.057341  efectores
..        …         …         …         …         …         …         …
492  0.017394  0.005481  0.019756  0.013222  0.002491  0.000765  efectores
495  0.060019  0.030931  0.019357  0.047242  0.011669  0.021858  efectores
497  0.034526  0.037499  0.018686  0.035580  0.008398  0.017883  efectores
498  0.057614  0.032870  0.005631  0.031765  0.049600  0.025513  efectores
499  0.016579  0.027770  0.020932  0.029662  0.030582  0.041902  efectores

[402 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean     0.043068    0.014697    0.034426    0.041513    0.028834    0.034849
std      0.015039    0.011293    0.017150    0.021978    0.016804    0.012147
min      0.006130    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.033586    0.006679    0.023101    0.025298    0.016770    0.026558
50%      0.043154    0.011780    0.033011    0.039349    0.026070    0.034111
```

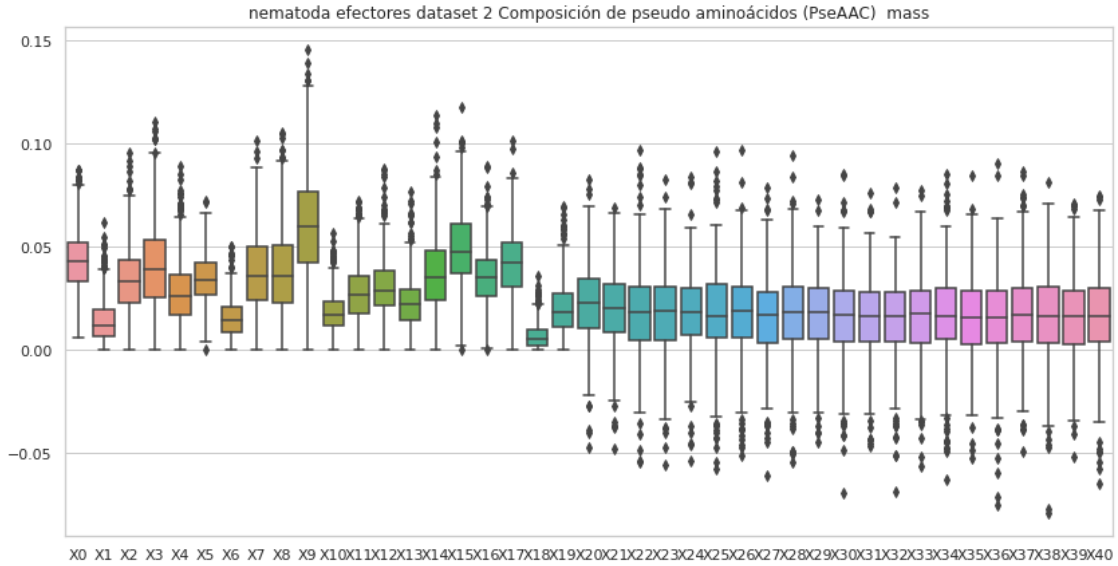
```
75%      0.052389    0.019849    0.043757    0.053528    0.036315    0.042664
max      0.087449    0.061667    0.095439    0.110487    0.089392    0.072488

               X6          X7          X8          X9  …         X31  \
count  402.000000  402.000000  402.000000  402.000000  …  402.000000
mean     0.015495    0.037949    0.038632    0.060801  …    0.015701
std      0.009497    0.018979    0.021226    0.027181  …    0.019215
min      0.000000    0.000000    0.000000    0.000000  …   -0.046841
25%      0.008780    0.024025    0.022794    0.042314  …    0.004019
50%      0.014149    0.036033    0.035950    0.059696  …    0.016585
75%      0.020720    0.050354    0.050863    0.077015  …    0.028011
max      0.050369    0.101483    0.105177    0.145561  …    0.075866

               X32         X33         X34         X35         X36         X37  \
count  402.000000  402.000000  402.000000  402.000000  402.000000  402.000000
mean     0.014953    0.016203    0.016769    0.015314    0.015579    0.016723
std      0.019686    0.021261    0.021937    0.020304    0.021352    0.021166
min     -0.068653   -0.056275   -0.063120   -0.052709   -0.074886   -0.049398
25%      0.003937    0.003632    0.005290    0.002784    0.003298    0.004185
50%      0.016489    0.017731    0.016188    0.015680    0.015655    0.017036
75%      0.028155    0.029030    0.029884    0.028720    0.028842    0.029764
max      0.078633    0.077106    0.085492    0.084755    0.090175    0.086326

               X38         X39         X40
count  402.000000  402.000000  402.000000
mean     0.016074    0.015418    0.016432
std      0.021628    0.020285    0.021234
min     -0.079064   -0.051661   -0.064561
25%      0.003159    0.002481    0.003834
50%      0.016167    0.016592    0.016431
75%      0.030985    0.028581    0.030223
max      0.081642    0.070635    0.074628

[8 rows x 41 columns]
```

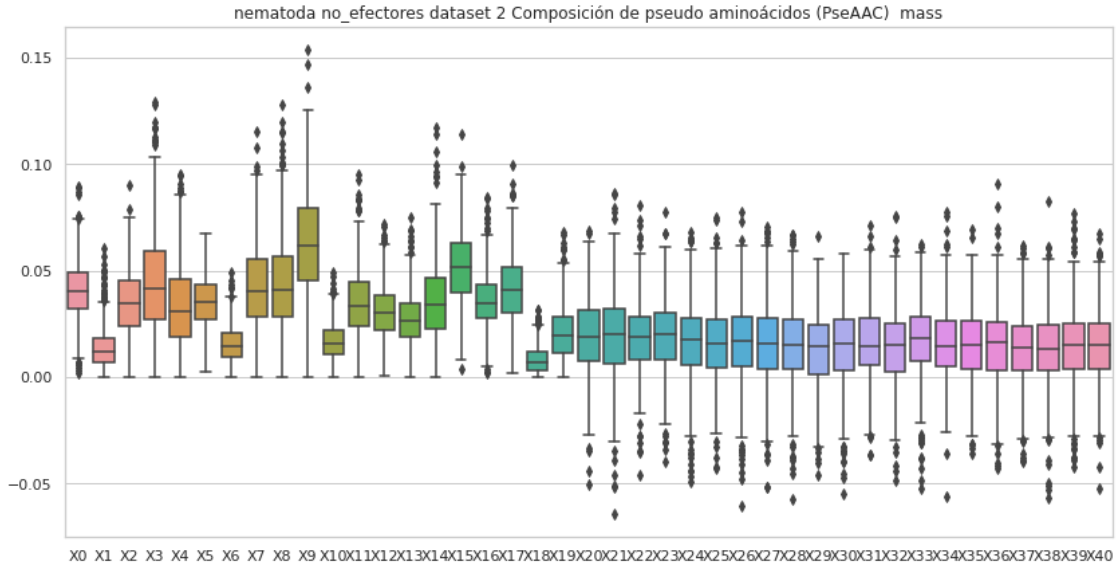
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
0    0.034627  0.004947  0.019787  0.022260  0.049468  0.056888  0.007420
2    0.049676  0.013548  0.027096  0.040644  0.058708  0.058708  0.027096
3    0.019340  0.017958  0.037298  0.037298  0.041442  0.033154  0.030391
4    0.045678  0.014534  0.022839  0.031144  0.060212  0.012458  0.029068
5    0.038173  0.015269  0.049079  0.046898  0.033810  0.057804  0.019632
..        …         …         …         …         …         …         …
```

```
493  0.045862  0.017198  0.030097  0.053028  0.027231  0.050162  0.017198
494  0.054052  0.025739  0.038608  0.061774  0.036035  0.041182  0.036035
495  0.034014  0.025510  0.053855  0.068028  0.043935  0.035431  0.017007
496  0.033325  0.004347  0.033325  0.034774  0.013040  0.028978  0.007245
498  0.044574  0.026183  0.027118  0.040833  0.015274  0.016209  0.008104

           X7        X8        X9  …       X32       X33       X34  \
0    0.046994  0.024734  0.056888  …  0.028875 -0.020245 -0.008724
2    0.063224  0.058708  0.153544  … -0.029635  0.021992  0.025927
3    0.062163  0.044205  0.113275  … -0.007813 -0.004888  0.031515
4    0.080974  0.035296  0.101737  …  0.003687  0.015463 -0.016523
5    0.027266  0.047988  0.071983  …  0.020240  0.028329  0.014238
..        …         …         …  …         …         …
493  0.040129  0.038696  0.047295  … -0.006049  0.029115  0.020193
494  0.066921  0.051478  0.095234  …  0.006623  0.020531  0.007971
495  0.028345  0.035431  0.062359  … -0.010329  0.020742  0.013356
496  0.018836  0.021734  0.037672  …  0.017081  0.019761  0.029700
498  0.020261  0.024625  0.038340  …  0.024787  0.023389  0.017417

           X35       X36       X37       X38       X39       X40           X41
0    0.068993  0.021776  0.019051 -0.005004  0.044253  0.016594  no_efectores
2   -0.000578 -0.026345  0.016201  0.014279 -0.037838  0.004077  no_efectores
3    0.020548  0.010312  0.023627  0.001454  0.013942  0.004769  no_efectores
4    0.005485 -0.030424 -0.026392 -0.005144  0.040186 -0.004108  no_efectores
5    0.011691  0.019153  0.002561 -0.005732  0.012801 -0.005049  no_efectores
..        …         …         …         …         …         …
493  0.041038  0.007866  0.037797  0.019813  0.014427  0.008446  no_efectores
494  0.014550 -0.010967 -0.024104  0.045949  0.014163 -0.005730  no_efectores
495  0.028819  0.003969 -0.002085  0.007820  0.010056  0.025843  no_efectores
496  0.023251  0.028827  0.022656  0.024889  0.019810  0.021751  no_efectores
498  0.026812  0.027117  0.031449  0.016471  0.029198  0.019545  no_efectores

[408 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  408.000000  408.000000  408.000000  408.000000  408.000000  408.000000
mean     0.040631    0.014282    0.035148    0.045150    0.033686    0.035692
std      0.014587    0.010625    0.015919    0.024495    0.018979    0.011653
min      0.002157    0.000000    0.000000    0.000000    0.000000    0.002690
25%      0.032245    0.006748    0.023778    0.027017    0.019018    0.027025
50%      0.040343    0.012047    0.034524    0.041239    0.030713    0.035345
75%      0.049087    0.018183    0.045437    0.059303    0.045818    0.043650
max      0.089239    0.060266    0.089686    0.128733    0.095091    0.067617
```

```
              X6          X7          X8          X9  …         X31  \
count  408.000000  408.000000  408.000000  408.000000  …  408.000000
mean     0.015538    0.043282    0.043889    0.063541  …    0.015779
std      0.008978    0.020758    0.023319    0.025950  …    0.017680
min      0.000000    0.000000    0.000000    0.000000  …   -0.036744
25%      0.009290    0.028262    0.028565    0.045463  …    0.005327
50%      0.014459    0.040527    0.041163    0.061713  …    0.014406
75%      0.020666    0.055479    0.056837    0.078986  …    0.027482
max      0.048829    0.115110    0.128038    0.153544  …    0.071275

              X32         X33         X34         X35         X36         X37  \
count  408.000000  408.000000  408.000000  408.000000  408.000000  408.000000
mean     0.013550    0.016641    0.015362    0.014678    0.014639    0.013452
std      0.018770    0.018428    0.018090    0.017638    0.019260    0.017724
min     -0.048548   -0.052162   -0.055879   -0.035645   -0.043111   -0.039759
25%      0.002358    0.007603    0.004704    0.003567    0.003009    0.002837
50%      0.014997    0.017957    0.014581    0.015248    0.016447    0.013943
75%      0.025161    0.028108    0.026166    0.026598    0.026076    0.024181
max      0.075799    0.062275    0.077200    0.068993    0.090645    0.061812

              X38         X39         X40
count  408.000000  408.000000  408.000000
mean     0.012920    0.014616    0.013999
std      0.018315    0.018531    0.017626
min     -0.056884   -0.042405   -0.052223
25%      0.003182    0.003743    0.003772
50%      0.013406    0.014951    0.014894
75%      0.024523    0.025403    0.024909
max      0.082438    0.076856    0.067615

[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 nematoda dataset 2,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.103481  0.022996  0.080485  0.000000  0.034494  0.091983  0.011498
1    0.050742  0.026706  0.045401  0.044065  0.040060  0.038724  0.020030
2    0.056324  0.006626  0.038654  0.044176  0.024297  0.030923  0.016566
3    0.064204  0.096307  0.022225  0.041980  0.019755  0.061735  0.007408
4    0.029359  0.002669  0.021352  0.029359  0.029359  0.024021  0.008007
..        …         …         …         …         …         …         …
495  0.055831  0.088154  0.049954  0.026446  0.023508  0.067584  0.017631
496  0.014873  0.104109  0.029745  0.037182  0.029745  0.074363  0.037182
497  0.050672  0.016002  0.040004  0.066673  0.034670  0.088009  0.024002
498  0.054498  0.022707  0.068122  0.027249  0.018166  0.063580  0.018166
499  0.038326  0.009581  0.024638  0.067070  0.010950  0.027376  0.008213

           X7        X8        X9  …       X53       X54       X55  \
0    0.034494  0.022996  0.080485  … -0.034894  0.008526 -0.034448
1    0.049407  0.042730  0.085460  … -0.015195 -0.002250 -0.010915
```

```
2    0.032027  0.035340  0.048593  …  0.039579  0.011391  0.002137
3    0.014816  0.019755  0.046919  …  0.003862  0.004155  0.016359
4    0.034697  0.008007  0.061386  …  0.002139  0.038138  0.023844
..        …         …        …  …        …         …
495  0.038200  0.058769  0.017631  …  0.039650 -0.032536 -0.017288
496  0.059491  0.081800  0.007436  …  0.024056  0.011135  0.004480
497  0.093343  0.082675  0.098676  … -0.039025 -0.016285 -0.010509
498  0.040873  0.027249  0.068122  …  0.007026 -0.020591 -0.011939
499  0.031482  0.034220  0.042432  …  0.016307 -0.029520  0.008183

          X56       X57       X58       X59       X60       X61        X62
0   -0.088784 -0.053566  0.074788  0.056890 -0.048048 -0.015070  efectores
1    0.014735  0.023161  0.014366 -0.008696  0.008561  0.012954  efectores
2   -0.007744  0.009131 -0.002888  0.001325 -0.009080  0.007899  efectores
3    0.002720 -0.005722 -0.015385 -0.015163 -0.001836  0.019187  efectores
4   -0.022677 -0.033526  0.038869  0.015391  0.057382  0.030830  efectores
..        …         …         …         …         …         …
495 -0.002062 -0.002671  0.003514 -0.018121  0.014231  0.020747  efectores
496 -0.049055 -0.066884  0.040855  0.029830 -0.003999 -0.009148  efectores
497  0.003920  0.069173 -0.010804  0.043348 -0.028522 -0.000919  efectores
498  0.054909  0.030177  0.037985  0.012821  0.011404  0.062941  efectores
499  0.014787  0.046076  0.024013  0.012757  0.003884  0.025231  efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.050116    0.019832    0.038429    0.043864    0.034160    0.044166
std      0.035336    0.027259    0.027318    0.027320    0.029895    0.040410
min     -0.047912    0.000000   -0.143737   -0.095824   -0.143737   -0.047912
25%      0.028553    0.005683    0.019318    0.024794    0.016470    0.020724
50%      0.043726    0.012565    0.035051    0.042000    0.028682    0.036266
75%      0.062755    0.023803    0.051442    0.059061    0.045162    0.054901
max      0.367307    0.244872    0.175915    0.162550    0.277191    0.489743

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.018545    0.043224    0.042122    0.070153  …    0.002701
std      0.018805    0.040029    0.032341    0.052973  …    0.054562
min     -0.143737   -0.191649   -0.143737   -0.191649  …   -0.391140
25%      0.007434    0.022949    0.022131    0.037602  …   -0.012753
50%      0.014694    0.038075    0.037074    0.061370  …    0.005125
75%      0.025015    0.054709    0.055513    0.091818  …    0.022978
```

```
max      0.166963    0.612179    0.367307    0.612179  …    0.457003

              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.007167    0.004477    0.010252    0.000116    0.006059    0.004644
std      0.042221    0.067394    0.048309    0.075752    0.054984    0.058333
min     -0.217208   -0.406266   -0.175849   -1.286692   -0.880785   -0.604859
25%     -0.006239   -0.010680   -0.007764   -0.015449   -0.006911   -0.012349
50%      0.009991    0.006323    0.010796    0.003876    0.011086    0.007461
75%      0.022998    0.020319    0.026672    0.022976    0.025930    0.023108
max      0.356761    0.971634    0.432989    0.364393    0.186498    0.373829

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.009589   -0.000534    0.005839
std      0.037426    0.067091    0.042875
min     -0.233844   -1.037363   -0.348295
25%     -0.005667   -0.014902   -0.009413
50%      0.010628    0.005113    0.008889
75%      0.025994    0.021543    0.025192
max      0.265486    0.183110    0.164922

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

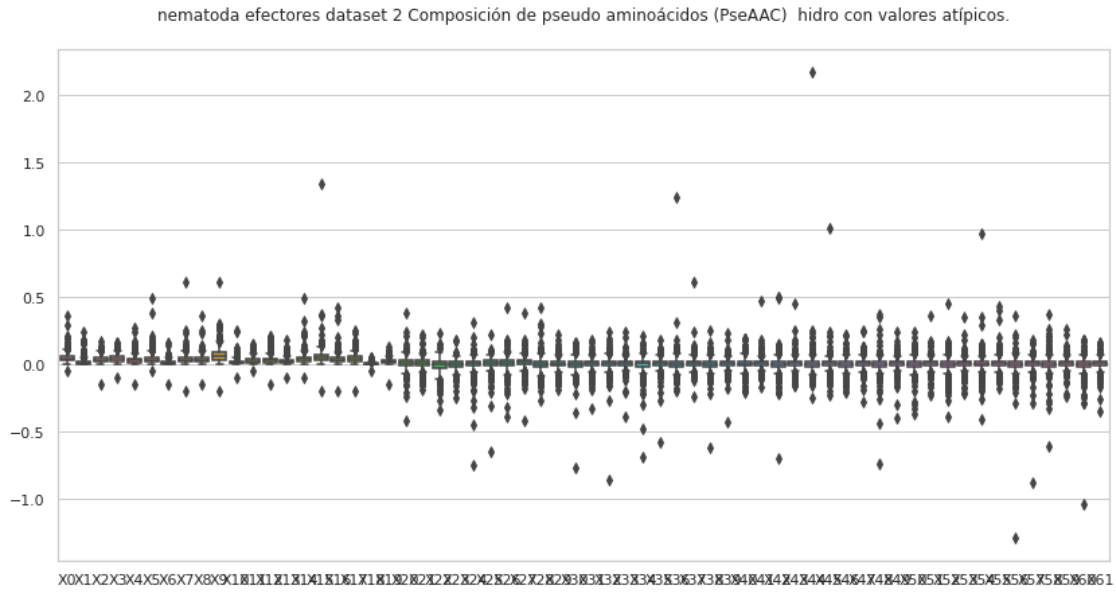
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.021753  0.003108  0.012430  0.013984  0.031076  0.035737  0.004661
1    0.025742  0.000000  0.018387  0.033097  0.011032  0.025742  0.029420
2    0.028705  0.007829  0.015657  0.023486  0.033924  0.033924  0.015657
3    0.024358  0.022618  0.046976  0.046976  0.052195  0.041756  0.038277
4    0.033027  0.010509  0.016513  0.022518  0.043535  0.009007  0.021017
..        …         …         …         …         …         …         …
495  0.030050  0.022538  0.047579  0.060100  0.038815  0.031302  0.015025
496  0.030576  0.003988  0.030576  0.031905  0.011964  0.026588  0.006647
497  0.012644  0.063220  0.021073  0.029503  0.033717  0.067435  0.004215
498  0.058963  0.034635  0.035872  0.054015  0.020204  0.021441  0.010720
499  0.004793  0.001307  0.002179  0.003486  0.026142  0.006971  0.002614

           X7        X8        X9  …       X53       X54       X55  \
0    0.029522  0.015538  0.035737  …  0.013069  0.004000  0.000376
1    0.018387  0.018387  0.036774  … -0.007126  0.009652 -0.012255
2    0.036533  0.033924  0.088723  …  0.016387  0.017635 -0.042630
```

