ds2_nematoda_limpieza_de_datos

February 1, 2021

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="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  \
0    5.758   4.848  1.818   7.879   0.909   4.848  3.030  11.515   3.030
1    6.599   5.584  4.061   5.076   3.046   5.076  7.107   4.569   1.523
2    7.258   4.839  5.645  10.484   2.419   4.032  2.419   1.613   2.419
3    9.143   5.143  4.000   3.429   0.571  11.429  2.286   5.143  12.571
4    2.457   5.160  5.405   6.634  12.285   5.651  3.194   8.845   3.194
..     …       …      …       …       …       …      …       …       …
995  5.115   8.951  3.581   2.813   3.069   2.558  2.046   4.092   1.023
996  7.042   2.817  4.225   0.000   8.451   5.634  1.408   0.000   2.817
997  6.109   3.859  5.145   6.431   1.286   8.039  5.466   3.859   4.502
998  4.878   3.659  3.659   6.098   3.659   4.878  3.659   6.098   2.439
999  4.587  12.844  4.587   7.339   0.917   4.587  0.000   3.670   2.752
```

```
        X9  …     X11    X12    X13    X14     X15    X16    X17    X18  \
0    5.152  …  10.303  3.333  5.758  6.970   4.848  4.848  0.303  3.030
1    4.569  …   5.584  2.538  5.584  3.046   5.584  5.584  0.000  1.523
2    8.065  …   7.258  2.419  4.839  4.839   8.871  3.226  0.000  3.226
3    5.714  …  12.571  1.714  4.571  2.286   0.571  3.429  1.143  4.000
4    4.423  …   4.914  2.457  3.440  3.440   6.880  6.143  0.737  2.948
..     …  …     …    …    …    …     …    …    …    …
995  6.905  …   3.836  3.581  9.463  1.790   7.928  5.627  2.046  3.836
996  4.225  …  14.085  5.634  4.225  1.408  12.676  8.451  0.000  5.634
997  6.109  …   3.537  2.894  2.251  4.823  10.611  6.109  0.000  1.608
998  1.220  …  13.415  2.439  8.537  6.098   8.537  4.878  2.439  2.439
999  5.505  …   2.752  2.752  1.835  4.587   8.257  6.422  1.835  3.670

        X19         X20
0     5.455   efectores
1     9.137   efectores
2     4.032   efectores
3     5.143   efectores
4     6.388   efectores
..      …         …
995  10.230   efectores
996   0.000   efectores
997   5.788   efectores
998   2.439   efectores
999  11.927   efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.00000
mean      6.810689     6.092888     4.368911     5.372661     2.32793
std       2.673343     2.813390     1.816448     2.307954     1.99220
min       0.000000     0.000000     0.000000     0.000000     0.00000
25%       5.167000     4.348000     3.251250     3.879750     1.07400
50%       6.610500     5.684000     4.240500     5.289500     1.92450
75%       8.231750     7.407000     5.389000     6.607500     2.99400
max      22.892000    36.364000    17.241000    17.045000    16.50300

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.377445     3.775600     5.738327     2.437203     5.659319
std       2.934035     2.030226     3.231297     1.421578     2.249569
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.519000	2.516000	3.794750	1.515000	4.082000
50%	6.043500	3.567000	5.288000	2.302000	5.501000
75%	7.819750	4.718000	7.059000	3.158000	7.080000
max	30.303000	28.684000	30.303000	12.571000	19.444000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.968827	5.966024	2.852144	4.287785	4.847484
std	2.855206	2.931394	1.440594	2.034207	2.987635
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.055750	3.974750	1.891000	2.964750	3.092000
50%	9.004500	5.682000	2.632000	4.133000	4.387500
75%	10.784000	7.375500	3.585500	5.492000	5.909000
max	21.311000	22.115000	12.500000	13.636000	26.718000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	7.678113	5.492777	1.204874	3.227920	6.513063
std	2.953119	2.220631	1.054852	2.084197	2.536191
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	5.710000	4.167000	0.454250	2.029750	4.972500
50%	7.341000	5.368500	1.043500	3.030000	6.364000
75%	9.333000	6.529500	1.716250	4.086250	7.864250
max	20.930000	19.780000	7.576000	28.571000	35.450000