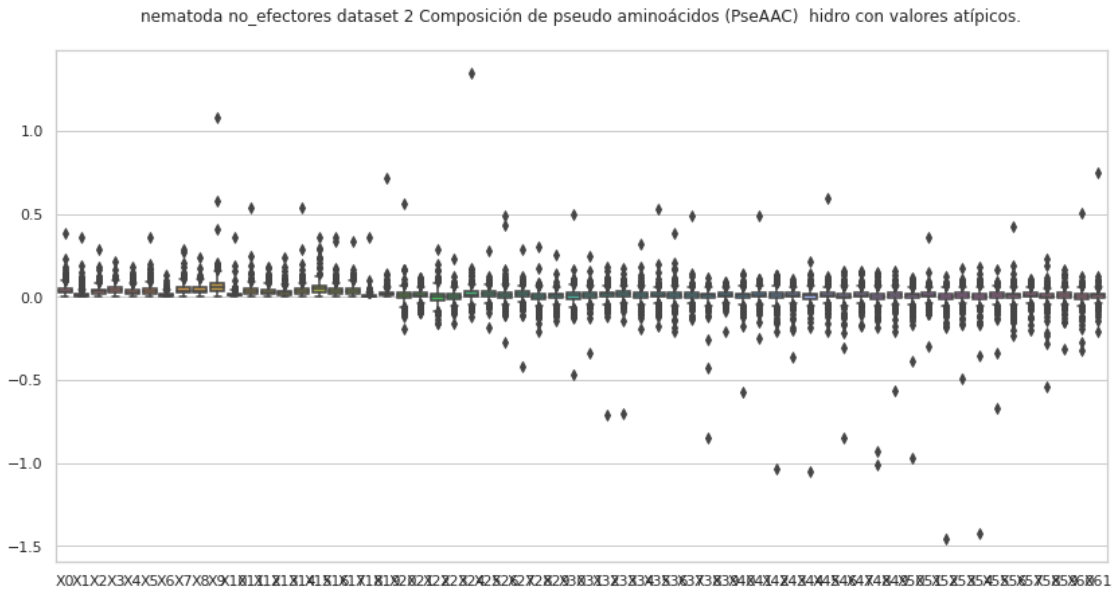
```
3    0.078293  0.055675  0.142667  … -0.011815  0.005447  0.011305
4    0.058548  0.025521  0.073560  … -0.017292 -0.051527 -0.023676
..        …         …         …  …         …         …
495  0.025042  0.031302  0.055092  … -0.025739  0.009544  0.037567
496  0.017282  0.019941  0.034564  …  0.030713  0.015605  0.025828
497  0.046361  0.067435  0.050576  …  0.082250  0.048016  0.058244
498  0.026801  0.032574  0.050716  …  0.004379  0.014322  0.035044
499  0.012200  0.005664  0.028321  …  0.011232  0.022274  0.010299

          X56       X57       X58       X59       X60       X61           X62
0    0.005793  0.004332  0.005604  0.008828  0.028278  0.011150  no_efectores
1    0.018424 -0.002264 -0.006656  0.016903 -0.028645 -0.027703  no_efectores
2    0.002267 -0.011192 -0.009157  0.005326  0.018979  0.008283  no_efectores
3    0.017126  0.012426 -0.038116 -0.030192  0.018374 -0.000764  no_efectores
4   -0.029026 -0.011045  0.008431 -0.007381  0.018356  0.009935  no_efectores
..        …         …         …         …         …         …           …
495  0.000551  0.025474  0.009938  0.021988 -0.000851  0.008559  no_efectores
496 -0.013875  0.003876  0.002111  0.026522  0.017928  0.010091  no_efectores
497 -0.011752 -0.010923 -0.030871 -0.009192  0.035082  0.000625  no_efectores
498 -0.006960  0.011018  0.001775  0.020953 -0.001905  0.007620  no_efectores
499  0.022455  0.010668  0.028035  0.015322  0.024114  0.012934  no_efectores

[500 rows x 63 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.041657    0.016777    0.034991    0.042539    0.033658    0.038528
std      0.031403    0.024773    0.024993    0.024613    0.023376    0.031163
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.023364    0.005740    0.017973    0.024864    0.017721    0.019237
50%      0.036106    0.010538    0.031931    0.039208    0.029471    0.030934
75%      0.053346    0.019753    0.046232    0.056647    0.043586    0.048314
max      0.382833    0.359877    0.287125    0.215700    0.179938    0.359877

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.016078    0.044184    0.042642    0.065503  …   -0.003412
std      0.013880    0.031779    0.027057    0.063162  …    0.073417
min      0.000000    0.000000    0.000000    0.000000  …   -1.457568
25%      0.006097    0.023310    0.023337    0.036768  …   -0.011260
50%      0.013226    0.037173    0.037735    0.055477  …    0.003880
75%      0.021750    0.058184    0.057986    0.084949  …    0.017225
max      0.134778    0.287125    0.239271    1.079631  …    0.122473
```

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.007382	-0.000337	0.007277	0.003383	0.009305	0.001055	
std	0.037160	0.074034	0.045013	0.041543	0.030831	0.044569	
min	-0.491181	-1.426855	-0.667527	-0.234964	-0.200971	-0.541885	
25%	-0.003003	-0.011306	-0.003368	-0.009172	-0.000942	-0.009281	
50%	0.012648	0.005623	0.010282	0.006712	0.010980	0.004738	
75%	0.024217	0.020084	0.024967	0.020106	0.024013	0.016636	
max	0.176061	0.181333	0.165500	0.422381	0.104177	0.233029	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.007943	0.001083	0.009294
std	0.033062	0.043670	0.043569
min	-0.310898	-0.321343	-0.206736
25%	-0.004336	-0.010168	-0.003857
50%	0.011341	0.004366	0.009669
75%	0.022483	0.018988	0.022001
max	0.162997	0.501258	0.745834

[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 nematoda dataset 2, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.103481  0.022996  0.080485  0.000000  0.034494  0.091983  0.011498
1    0.050742  0.026706  0.045401  0.044065  0.040060  0.038724  0.020030
2    0.056324  0.006626  0.038654  0.044176  0.024297  0.030923  0.016566
3    0.064204  0.096307  0.022225  0.041980  0.019755  0.061735  0.007408
4    0.029359  0.002669  0.021352  0.029359  0.029359  0.024021  0.008007
..        …         …         …         …         …         …         …
494  0.004842  0.033891  0.019367  0.043575  0.009683  0.000000  0.009683
495  0.055831  0.088154  0.049954  0.026446  0.023508  0.067584  0.017631
497  0.050672  0.016002  0.040004  0.066673  0.034670  0.088009  0.024002
498  0.054498  0.022707  0.068122  0.027249  0.018166  0.063580  0.018166
499  0.038326  0.009581  0.024638  0.067070  0.010950  0.027376  0.008213

           X7        X8        X9  …       X53       X54       X55  \
0    0.034494  0.022996  0.080485  … -0.034894  0.008526 -0.034448
1    0.049407  0.042730  0.085460  … -0.015195 -0.002250 -0.010915
2    0.032027  0.035340  0.048593  …  0.039579  0.011391  0.002137
3    0.014816  0.019755  0.046919  …  0.003862  0.004155  0.016359
4    0.034697  0.008007  0.061386  …  0.002139  0.038138  0.023844
```

```
..        …       …       …  …       …       …       …
494  0.014525  0.019367  0.024208  …  0.004474 -0.025325 -0.003596
495  0.038200  0.058769  0.017631  …  0.039650 -0.032536 -0.017288
497  0.093343  0.082675  0.098676  … -0.039025 -0.016285 -0.010509
498  0.040873  0.027249  0.068122  …  0.007026 -0.020591 -0.011939
499  0.031482  0.034220  0.042432  …  0.016307 -0.029520  0.008183

          X56       X57       X58       X59       X60       X61        X62
0   -0.088784 -0.053566  0.074788  0.056890 -0.048048 -0.015070  efectores
1    0.014735  0.023161  0.014366 -0.008696  0.008561  0.012954  efectores
2   -0.007744  0.009131 -0.002888  0.001325 -0.009080  0.007899  efectores
3    0.002720 -0.005722 -0.015385 -0.015163 -0.001836  0.019187  efectores
4   -0.022677 -0.033526  0.038869  0.015391  0.057382  0.030830  efectores
..        …       …       …       …       …       …       …
494 -0.046804  0.009257 -0.057885 -0.048284 -0.032396 -0.017536  efectores
495 -0.002062 -0.002671  0.003514 -0.018121  0.014231  0.020747  efectores
497  0.003920  0.069173 -0.010804  0.043348 -0.028522 -0.000919  efectores
498  0.054909  0.030177  0.037985  0.012821  0.011404  0.062941  efectores
499  0.014787  0.046076  0.024013  0.012757  0.003884  0.025231  efectores

[419 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 2, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  419.000000  419.000000  419.000000  419.000000  419.000000  419.000000
mean     0.044351    0.015740    0.033605    0.040471    0.028986    0.036763
std      0.024542    0.015330    0.019069    0.022183    0.017467    0.023262
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.026755    0.005412    0.018298    0.023345    0.015894    0.019227
50%      0.039854    0.011661    0.032213    0.038778    0.026729    0.031745
75%      0.058054    0.021671    0.047127    0.054728    0.038644    0.050239
max      0.134294    0.096307    0.094050    0.121358    0.097291    0.157222

                X6          X7          X8          X9  …         X52  \
count  419.000000  419.000000  419.000000  419.000000  …  419.000000
mean     0.016236    0.037960    0.037730    0.060715  …    0.004564
std      0.012095    0.021423    0.021040    0.031770  …    0.031491
min      0.000000    0.000000    0.000000    0.000000  …   -0.124669
25%      0.007239    0.022195    0.021631    0.035867  …   -0.011204
50%      0.014073    0.034686    0.034185    0.057255  …    0.004553
75%      0.022471    0.049392    0.049352    0.082265  …    0.019149
max      0.074142    0.114787    0.130434    0.176087  …    0.150913

               X53         X54         X55         X56         X57         X58  \
```

```
count  419.000000  419.000000  419.000000  419.000000  419.000000  419.000000
mean     0.009307    0.005844    0.010204    0.005351    0.010723    0.006300
std      0.027291    0.032252    0.029007    0.033631    0.026575    0.029979
min     -0.097682   -0.152945   -0.121091   -0.167120   -0.105869   -0.094522
25%     -0.003094   -0.007297   -0.003990   -0.010713   -0.002815   -0.008299
50%      0.010849    0.007927    0.012171    0.005290    0.011647    0.007892
75%      0.022681    0.019961    0.026434    0.022387    0.025876    0.020888
max      0.121500    0.149413    0.122624    0.127896    0.100044    0.130292

              X59         X60         X61
count  419.000000  419.000000  419.000000
mean     0.010622    0.004505    0.008986
std      0.023242    0.029538    0.027476
min     -0.082661   -0.135308   -0.116130
25%     -0.003495   -0.011684   -0.005795
50%      0.011356    0.006411    0.010004
75%      0.024669    0.020083    0.024577
max      0.095687    0.098019    0.095379

[8 rows x 62 columns]
```

no_efectores

Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 2, sin valores atípicos.

Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.021753  0.003108  0.012430  0.013984  0.031076  0.035737  0.004661
1    0.025742  0.000000  0.018387  0.033097  0.011032  0.025742  0.029420
2    0.028705  0.007829  0.015657  0.023486  0.033924  0.033924  0.015657
3    0.024358  0.022618  0.046976  0.046976  0.052195  0.041756  0.038277
4    0.033027  0.010509  0.016513  0.022518  0.043535  0.009007  0.021017
..        …         …         …         …         …         …         …
495  0.030050  0.022538  0.047579  0.060100  0.038815  0.031302  0.015025
496  0.030576  0.003988  0.030576  0.031905  0.011964  0.026588  0.006647
497  0.012644  0.063220  0.021073  0.029503  0.033717  0.067435  0.004215
498  0.058963  0.034635  0.035872  0.054015  0.020204  0.021441  0.010720
499  0.004793  0.001307  0.002179  0.003486  0.026142  0.006971  0.002614

           X7        X8        X9  …       X53       X54       X55  \
0    0.029522  0.015538  0.035737  …  0.013069  0.004000  0.000376
1    0.018387  0.018387  0.036774  … -0.007126  0.009652 -0.012255
2    0.036533  0.033924  0.088723  …  0.016387  0.017635 -0.042630
3    0.078293  0.055675  0.142667  … -0.011815  0.005447  0.011305
4    0.058548  0.025521  0.073560  … -0.017292 -0.051527 -0.023676
..        …         …         …  …       …         …         …
```

```
495  0.025042  0.031302  0.055092  … -0.025739  0.009544  0.037567
496  0.017282  0.019941  0.034564  …  0.030713  0.015605  0.025828
497  0.046361  0.067435  0.050576  …  0.082250  0.048016  0.058244
498  0.026801  0.032574  0.050716  …  0.004379  0.014322  0.035044
499  0.012200  0.005664  0.028321  …  0.011232  0.022274  0.010299

          X56       X57       X58       X59       X60       X61           X62
0    0.005793  0.004332  0.005604  0.008828  0.028278  0.011150  no_efectores
1    0.018424 -0.002264 -0.006656  0.016903 -0.028645 -0.027703  no_efectores
2    0.002267 -0.011192 -0.009157  0.005326  0.018979  0.008283  no_efectores
3    0.017126  0.012426 -0.038116 -0.030192  0.018374 -0.000764  no_efectores
4   -0.029026 -0.011045  0.008431 -0.007381  0.018356  0.009935  no_efectores
..        …         …         …         …         …         …           …
495  0.000551  0.025474  0.009938  0.021988 -0.000851  0.008559  no_efectores
496 -0.013875  0.003876  0.002111  0.026522  0.017928  0.010091  no_efectores
497 -0.011752 -0.010923 -0.030871 -0.009192  0.035082  0.000625  no_efectores
498 -0.006960  0.011018  0.001775  0.020953 -0.001905  0.007620  no_efectores
499  0.022455  0.010668  0.028035  0.015322  0.024114  0.012934  no_efectores

[435 rows x 63 columns]
```

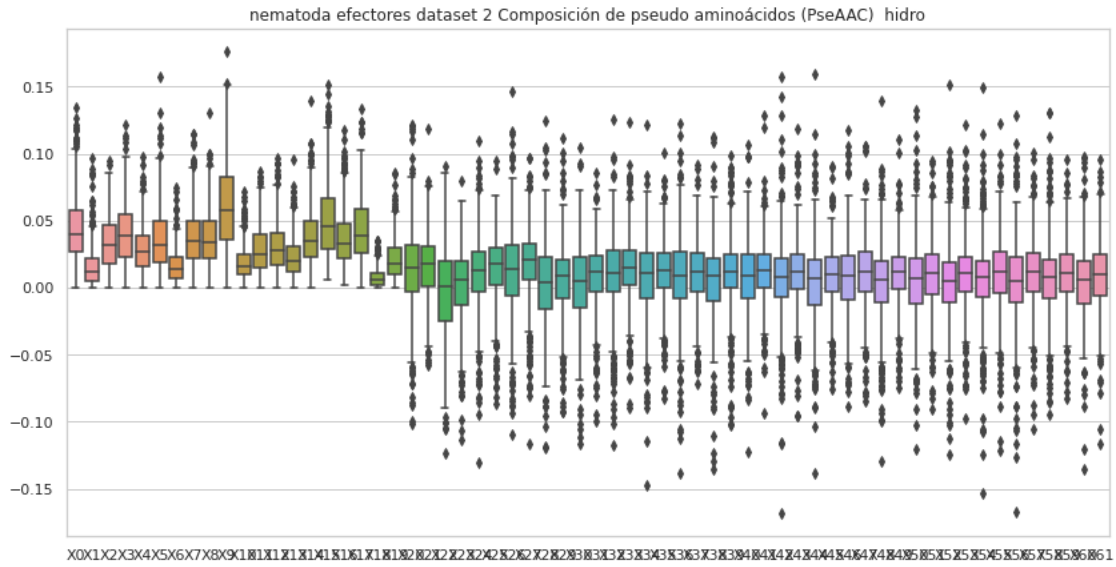
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

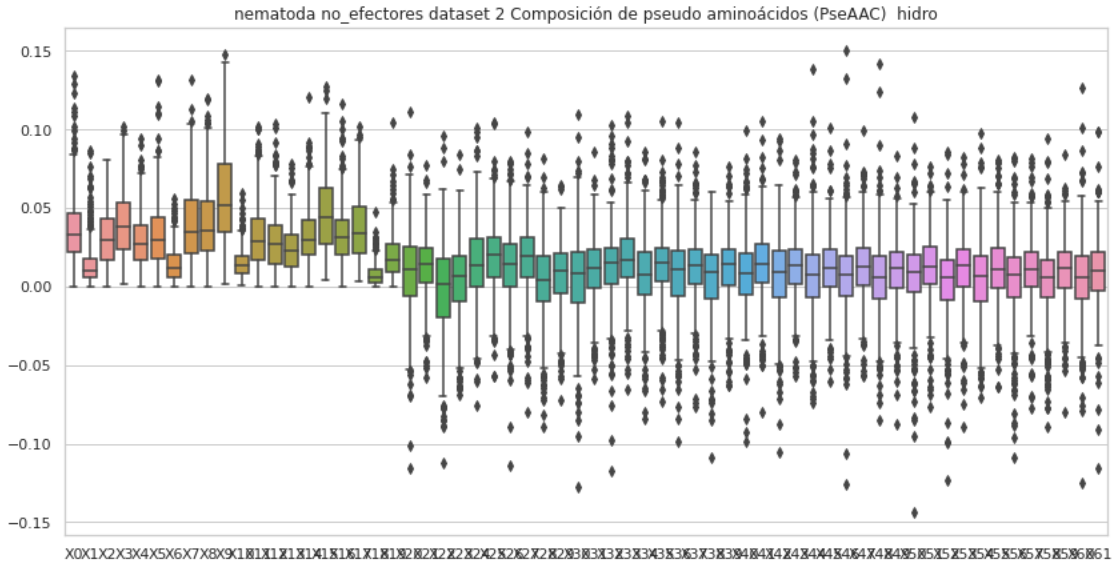
```
                X0          X1          X2          X3          X4          X5  \
count  435.000000  435.000000  435.000000  435.000000  435.000000  435.000000
mean     0.036266    0.013779    0.031046    0.039278    0.029748    0.033278
std      0.020939    0.013467    0.017669    0.020232    0.017166    0.021003
min      0.000000    0.000000    0.000000    0.001505    0.000000    0.000000
25%      0.022099    0.005451    0.016889    0.023374    0.017205    0.018114
50%      0.033252    0.009911    0.029733    0.037790    0.027401    0.029674
75%      0.047044    0.017964    0.043226    0.053791    0.039366    0.043972
max      0.134504    0.086216    0.080956    0.101735    0.093910    0.131263

                X6          X7          X8          X9  …         X52  \
count  435.000000  435.000000  435.000000  435.000000  …  435.000000
mean     0.013983    0.038447    0.039112    0.056477  …    0.003246
std      0.010061    0.022712    0.022192    0.028783  …    0.024195
min      0.000000    0.000000    0.000000    0.001866  …   -0.123273
25%      0.005743    0.021550    0.022532    0.034730  …   -0.008265
50%      0.012006    0.034354    0.035243    0.051367  …    0.005464
75%      0.019896    0.054809    0.053969    0.077986  …    0.017241
max      0.055952    0.131646    0.119677    0.147873  …    0.086067

               X53         X54         X55         X56         X57         X58  \
count  435.000000  435.000000  435.000000  435.000000  435.000000  435.000000
```

mean	0.010859	0.005049	0.011420	0.004760	0.010952	0.003632
std	0.021541	0.024015	0.022185	0.025765	0.020817	0.024265
min	-0.089275	-0.070529	-0.064028	-0.108781	-0.077749	-0.089295
25%	0.000095	-0.009359	-0.000583	-0.006566	0.000296	-0.007103
50%	0.013069	0.006331	0.010988	0.007204	0.011243	0.006081
75%	0.024114	0.019525	0.024822	0.018683	0.022225	0.016477
max	0.082250	0.097810	0.080997	0.082031	0.081183	0.094521