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	7.595	3.376	4.219	5.063	0.000	7.595	7.595	4.219	0.844	5.485
1	7.407	2.469	4.012	3.086	2.469	3.704	2.160	3.086	1.852	7.716
2	6.908	9.211	3.618	7.895	0.000	14.474	6.250	4.605	1.316	4.276
3	1.562	3.125	7.812	1.562	0.000	0.000	15.625	3.125	1.562	9.375
4	8.772	2.924	5.848	5.263	1.170	9.357	3.509	3.509	3.509	8.187
..	…	…	…	…	…	…	…	…	…	
995	7.973	4.784	2.733	2.506	1.822	2.506	2.050	6.378	1.139	8.428
996	7.384	6.228	4.804	5.961	3.648	6.673	5.249	8.007	1.601	5.783
997	4.930	21.831	2.817	4.930	0.000	13.380	8.451	1.408	1.408	3.521
998	0.948	1.896	11.374	2.370	2.370	5.687	5.213	5.213	2.370	6.635
999	6.625	5.678	5.047	5.363	0.946	4.101	3.785	4.416	1.577	7.886

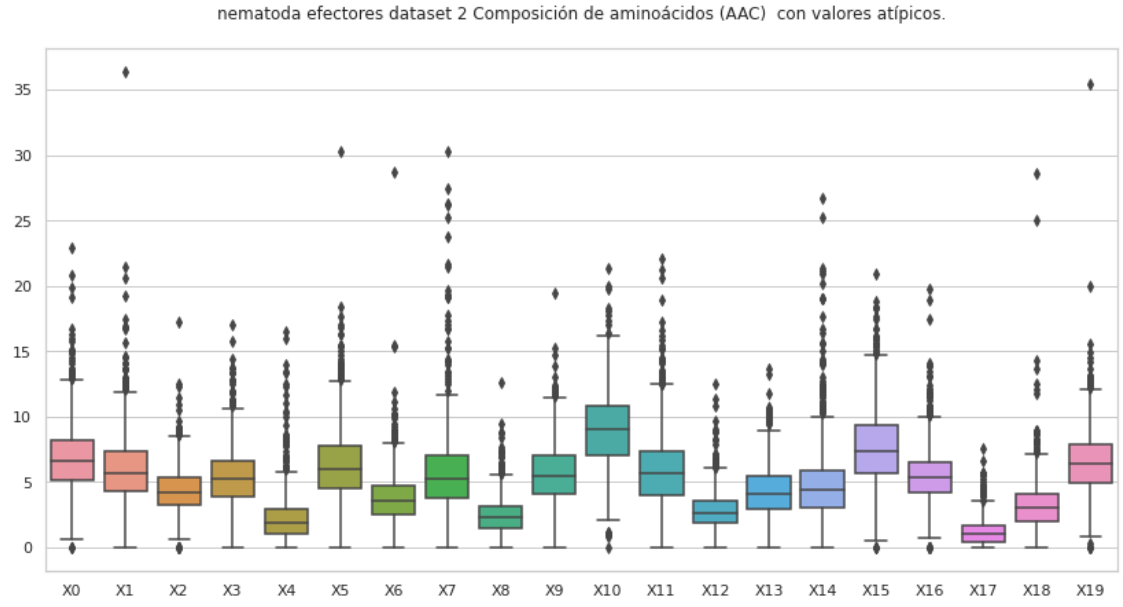
	…	X11	X12	X13	X14	X15	X16	X17	X18	X19 \
0	…	6.329	5.063	5.063	5.907	8.017	4.641	0.422	2.532	6.751