	X59	X60	X61
count	435.000000	435.000000	435.000000
mean	0.009465	0.005061	0.010076
std	0.021011	0.024459	0.022343
min	-0.079916	-0.125058	-0.115689
25%	-0.001306	-0.007416	-0.002248
50%	0.011602	0.005531	0.010464
75%	0.021831	0.019034	0.022100
max	0.083970	0.126513	0.098392

[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 nematoda dataset 2,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.015752  0.066245  0.155461  0.150801 -0.002553  0.121158  0.124998
1   -0.041696 -0.018513  0.027512  0.014962 -0.004789 -0.000683  0.025180
2    0.079974  0.043261 -0.012947 -0.001102  0.037922  0.046708  0.049708
3    0.192504 -0.001997 -0.027573 -0.025091  0.030296  0.042974  0.065094
4    0.033449  0.068663 -0.055571  0.011503 -0.004991  0.046743 -0.053794
..        …       …       …       …       …       …       …
495 -0.020319 -0.038151  0.051879 -0.003985  0.002508  0.005074  0.024917
496 -0.117168 -0.212870  0.106625  0.042109 -0.066790 -0.030325  0.095807
497 -0.028852  0.062159  0.038805  0.008475  0.013419  0.043000 -0.048395
498  0.021627 -0.017820  0.065489  0.107262  0.074373  0.006544 -0.016983
499  0.097856  0.020170  0.039437  0.031434  0.052583  0.005565  0.053110

           X7        X8        X9       X10       X11       X12        X13
0   -0.267626  0.009404  0.004535  0.077451 -0.325985  0.007528  efectores
1    0.009671 -0.023950  0.016948 -0.011896  0.028614 -0.038984  efectores
2    0.025106  0.023382  0.003456  0.026609  0.018083  0.071509  efectores
3    0.076930  0.069509  0.064827  0.000151  0.009315  0.014948  efectores
4   -0.002302  0.012039  0.043136 -0.089382  0.062212 -0.000510  efectores
..        …       …       …       …       …       …        …
495 -0.060971 -0.064822 -0.010204 -0.025697  0.044815 -0.031672  efectores
496 -0.046317 -0.176067  0.113132  0.179464 -0.181108 -0.082707  efectores
497 -0.012851 -0.023529  0.000618  0.029189  0.016163 -0.000878  efectores
498  0.082887  0.025198 -0.007416  0.054706  0.026473  0.109993  efectores
499  0.036943  0.055982  0.026346  0.098564  0.087351 -0.024802  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 2,
con valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
```

```
mean     0.016236    0.005568    0.010125    0.010517    0.006604    0.010995
std      0.068588    0.077462    0.086226    0.079443    0.075845    0.096205
min     -0.385441   -0.642256   -0.331947   -0.308461   -0.323472   -0.252643
25%     -0.018501   -0.026757   -0.033645   -0.029405   -0.025851   -0.036998
50%      0.016104    0.011084    0.007919    0.008475    0.005494    0.003898
75%      0.055676    0.046545    0.047135    0.045007    0.039984    0.045918
max      0.235907    0.308342    0.673707    0.594095    0.764398    0.925916

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.009254   -0.000495    0.011176    0.007563    0.003474    0.005786
std      0.090506    0.089105    0.101368    0.100428    0.100391    0.112432
min     -0.338781   -0.285119   -0.227640   -0.266591   -0.336578   -0.325985
25%     -0.030131   -0.040803   -0.033919   -0.032587   -0.035047   -0.036542
50%      0.007268    0.000632    0.005274    0.005553    0.002074    0.002046
75%      0.044064    0.036428    0.046103    0.043403    0.038256    0.042095
max      1.077387    1.221491    1.361233    1.495679    1.623606    1.743385

              X12
count  500.000000
mean     0.003462
std      0.107725
min     -0.273328
25%     -0.033378
50%      0.001048
75%      0.039578
max      1.872875
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.000167 -0.055976 -0.018631  0.120863  0.056236 -0.081618  0.071315
1    0.109958 -0.074934  0.024773  0.083190 -0.037152 -0.004095  0.101269
2    0.008312 -0.029268 -0.030858  0.047805  0.048846  0.034607 -0.083745
3    0.032531  0.028539  0.040347  0.003388 -0.008967  0.024495 -0.028408
4   -0.000250  0.083156 -0.031320 -0.000377 -0.009948  0.021810  0.029932
..        …         …         …         …         …         …         …
495  0.010204  0.031849  0.043003  0.020430  0.020456  0.012191  0.012470
496  0.070957  0.001946 -0.047749 -0.051711  0.046382  0.049022 -0.021044
497 -0.034792 -0.003683 -0.110819  0.031979 -0.053714 -0.018849  0.037101
498 -0.010908  0.026084  0.002328  0.017037  0.012033 -0.001157 -0.021899
499  0.018003  0.046014 -0.000730 -0.011415 -0.019507 -0.001468 -0.033666
```

	X7	X8	X9	X10	X11	X12	X13
0	0.015817	0.044492	-0.167400	-0.068911	0.035862	0.019047	no_efectores
1	0.123263	0.117590	0.012562	0.004407	0.023166	-0.077816	no_efectores
2	0.105819	0.032926	0.000669	-0.081337	-0.075501	-0.086625	no_efectores
3	-0.067922	-0.000237	-0.008066	-0.021767	0.009995	-0.040525	no_efectores
4	0.040939	0.001385	0.054688	0.038300	0.056941	0.002127	no_efectores
..	…	…	…	…	…	…	…
495	-0.026703	0.052832	-0.016080	-0.003467	0.016318	-0.045336	no_efectores
496	-0.016073	-0.031247	0.051904	0.043882	0.046694	-0.025329	no_efectores
497	0.064931	0.045060	0.050840	-0.072741	-0.104483	0.017203	no_efectores
498	0.052369	-0.003217	0.036388	-0.004217	0.046137	0.002267	no_efectores
499	-0.101918	0.029696	0.064029	-0.028657	0.012241	0.026371	no_efectores

[500 rows x 14 columns]

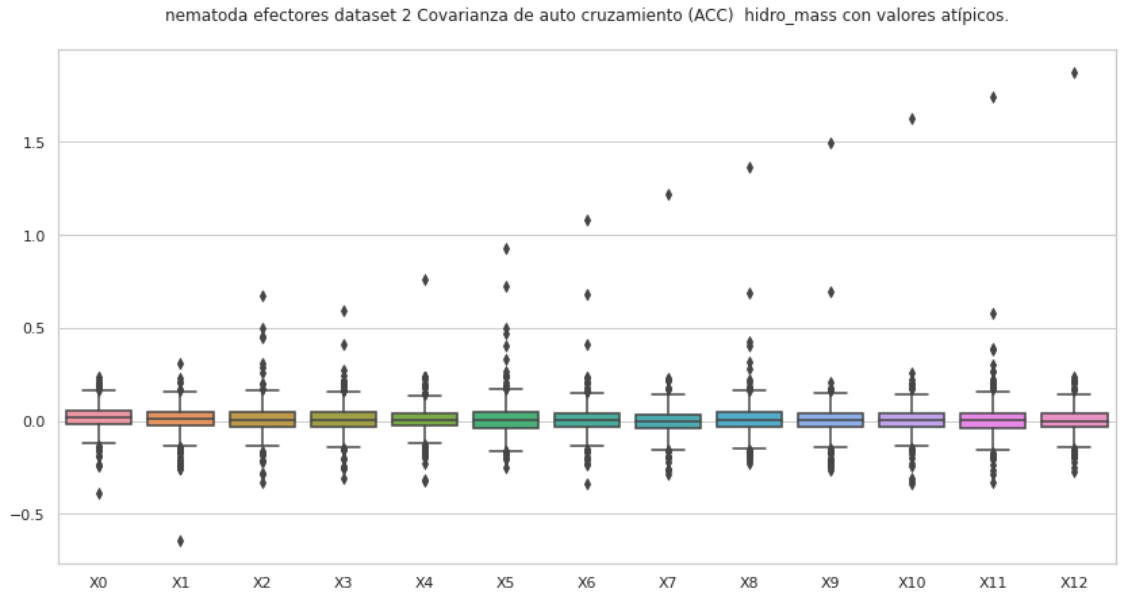
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

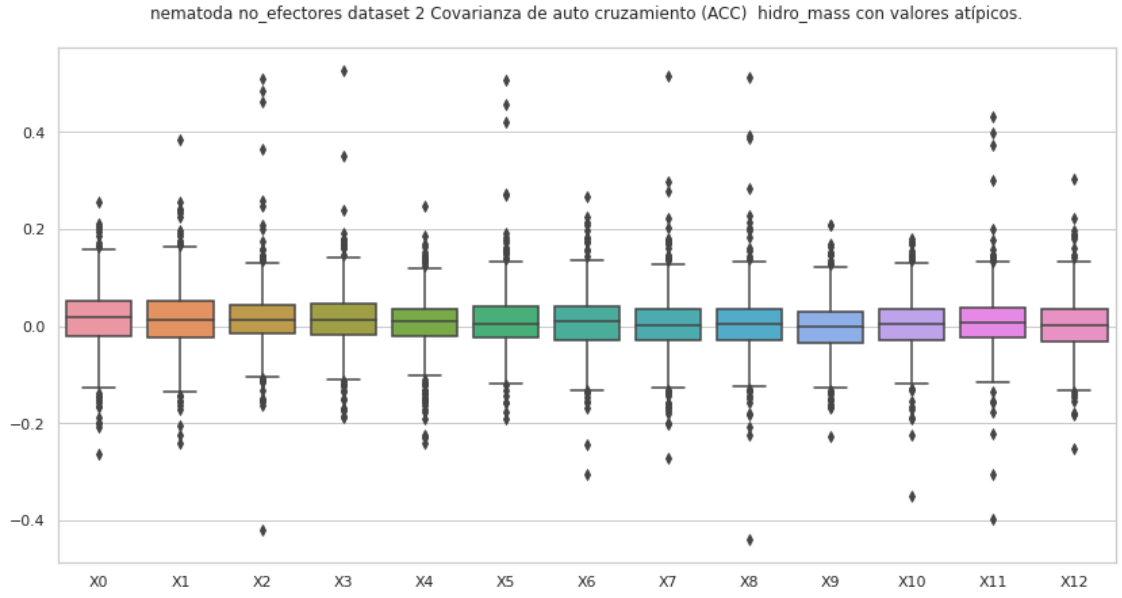
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.013983	0.016572	0.017787	0.015388	0.004931	0.011451	
std	0.064990	0.070314	0.071269	0.065266	0.061600	0.067983	
min	-0.264418	-0.241314	-0.419134	-0.187454	-0.239841	-0.190333	
25%	-0.021353	-0.022781	-0.015871	-0.017133	-0.021780	-0.023273	
50%	0.018119	0.013920	0.013482	0.014025	0.009222	0.004239	
75%	0.051248	0.051982	0.043258	0.046571	0.034419	0.040248	
max	0.255911	0.384953	0.510402	0.526831	0.246661	0.508689	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008431	0.003288	0.005450	-0.002891	0.002100	0.007134	
std	0.064338	0.069019	0.071135	0.058931	0.059401	0.068898	
min	-0.304470	-0.272421	-0.439566	-0.226175	-0.349484	-0.398331	
25%	-0.028064	-0.029777	-0.028496	-0.033645	-0.030084	-0.024533	
50%	0.011323	0.001348	0.004462	-0.000188	0.003411	0.006786	
75%	0.041003	0.036346	0.036363	0.029976	0.035406	0.038652	
max	0.266736	0.515027	0.512435	0.208619	0.180608	0.432570	

	X12
count	500.000000
mean	0.002292
std	0.060231
min	-0.252798
25%	-0.032602
50%	0.000861
75%	0.035299

```
max       0.303404
```


nematoda efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
1.5
1.0
0.5
0.0
−0.5
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
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 nematoda dataset 2,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
1	-0.041696	-0.018513	0.027512	0.014962	-0.004789	-0.000683	0.025180	
2	0.079974	0.043261	-0.012947	-0.001102	0.037922	0.046708	0.049708	
3	0.192504	-0.001997	-0.027573	-0.025091	0.030296	0.042974	0.065094	
4	0.033449	0.068663	-0.055571	0.011503	-0.004991	0.046743	-0.053794	
5	0.031698	-0.014385	-0.037674	0.062170	-0.022200	0.010384	-0.060576	
..	…	…	…	…	…	…	…	
495	-0.020319	-0.038151	0.051879	-0.003985	0.002508	0.005074	0.024917	
496	-0.117168	-0.212870	0.106625	0.042109	-0.066790	-0.030325	0.095807	
497	-0.028852	0.062159	0.038805	0.008475	0.013419	0.043000	-0.048395	
498	0.021627	-0.017820	0.065489	0.107262	0.074373	0.006544	-0.016983	
499	0.097856	0.020170	0.039437	0.031434	0.052583	0.005565	0.053110	

	X7	X8	X9	X10	X11	X12	X13
1	0.009671	-0.023950	0.016948	-0.011896	0.028614	-0.038984	efectores
2	0.025106	0.023382	0.003456	0.026609	0.018083	0.071509	efectores
3	0.076930	0.069509	0.064827	0.000151	0.009315	0.014948	efectores
4	-0.002302	0.012039	0.043136	-0.089382	0.062212	-0.000510	efectores
5	0.074264	0.005444	-0.020151	-0.021700	0.087582	0.063344	efectores
..	…	…	…	…	…	…	
495	-0.060971	-0.064822	-0.010204	-0.025697	0.044815	-0.031672	efectores
496	-0.046317	-0.176067	0.113132	0.179464	-0.181108	-0.082707	efectores
497	-0.012851	-0.023529	0.000618	0.029189	0.016163	-0.000878	efectores
498	0.082887	0.025198	-0.007416	0.054706	0.026473	0.109993	efectores
499	0.036943	0.055982	0.026346	0.098564	0.087351	-0.024802	efectores

[474 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 2,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	474.000000	474.000000	474.000000	474.000000	474.000000	474.000000	
mean	0.019094	0.009858	0.005513	0.009460	0.005308	0.003215	
std	0.060107	0.065264	0.063416	0.065039	0.059381	0.066998	
min	-0.163222	-0.214657	-0.217979	-0.198038	-0.185511	-0.252643	
25%	-0.016906	-0.023419	-0.033331	-0.027628	-0.023198	-0.036530	
50%	0.018809	0.011474	0.006895	0.008167	0.005553	0.003292	
75%	0.055397	0.046340	0.041995	0.043662	0.039742	0.042343	
max	0.219976	0.233298	0.203186	0.218379	0.221987	0.244085	

```
                X6           X7           X8           X9          X10          X11  \
count   474.000000   474.000000   474.000000   474.000000   474.000000   474.000000
mean      0.007715    -0.002074     0.005057     0.005334     0.003085    -0.000700
std       0.063188     0.064358     0.067020     0.064851     0.061364     0.066650
min      -0.224970    -0.259215    -0.212727    -0.250776    -0.199729    -0.285880
25%      -0.028792    -0.039770    -0.033107    -0.030189    -0.033346    -0.036031
50%       0.007268     0.000652     0.004456     0.007134     0.002371     0.001880
75%       0.043966     0.034510     0.044118     0.043248     0.037377     0.038839
max       0.235757     0.233122     0.279372     0.206226     0.260508     0.268689

               X12
count   474.000000
mean      0.000968
std       0.062691
min      -0.249075
25%      -0.032051
50%       0.001181
75%       0.038805
max       0.240576
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.000167 -0.055976 -0.018631  0.120863  0.056236 -0.081618  0.071315
1    0.109958 -0.074934  0.024773  0.083190 -0.037152 -0.004095  0.101269
2    0.008312 -0.029268 -0.030858  0.047805  0.048846  0.034607 -0.083745
3    0.032531  0.028539  0.040347  0.003388 -0.008967  0.024495 -0.028408
4   -0.000250  0.083156 -0.031320 -0.000377 -0.009948  0.021810  0.029932
..        …         …         …         …         …         …
495  0.010204  0.031849  0.043003  0.020430  0.020456  0.012191  0.012470
496  0.070957  0.001946 -0.047749 -0.051711  0.046382  0.049022 -0.021044
497 -0.034792 -0.003683 -0.110819  0.031979 -0.053714 -0.018849  0.037101
498 -0.010908  0.026084  0.002328  0.017037  0.012033 -0.001157 -0.021899
499  0.018003  0.046014 -0.000730 -0.011415 -0.019507 -0.001468 -0.033666

           X7        X8        X9       X10       X11       X12           X13
0    0.015817  0.044492 -0.167400 -0.068911  0.035862  0.019047  no_efectores
1    0.123263  0.117590  0.012562  0.004407  0.023166 -0.077816  no_efectores
2    0.105819  0.032926  0.000669 -0.081337 -0.075501 -0.086625  no_efectores
3   -0.067922 -0.000237 -0.008066 -0.021767  0.009995 -0.040525  no_efectores
4    0.040939  0.001385  0.054688  0.038300  0.056941  0.002127  no_efectores
..        …         …         …         …         …         …
495 -0.026703  0.052832 -0.016080 -0.003467  0.016318 -0.045336  no_efectores
```

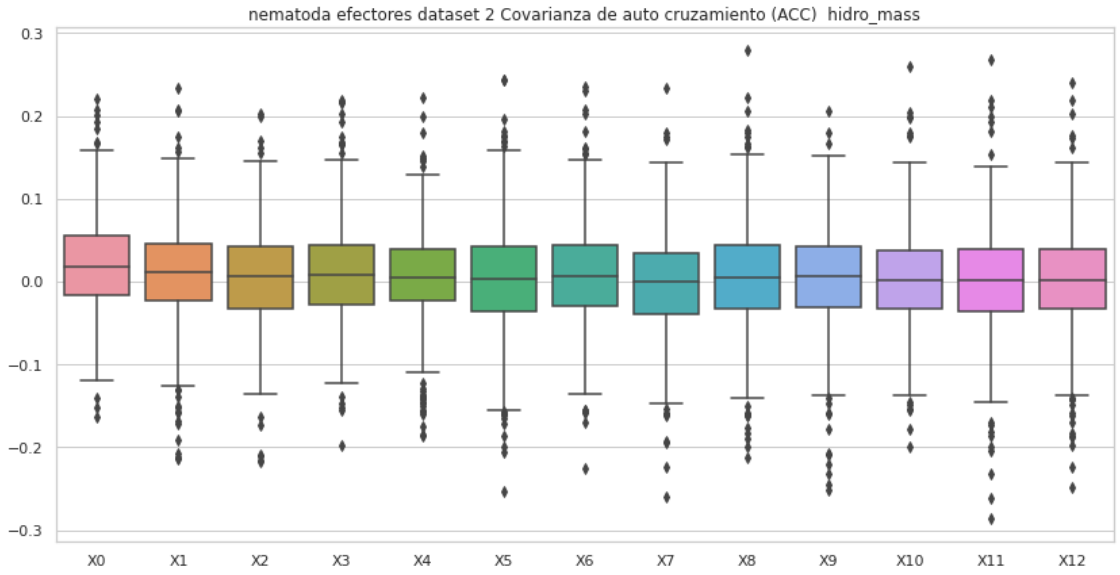
```
496 -0.016073 -0.031247  0.051904  0.043882  0.046694 -0.025329  no_efectores
497  0.064931  0.045060  0.050840 -0.072741 -0.104483  0.017203  no_efectores
498  0.052369 -0.003217  0.036388 -0.004217  0.046137  0.002267  no_efectores
499 -0.101918  0.029696  0.064029 -0.028657  0.012241  0.026371  no_efectores

[459 rows x 14 columns]
```

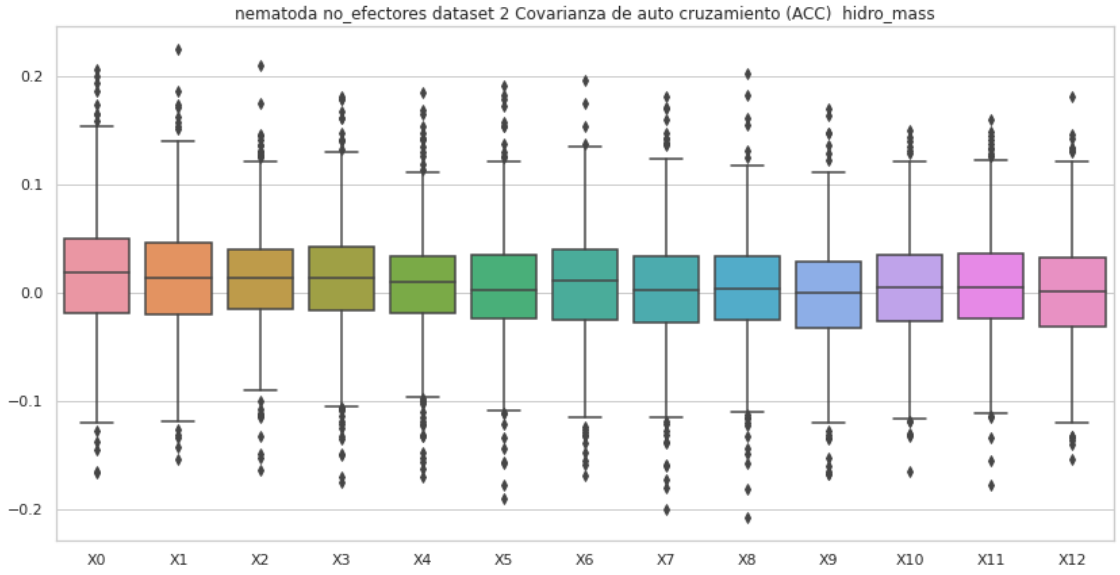
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.016613	0.014864	0.013882	0.012195	0.007501	0.005513	
std	0.057006	0.057283	0.050629	0.055095	0.053553	0.052333	
min	-0.166576	-0.154383	-0.163944	-0.175061	-0.170591	-0.190333	
25%	-0.018559	-0.020622	-0.014712	-0.016494	-0.018571	-0.023169	
50%	0.018217	0.013719	0.013437	0.013163	0.010047	0.001872	
75%	0.050523	0.046476	0.039981	0.042891	0.033968	0.035235	
max	0.206093	0.224908	0.209744	0.181726	0.185438	0.191338	

	X6	X7	X8	X9	X10	X11	\
count	459.000000	459.000000	459.000000	459.000000	459.000000	459.000000	
mean	0.007485	0.002949	0.002786	-0.003104	0.003700	0.004519	
std	0.054979	0.055875	0.052176	0.053697	0.049599	0.051917	
min	-0.169316	-0.200364	-0.207573	-0.167400	-0.164995	-0.177132	
25%	-0.024855	-0.027327	-0.024679	-0.032180	-0.025981	-0.023604	
50%	0.011234	0.001972	0.004248	-0.000242	0.004407	0.004884	
75%	0.039587	0.033496	0.033617	0.028713	0.034517	0.036184	
max	0.196800	0.181811	0.202233	0.170084	0.150091	0.159387	

	X12
count	459.000000
mean	0.000484
std	0.051149
min	-0.153864
25%	-0.031501
50%	0.000810
75%	0.031993
max	0.181441





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 nematoda dataset 2, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.015752  0.066245  0.155461  0.150801 -0.002553  0.121158  0.124998
1   -0.041696 -0.018513  0.027512  0.014962 -0.004789 -0.000683  0.025180
2    0.079974  0.043261 -0.012947 -0.001102  0.037922  0.046708  0.049708
3    0.192504 -0.001997 -0.027573 -0.025091  0.030296  0.042974  0.065094
4    0.033449  0.068663 -0.055571  0.011503 -0.004991  0.046743 -0.053794
..        …         …         …         …         …         …         …
495 -0.020319 -0.038151  0.051879 -0.003985  0.002508  0.005074  0.024917
496 -0.117168 -0.212870  0.106625  0.042109 -0.066790 -0.030325  0.095807
497 -0.028852  0.062159  0.038805  0.008475  0.013419  0.043000 -0.048395
498  0.021627 -0.017820  0.065489  0.107262  0.074373  0.006544 -0.016983
499  0.097856  0.020170  0.039437  0.031434  0.052583  0.005565  0.053110

           X7        X8        X9       X10       X11       X12        X13
0   -0.267626  0.009404  0.004535  0.077451 -0.325985  0.007528  efectores
1    0.009671 -0.023950  0.016948 -0.011896  0.028614 -0.038984  efectores
```

```
2    0.025106  0.023382  0.003456  0.026609  0.018083  0.071509  efectores
3    0.076930  0.069509  0.064827  0.000151  0.009315  0.014948  efectores
4   -0.002302  0.012039  0.043136 -0.089382  0.062212 -0.000510  efectores
..        …         …         …         …         …         …
495 -0.060971 -0.064822 -0.010204 -0.025697  0.044815 -0.031672  efectores
496 -0.046317 -0.176067  0.113132  0.179464 -0.181108 -0.082707  efectores
497 -0.012851 -0.023529  0.000618  0.029189  0.016163 -0.000878  efectores
498  0.082887  0.025198 -0.007416  0.054706  0.026473  0.109993  efectores
499  0.036943  0.055982  0.026346  0.098564  0.087351 -0.024802  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.016236    0.005568    0.010125    0.010517    0.006604    0.010995
std      0.068588    0.077462    0.086226    0.079443    0.075845    0.096205
min     -0.385441   -0.642256   -0.331947   -0.308461   -0.323472   -0.252643
25%     -0.018501   -0.026757   -0.033645   -0.029405   -0.025851   -0.036998
50%      0.016104    0.011084    0.007919    0.008475    0.005494    0.003898
75%      0.055676    0.046545    0.047135    0.045007    0.039984    0.045918
max      0.235907    0.308342    0.673707    0.594095    0.764398    0.925916

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.009254   -0.000495    0.011176    0.007563    0.003474    0.005786
std      0.090506    0.089105    0.101368    0.100428    0.100391    0.112432
min     -0.338781   -0.285119   -0.227640   -0.266591   -0.336578   -0.325985
25%     -0.030131   -0.040803   -0.033919   -0.032587   -0.035047   -0.036542
50%      0.007268    0.000632    0.005274    0.005553    0.002074    0.002046
75%      0.044064    0.036428    0.046103    0.043403    0.038256    0.042095
max      1.077387    1.221491    1.361233    1.495679    1.623606    1.743385

               X12
count  500.000000
mean     0.003462
std      0.107725
min     -0.273328
25%     -0.033378
50%      0.001048
75%      0.039578
max      1.872875
```

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.000167	-0.055976	-0.018631	0.120863	0.056236	-0.081618	0.071315	
1	0.109958	-0.074934	0.024773	0.083190	-0.037152	-0.004095	0.101269	
2	0.008312	-0.029268	-0.030858	0.047805	0.048846	0.034607	-0.083745	
3	0.032531	0.028539	0.040347	0.003388	-0.008967	0.024495	-0.028408	
4	-0.000250	0.083156	-0.031320	-0.000377	-0.009948	0.021810	0.029932	
..	…	…	…	…	…	…	…	
495	0.010204	0.031849	0.043003	0.020430	0.020456	0.012191	0.012470	
496	0.070957	0.001946	-0.047749	-0.051711	0.046382	0.049022	-0.021044	
497	-0.034792	-0.003683	-0.110819	0.031979	-0.053714	-0.018849	0.037101	
498	-0.010908	0.026084	0.002328	0.017037	0.012033	-0.001157	-0.021899	
499	0.018003	0.046014	-0.000730	-0.011415	-0.019507	-0.001468	-0.033666	

	X7	X8	X9	X10	X11	X12	X13
0	0.015817	0.044492	-0.167400	-0.068911	0.035862	0.019047	no_efectores
1	0.123263	0.117590	0.012562	0.004407	0.023166	-0.077816	no_efectores
2	0.105819	0.032926	0.000669	-0.081337	-0.075501	-0.086625	no_efectores
3	-0.067922	-0.000237	-0.008066	-0.021767	0.009995	-0.040525	no_efectores
4	0.040939	0.001385	0.054688	0.038300	0.056941	0.002127	no_efectores
..	…	…	…	…	…	…	…
495	-0.026703	0.052832	-0.016080	-0.003467	0.016318	-0.045336	no_efectores
496	-0.016073	-0.031247	0.051904	0.043882	0.046694	-0.025329	no_efectores
497	0.064931	0.045060	0.050840	-0.072741	-0.104483	0.017203	no_efectores
498	0.052369	-0.003217	0.036388	-0.004217	0.046137	0.002267	no_efectores
499	-0.101918	0.029696	0.064029	-0.028657	0.012241	0.026371	no_efectores

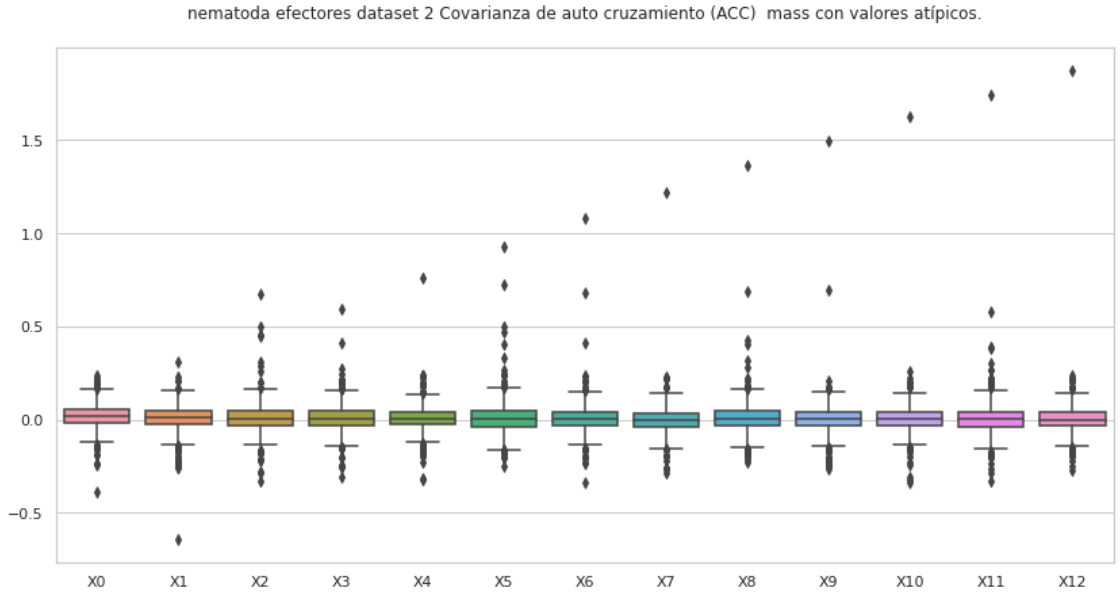
[500 rows x 14 columns]

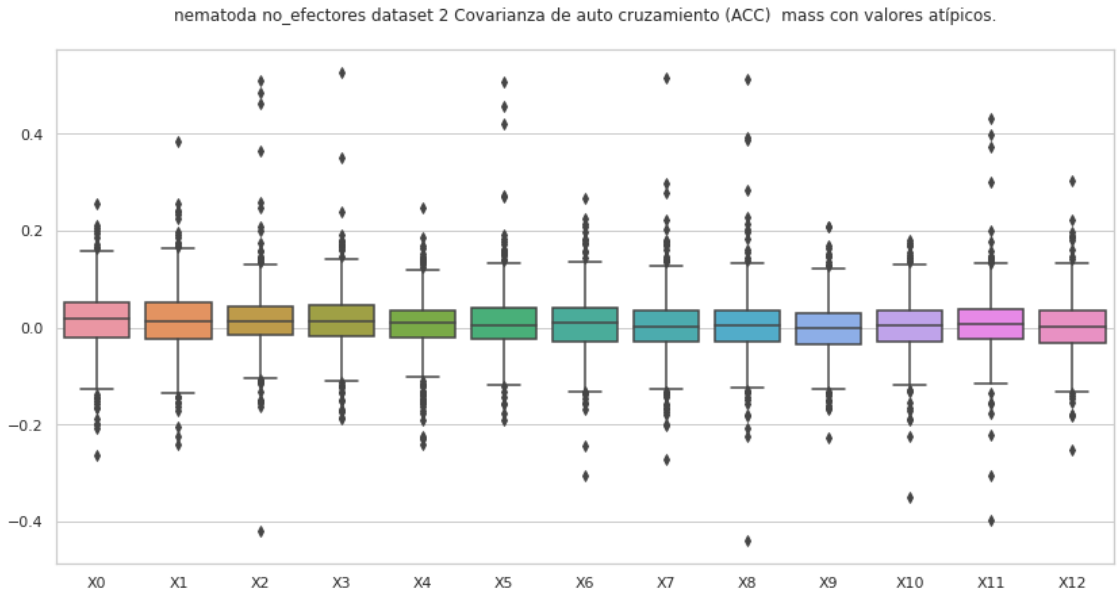
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.013983	0.016572	0.017787	0.015388	0.004931	0.011451	
std	0.064990	0.070314	0.071269	0.065266	0.061600	0.067983	
min	-0.264418	-0.241314	-0.419134	-0.187454	-0.239841	-0.190333	
25%	-0.021353	-0.022781	-0.015871	-0.017133	-0.021780	-0.023273	
50%	0.018119	0.013920	0.013482	0.014025	0.009222	0.004239	
75%	0.051248	0.051982	0.043258	0.046571	0.034419	0.040248	
max	0.255911	0.384953	0.510402	0.526831	0.246661	0.508689	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.008431	0.003288	0.005450	-0.002891	0.002100	0.007134	
std	0.064338	0.069019	0.071135	0.058931	0.059401	0.068898	
min	-0.304470	-0.272421	-0.439566	-0.226175	-0.349484	-0.398331	
25%	-0.028064	-0.029777	-0.028496	-0.033645	-0.030084	-0.024533	
50%	0.011323	0.001348	0.004462	-0.000188	0.003411	0.006786	
75%	0.041003	0.036346	0.036363	0.029976	0.035406	0.038652	
max	0.266736	0.515027	0.512435	0.208619	0.180608	0.432570	

	X12
count	500.000000
mean	0.002292
std	0.060231
min	-0.252798
25%	-0.032602
50%	0.000861
75%	0.035299
max	0.303404





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 nematoda dataset 2, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
1   -0.041696 -0.018513  0.027512  0.014962 -0.004789 -0.000683  0.025180
2    0.079974  0.043261 -0.012947 -0.001102  0.037922  0.046708  0.049708
3    0.192504 -0.001997 -0.027573 -0.025091  0.030296  0.042974  0.065094
4    0.033449  0.068663 -0.055571  0.011503 -0.004991  0.046743 -0.053794
5    0.031698 -0.014385 -0.037674  0.062170 -0.022200  0.010384 -0.060576
..        …         …         …         …         …         …         …
495 -0.020319 -0.038151  0.051879 -0.003985  0.002508  0.005074  0.024917
496 -0.117168 -0.212870  0.106625  0.042109 -0.066790 -0.030325  0.095807
497 -0.028852  0.062159  0.038805  0.008475  0.013419  0.043000 -0.048395
498  0.021627 -0.017820  0.065489  0.107262  0.074373  0.006544 -0.016983
499  0.097856  0.020170  0.039437  0.031434  0.052583  0.005565  0.053110

           X7        X8        X9       X10       X11       X12        X13
1    0.009671 -0.023950  0.016948 -0.011896  0.028614 -0.038984  efectores
2    0.025106  0.023382  0.003456  0.026609  0.018083  0.071509  efectores
3    0.076930  0.069509  0.064827  0.000151  0.009315  0.014948  efectores
4   -0.002302  0.012039  0.043136 -0.089382  0.062212 -0.000510  efectores
5    0.074264  0.005444 -0.020151 -0.021700  0.087582  0.063344  efectores
..        …         …         …         …         …         …         …
```

```
495 -0.060971 -0.064822 -0.010204 -0.025697  0.044815 -0.031672  efectores
496 -0.046317 -0.176067  0.113132  0.179464 -0.181108 -0.082707  efectores
497 -0.012851 -0.023529  0.000618  0.029189  0.016163 -0.000878  efectores
498  0.082887  0.025198 -0.007416  0.054706  0.026473  0.109993  efectores
499  0.036943  0.055982  0.026346  0.098564  0.087351 -0.024802  efectores

[474 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5
count	474.000000	474.000000	474.000000	474.000000	474.000000	474.000000
mean	0.019094	0.009858	0.005513	0.009460	0.005308	0.003215
std	0.060107	0.065264	0.063416	0.065039	0.059381	0.066998
min	-0.163222	-0.214657	-0.217979	-0.198038	-0.185511	-0.252643
25%	-0.016906	-0.023419	-0.033331	-0.027628	-0.023198	-0.036530
50%	0.018809	0.011474	0.006895	0.008167	0.005553	0.003292
75%	0.055397	0.046340	0.041995	0.043662	0.039742	0.042343
max	0.219976	0.233298	0.203186	0.218379	0.221987	0.244085

	X6	X7	X8	X9	X10	X11
count	474.000000	474.000000	474.000000	474.000000	474.000000	474.000000
mean	0.007715	-0.002074	0.005057	0.005334	0.003085	-0.000700
std	0.063188	0.064358	0.067020	0.064851	0.061364	0.066650
min	-0.224970	-0.259215	-0.212727	-0.250776	-0.199729	-0.285880
25%	-0.028792	-0.039770	-0.033107	-0.030189	-0.033346	-0.036031
50%	0.007268	0.000652	0.004456	0.007134	0.002371	0.001880
75%	0.043966	0.034510	0.044118	0.043248	0.037377	0.038839
max	0.235757	0.233122	0.279372	0.206226	0.260508	0.268689

	X12
count	474.000000
mean	0.000968
std	0.062691
min	-0.249075
25%	-0.032051
50%	0.001181
75%	0.038805
max	0.240576

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

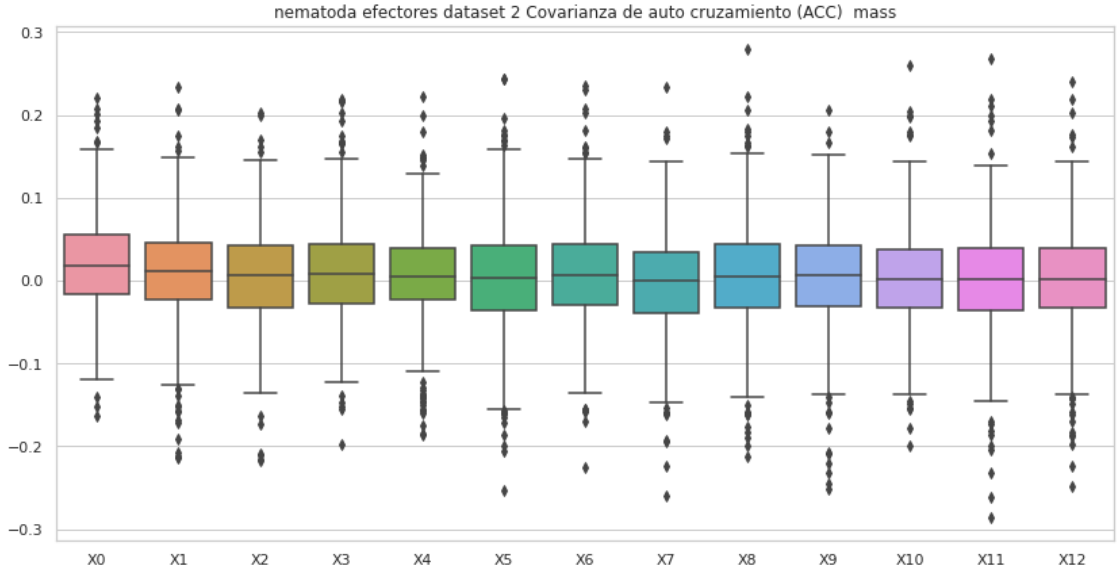
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.000167 -0.055976 -0.018631  0.120863  0.056236 -0.081618  0.071315
1    0.109958 -0.074934  0.024773  0.083190 -0.037152 -0.004095  0.101269
2    0.008312 -0.029268 -0.030858  0.047805  0.048846  0.034607 -0.083745
3    0.032531  0.028539  0.040347  0.003388 -0.008967  0.024495 -0.028408
4   -0.000250  0.083156 -0.031320 -0.000377 -0.009948  0.021810  0.029932
..        …         …         …         …         …         …         …
495  0.010204  0.031849  0.043003  0.020430  0.020456  0.012191  0.012470
496  0.070957  0.001946 -0.047749 -0.051711  0.046382  0.049022 -0.021044
497 -0.034792 -0.003683 -0.110819  0.031979 -0.053714 -0.018849  0.037101
498 -0.010908  0.026084  0.002328  0.017037  0.012033 -0.001157 -0.021899
499  0.018003  0.046014 -0.000730 -0.011415 -0.019507 -0.001468 -0.033666

           X7        X8        X9       X10       X11       X12           X13
0    0.015817  0.044492 -0.167400 -0.068911  0.035862  0.019047  no_efectores
1    0.123263  0.117590  0.012562  0.004407  0.023166 -0.077816  no_efectores
2    0.105819  0.032926  0.000669 -0.081337 -0.075501 -0.086625  no_efectores
3   -0.067922 -0.000237 -0.008066 -0.021767  0.009995 -0.040525  no_efectores
4    0.040939  0.001385  0.054688  0.038300  0.056941  0.002127  no_efectores
..        …         …         …         …         …         …           …
495 -0.026703  0.052832 -0.016080 -0.003467  0.016318 -0.045336  no_efectores
496 -0.016073 -0.031247  0.051904  0.043882  0.046694 -0.025329  no_efectores
497  0.064931  0.045060  0.050840 -0.072741 -0.104483  0.017203  no_efectores
498  0.052369 -0.003217  0.036388 -0.004217  0.046137  0.002267  no_efectores
499 -0.101918  0.029696  0.064029 -0.028657  0.012241  0.026371  no_efectores

[459 rows x 14 columns]
```

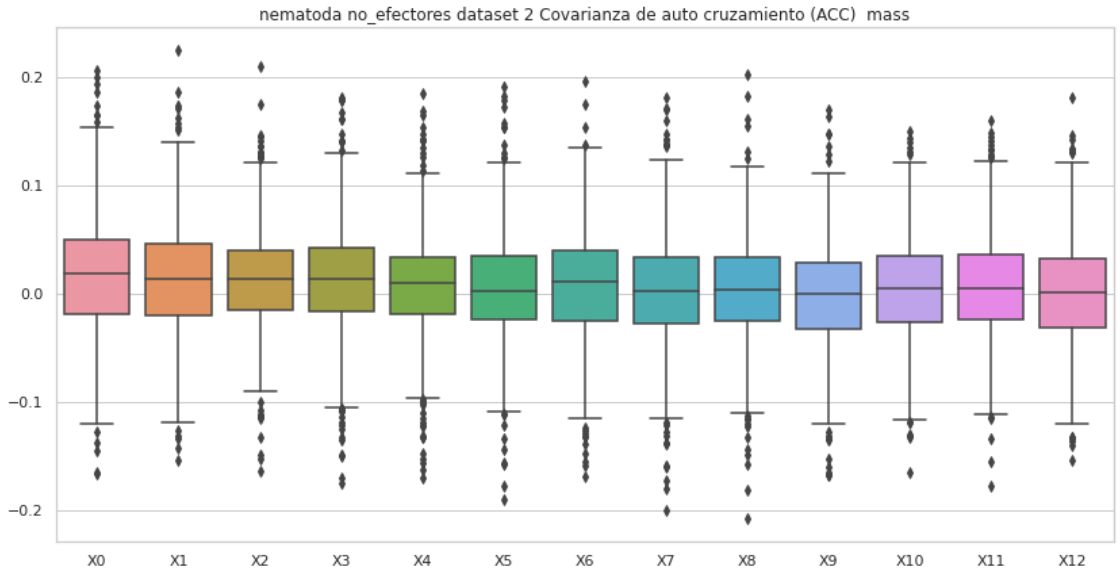
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.016613    0.014864    0.013882    0.012195    0.007501    0.005513
std      0.057006    0.057283    0.050629    0.055095    0.053553    0.052333
min     -0.166576   -0.154383   -0.163944   -0.175061   -0.170591   -0.190333
25%     -0.018559   -0.020622   -0.014712   -0.016494   -0.018571   -0.023169
50%      0.018217    0.013719    0.013437    0.013163    0.010047    0.001872
75%      0.050523    0.046476    0.039981    0.042891    0.033968    0.035235
max      0.206093    0.224908    0.209744    0.181726    0.185438    0.191338

                X6          X7          X8          X9         X10         X11  \
count  459.000000  459.000000  459.000000  459.000000  459.000000  459.000000
mean     0.007485    0.002949    0.002786   -0.003104    0.003700    0.004519
std      0.054979    0.055875    0.052176    0.053697    0.049599    0.051917
```

min	-0.169316	-0.200364	-0.207573	-0.167400	-0.164995	-0.177132
25%	-0.024855	-0.027327	-0.024679	-0.032180	-0.025981	-0.023604
50%	0.011234	0.001972	0.004248	-0.000242	0.004407	0.004884
75%	0.039587	0.033496	0.033617	0.028713	0.034517	0.036184
max	0.196800	0.181811	0.202233	0.170084	0.150091	0.159387

	X12
count	459.000000
mean	0.000484
std	0.051149
min	-0.153864
25%	-0.031501
50%	0.000810
75%	0.031993
max	0.181441





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 nematoda dataset 2, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.006187 -0.125155  0.033839  0.018773 -0.071007 -0.181102  0.069753
1   -0.078093 -0.030251  0.022794  0.014011 -0.029541 -0.075016 -0.039385
2    0.030181 -0.057173  0.015594  0.003967 -0.013088  0.036819  0.111050
3   -0.027193 -0.043489  0.012030 -0.005834  0.032613  0.029432 -0.001346
4    0.042989 -0.000286  0.082290  0.000165 -0.038623  0.114976  0.068753
..        …       …       …       …       …       …       …
495  0.014909 -0.031728 -0.052242  0.017597 -0.059403  0.063978  0.085468
496 -0.077790  0.017915 -0.036721 -0.049376  0.151838 -0.120748 -0.019889
497 -0.045581 -0.029449 -0.007293 -0.080780 -0.061067  0.017859  0.048828
498 -0.009005  0.009821 -0.101282 -0.032822  0.036063 -0.087909 -0.011789
499 -0.101038 -0.024428 -0.059178  0.031380 -0.064509 -0.022418  0.004795

           X7        X8        X9       X10       X11       X12        X13
0   -0.028109  0.032399 -0.151048 -0.072942  0.027194  0.099343  efectores
1   -0.002494  0.031486 -0.039131  0.065615  0.034885 -0.018575  efectores
2    0.067301  0.040717  0.010112 -0.014032  0.042724 -0.020254  efectores
3   -0.011335  0.026719  0.041951  0.030536  0.004215  0.005172  efectores
4   -0.129282  0.064176  0.196030  0.015860 -0.004436  0.089242  efectores
..        …       …       …       …       …       …        …
495  0.042238 -0.005962 -0.020625  0.020028 -0.072402  0.019920  efectores
496  0.115157 -0.010666 -0.071201 -0.048334 -0.159851  0.046099  efectores
497 -0.010938 -0.094156  0.000113 -0.104564  0.033911  0.053827  efectores
498  0.022520  0.143028  0.030743 -0.006784  0.024540 -0.037666  efectores
499 -0.048827 -0.023711 -0.060903  0.159221 -0.021331 -0.003179  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 2, con
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
```

```
mean     0.019770   -0.016509    0.017283    0.031563   -0.008870   -0.000936
std      0.087812    0.086211    0.085293    0.090382    0.087202    0.085015
min     -0.285560   -0.332039   -0.296429   -0.354681   -0.295669   -0.431173
25%     -0.032054   -0.072575   -0.033751   -0.024237   -0.059415   -0.047293
50%      0.015506   -0.015239    0.019544    0.033638   -0.004868   -0.000677
75%      0.073501    0.033100    0.072263    0.082149    0.040489    0.049440
max      0.349453    0.351250    0.334136    0.344887    0.312213    0.300667

               X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.016146    0.010305    0.006038    0.008456    0.012750    0.006169
std      0.082519    0.078123    0.083516    0.079866    0.075439    0.085118
min     -0.405379   -0.313475   -0.295186   -0.341541   -0.271427   -0.268242
25%     -0.034167   -0.033060   -0.046017   -0.032337   -0.030092   -0.039545
50%      0.017532    0.007727   -0.000187    0.012635    0.011316    0.008667
75%      0.063298    0.058399    0.052863    0.049870    0.053744    0.048589
max      0.313628    0.261067    0.286526    0.297441    0.314624    0.425147

              X12
count  500.000000
mean    -0.000746
std      0.081152
min     -0.312595
25%     -0.046324
50%     -0.000118
75%      0.045873
max      0.407685
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

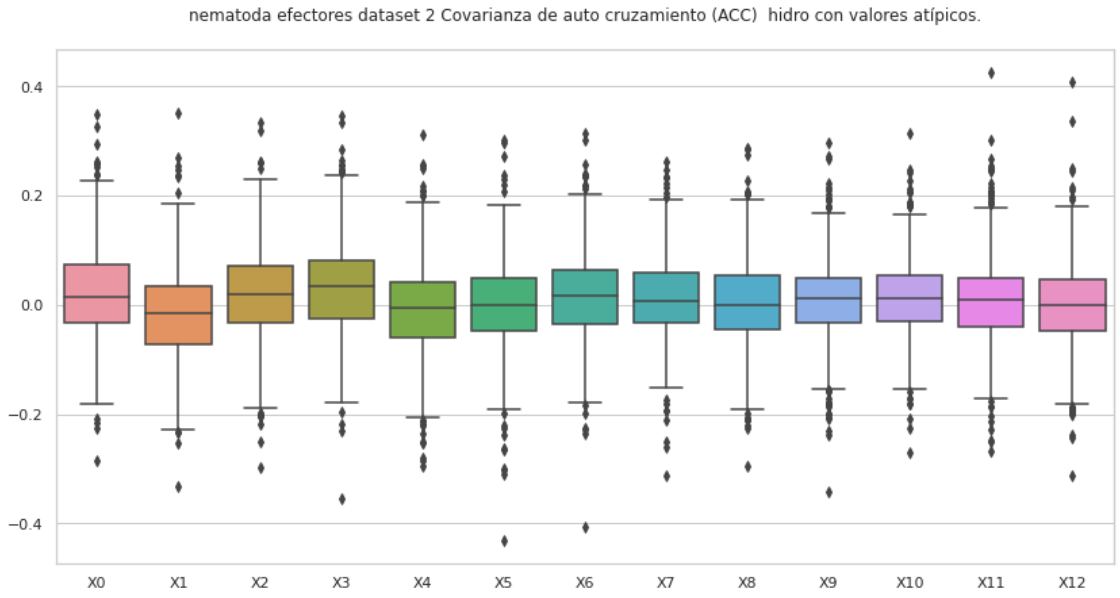
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.067759  0.102590  0.209716 -0.031470  0.170139  0.063158  0.075972
1    0.176826  0.003693  0.027015  0.043498 -0.166570  0.022397  0.010601
2    0.075383  0.024151 -0.007323 -0.144325 -0.037006  0.022531 -0.025016
3   -0.100044 -0.002005  0.062260 -0.028153  0.041534 -0.017321 -0.007624
4    0.051857  0.052449  0.029277  0.058180  0.004323  0.079580  0.045657
..        …         …         …         …         …         …         …
495  0.007260  0.007043  0.022844  0.043288  0.045162 -0.035897  0.059673
496  0.029313  0.016775 -0.076323  0.096843  0.025993  0.078836 -0.023691
497  0.041236 -0.040215 -0.051956  0.002814 -0.123644 -0.040088  0.050274
498 -0.038293 -0.036308  0.041323  0.110575 -0.023436  0.058752  0.029255
499  0.046979 -0.006354  0.019497 -0.020822 -0.012787  0.018020 -0.025322
```

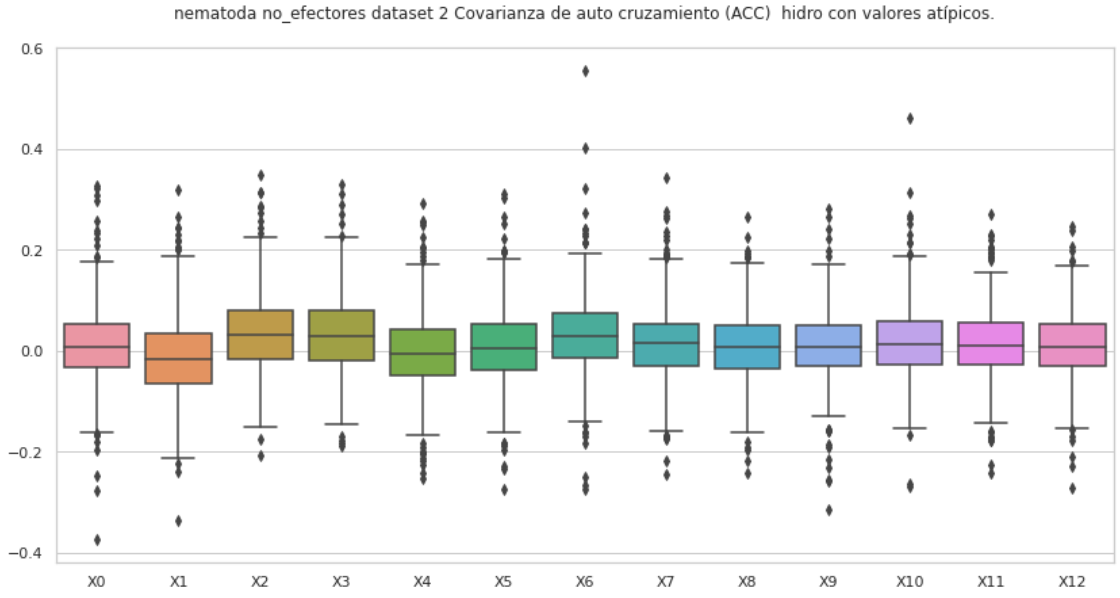
```
           X7        X8        X9       X10       X11       X12           X13
0    0.127825  0.036293  0.033626  0.051353  0.074962  0.060052  no_efectores
1   -0.112142  0.064831  0.108393 -0.088807 -0.008406  0.197020  no_efectores
2   -0.041732 -0.008749  0.017615 -0.023953 -0.137080 -0.029936  no_efectores
3    0.024910 -0.045938  0.052320 -0.022344 -0.035837  0.034336  no_efectores
4    0.046194 -0.021556  0.008846  0.051189  0.021477  0.018822  no_efectores
..        …        …         …         …         …         …           …
495 -0.011730  0.082037  0.014087  0.037905  0.030883 -0.014374  no_efectores
496  0.120516  0.056600  0.046585  0.035034  0.025066  0.004344  no_efectores
497  0.155655  0.172960 -0.035594 -0.030713 -0.059818 -0.066543  no_efectores
498  0.035391 -0.048375  0.067678  0.009785  0.016804 -0.002620  no_efectores
499  0.019581 -0.001868  0.016257  0.058826  0.001556 -0.046994  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.012506   -0.012523    0.035108    0.032629   -0.003319    0.008389
std      0.079878    0.084188    0.078127    0.074232    0.076760    0.076695
min     -0.373576   -0.336981   -0.207617   -0.187691   -0.252038   -0.273091
25%     -0.033353   -0.065610   -0.016211   -0.017924   -0.047873   -0.038413
50%      0.008598   -0.016888    0.032054    0.030411   -0.005132    0.004929
75%      0.053030    0.035809    0.080913    0.079458    0.041599    0.052174
max      0.326746    0.319417    0.347899    0.329998    0.292827    0.311377

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.030941    0.014522    0.007409    0.009271    0.019263    0.012210
std      0.079570    0.077230    0.068910    0.069170    0.075094    0.070055
min     -0.273899   -0.245900   -0.243055   -0.314797   -0.269470   -0.240721
25%     -0.012727   -0.031179   -0.035282   -0.029338   -0.026527   -0.027580
50%      0.028378    0.014607    0.006755    0.007262    0.013665    0.011122
75%      0.074177    0.053903    0.050976    0.051390    0.059790    0.054835
max      0.555077    0.342457    0.264565    0.280166    0.461789    0.271987

               X12
count  500.000000
mean     0.010492
std      0.069076
min     -0.270572
25%     -0.030197
50%      0.006900
75%      0.051934
```

```
max       0.246517
```


nematoda efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
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 nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.006187	-0.125155	0.033839	0.018773	-0.071007	-0.181102	0.069753	
1	-0.078093	-0.030251	0.022794	0.014011	-0.029541	-0.075016	-0.039385	
2	0.030181	-0.057173	0.015594	0.003967	-0.013088	0.036819	0.111050	
3	-0.027193	-0.043489	0.012030	-0.005834	0.032613	0.029432	-0.001346	
4	0.042989	-0.000286	0.082290	0.000165	-0.038623	0.114976	0.068753	
..	…	…	…	…	…	…	…	
495	0.014909	-0.031728	-0.052242	0.017597	-0.059403	0.063978	0.085468	
496	-0.077790	0.017915	-0.036721	-0.049376	0.151838	-0.120748	-0.019889	
497	-0.045581	-0.029449	-0.007293	-0.080780	-0.061067	0.017859	0.048828	
498	-0.009005	0.009821	-0.101282	-0.032822	0.036063	-0.087909	-0.011789	
499	-0.101038	-0.024428	-0.059178	0.031380	-0.064509	-0.022418	0.004795	

	X7	X8	X9	X10	X11	X12	X13
0	-0.028109	0.032399	-0.151048	-0.072942	0.027194	0.099343	efectores
1	-0.002494	0.031486	-0.039131	0.065615	0.034885	-0.018575	efectores
2	0.067301	0.040717	0.010112	-0.014032	0.042724	-0.020254	efectores
3	-0.011335	0.026719	0.041951	0.030536	0.004215	0.005172	efectores
4	-0.129282	0.064176	0.196030	0.015860	-0.004436	0.089242	efectores
..	…	…	…	…	…	…	
495	0.042238	-0.005962	-0.020625	0.020028	-0.072402	0.019920	efectores
496	0.115157	-0.010666	-0.071201	-0.048334	-0.159851	0.046099	efectores
497	-0.010938	-0.094156	0.000113	-0.104564	0.033911	0.053827	efectores
498	0.022520	0.143028	0.030743	-0.006784	0.024540	-0.037666	efectores
499	-0.048827	-0.023711	-0.060903	0.159221	-0.021331	-0.003179	efectores

[454 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	454.000000	454.000000	454.000000	454.000000	454.000000	454.000000	
mean	0.019201	-0.020397	0.018513	0.030427	-0.007561	0.002109	
std	0.080189	0.078459	0.076203	0.081497	0.079129	0.070720	
min	-0.216058	-0.252134	-0.204309	-0.230174	-0.252873	-0.225903	
25%	-0.028590	-0.072720	-0.031675	-0.018733	-0.056114	-0.039727	
50%	0.014705	-0.019676	0.019760	0.032504	-0.004868	0.000673	
75%	0.071680	0.027165	0.070405	0.076812	0.040535	0.048278	
max	0.262133	0.185876	0.260263	0.284520	0.249483	0.230289	

```
                X6           X7           X8           X9          X10          X11  \
count   454.000000   454.000000   454.000000   454.000000   454.000000   454.000000
mean      0.016722     0.013400     0.006945     0.009841     0.010526     0.004530
std       0.072267     0.069954     0.074948     0.069028     0.067459     0.076556
min      -0.226498    -0.209967    -0.220769    -0.209078    -0.209484    -0.229291
25%      -0.029778    -0.028962    -0.041993    -0.027282    -0.029198    -0.038699
50%       0.019032     0.010476    -0.000187     0.012765     0.011203     0.008369
75%       0.062791     0.058176     0.052631     0.049245     0.050111     0.047724
max       0.237519     0.235201     0.227553     0.215800     0.227265     0.247112

               X12
count   454.000000
mean      0.000051
std       0.069696
min      -0.238930
25%      -0.042408
50%       0.001548
75%       0.045254
max       0.215602
```

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.067759  0.102590  0.209716 -0.031470  0.170139  0.063158  0.075972
1    0.176826  0.003693  0.027015  0.043498 -0.166570  0.022397  0.010601
2    0.075383  0.024151 -0.007323 -0.144325 -0.037006  0.022531 -0.025016
3   -0.100044 -0.002005  0.062260 -0.028153  0.041534 -0.017321 -0.007624
4    0.051857  0.052449  0.029277  0.058180  0.004323  0.079580  0.045657
..        …         …         …         …         …         …         …
495  0.007260  0.007043  0.022844  0.043288  0.045162 -0.035897  0.059673
496  0.029313  0.016775 -0.076323  0.096843  0.025993  0.078836 -0.023691
497  0.041236 -0.040215 -0.051956  0.002814 -0.123644 -0.040088  0.050274
498 -0.038293 -0.036308  0.041323  0.110575 -0.023436  0.058752  0.029255
499  0.046979 -0.006354  0.019497 -0.020822 -0.012787  0.018020 -0.025322

           X7        X8        X9       X10       X11       X12           X13
0    0.127825  0.036293  0.033626  0.051353  0.074962  0.060052  no_efectores
1   -0.112142  0.064831  0.108393 -0.088807 -0.008406  0.197020  no_efectores
2   -0.041732 -0.008749  0.017615 -0.023953 -0.137080 -0.029936  no_efectores
3    0.024910 -0.045938  0.052320 -0.022344 -0.035837  0.034336  no_efectores
4    0.046194 -0.021556  0.008846  0.051189  0.021477  0.018822  no_efectores
..        …         …         …         …         …         …           …
495 -0.011730  0.082037  0.014087  0.037905  0.030883 -0.014374  no_efectores
```

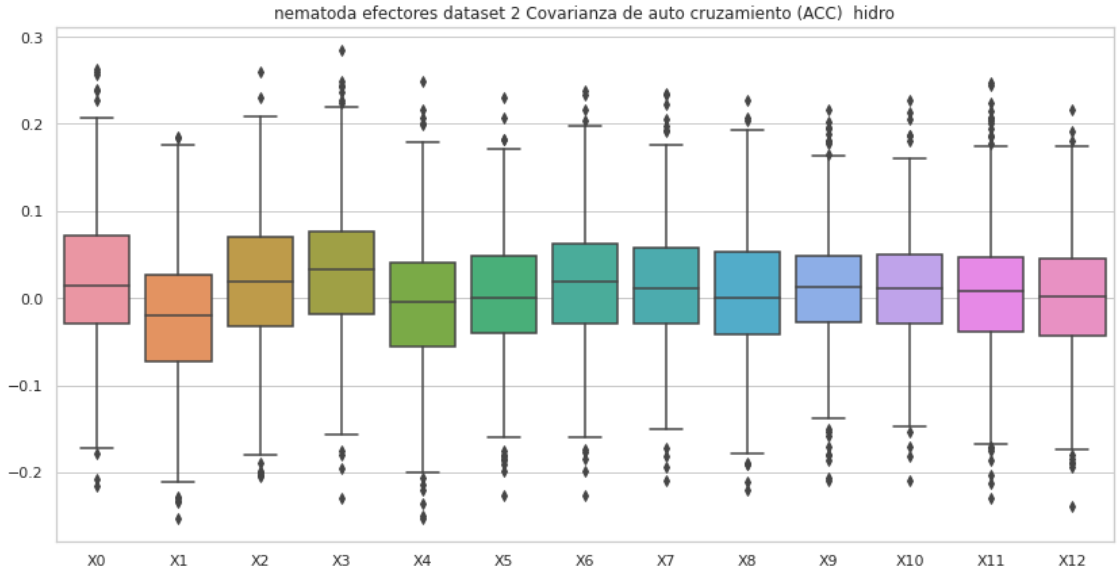
```
496  0.120516  0.056600  0.046585  0.035034  0.025066  0.004344  no_efectores
497  0.155655  0.172960 -0.035594 -0.030713 -0.059818 -0.066543  no_efectores
498  0.035391 -0.048375  0.067678  0.009785  0.016804 -0.002620  no_efectores
499  0.019581 -0.001868  0.016257  0.058826  0.001556 -0.046994  no_efectores

[462 rows x 14 columns]
```

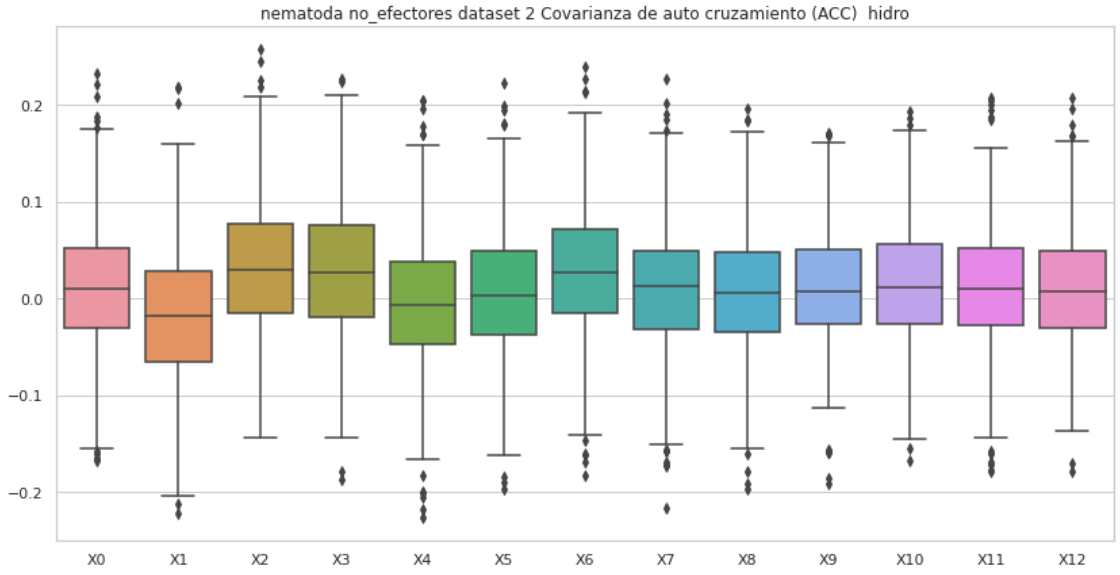
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.013454	-0.016151	0.032230	0.028730	-0.004780	0.005909	
std	0.068440	0.074555	0.069524	0.067403	0.069287	0.067373	
min	-0.167577	-0.222113	-0.143874	-0.187691	-0.226420	-0.196430	
25%	-0.030709	-0.065471	-0.015028	-0.019097	-0.046619	-0.037292	
50%	0.010139	-0.017771	0.029221	0.026368	-0.006677	0.003212	
75%	0.052094	0.028122	0.076779	0.075776	0.037984	0.049033	
max	0.232471	0.219311	0.257940	0.227368	0.204966	0.223516	

	X6	X7	X8	X9	X10	X11	\
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	
mean	0.027502	0.009810	0.005932	0.010059	0.014279	0.010548	
std	0.066633	0.068233	0.063939	0.058467	0.062898	0.063114	
min	-0.182946	-0.216531	-0.197122	-0.190994	-0.166991	-0.178492	
25%	-0.015141	-0.032125	-0.035123	-0.026014	-0.026739	-0.027106	
50%	0.027464	0.012565	0.006217	0.007869	0.011353	0.010346	
75%	0.071140	0.049519	0.048219	0.050886	0.055907	0.052014	
max	0.239965	0.227822	0.196927	0.171235	0.193343	0.207202	

	X12
count	462.000000
mean	0.009849
std	0.061584
min	-0.178554
25%	-0.029861
50%	0.006786
75%	0.049132
max	0.207172


nematoda efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12