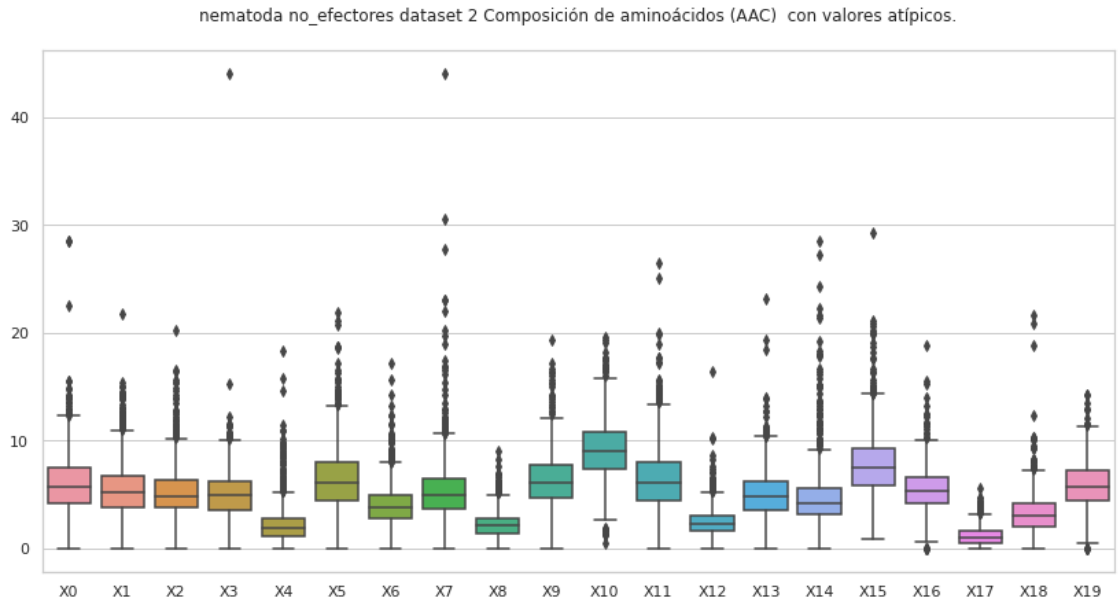
```
1    …   4.321  4.321  8.333  4.012   7.407   7.407  1.235  6.173  8.951
2    …   8.224  3.289  2.961  2.632   4.934   5.592  0.658  2.632  4.276
3    …   6.250  1.562  4.688  3.125  15.625  12.500  0.000  3.125  1.562
4    …   9.942  4.094  1.754  4.094   5.263   4.678  0.585  1.754  5.848
..   …     …    …    …    …     …     …    …    …    …
995  …   4.100  2.733  9.112  5.011   8.200   5.923  0.456  3.872  7.289
996  …   5.249  1.779  5.249  3.381   5.338   6.139  1.423  2.936  6.495
997  …  10.563  1.408  0.704  0.000  11.972   4.930  0.000  0.000  1.408
998  …   9.479  0.948  9.005  2.370  11.848   7.109  0.474  6.635  2.370
999  …   6.309  2.524  5.678  4.101   6.309   5.994  1.262  4.732  8.202

                X20
0     no_efectores
1     no_efectores
2     no_efectores
3     no_efectores
4     no_efectores
..             …
995   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[1000 rows x 21 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.968600     5.482970     5.273291     4.948548     2.268491
std       2.689825     2.536327     2.423353     2.407029     1.992934
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.225000     3.870000     3.791750     3.528250     1.091500
50%       5.761500     5.239000     4.875000     4.972000     1.852000
75%       7.465000     6.692750     6.381500     6.193500     2.779250
max      28.571000    21.831000    20.238000    44.048000    18.293000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.418393     4.037261     5.457695     2.188899     6.346578
std       3.120173     1.985836     3.141800     1.279182     2.584189
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.447750     2.829250     3.741000     1.366250     4.651000
50%       6.098000     3.813000     4.986000     2.108000     6.144500
```

```
75%       8.000000     4.920250     6.534500     2.857000     7.752000
max      21.918000    17.143000    44.000000     9.091000    19.403000

                X10          X11          X12          X13          X14  \
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       9.181312     6.560000     2.509824     5.021513     4.757538
std        2.879302     3.147508     1.342237     2.347986     3.046853
min        0.461000     0.000000     0.000000     0.000000     0.000000
25%        7.402000     4.467750     1.625500     3.503750     3.175000
50%        9.065500     6.158000     2.324500     4.816000     4.149000
75%       10.824000     8.041500     3.103750     6.290250     5.623000
max       19.643000    26.531000    16.444000    23.121000    28.571000

                X15          X16          X17          X18          X19
count   1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean       7.912483     5.433426     1.131518     3.290480     5.811169
std        2.979190     2.057532     0.905266     1.987878     2.148712
min        0.870000     0.000000     0.000000     0.000000     0.000000
25%        5.920000     4.167000     0.485000     2.093000     4.436500
50%        7.522000     5.319000     0.982000     3.077000     5.772500
75%        9.302750     6.576750     1.604750     4.214500     7.221250
max       29.302000    18.868000     5.660000    21.622000    14.286000
```





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    5.758   4.848  1.818   7.879  0.909   4.848  3.030  11.515  3.030  5.152
1    6.599   5.584  4.061   5.076  3.046   5.076  7.107   4.569  1.523  4.569
2    7.258   4.839  5.645  10.484  2.419   4.032  2.419   1.613  2.419  8.065
6    4.878   3.049  3.354   8.841  3.659   6.098  2.439   7.012  3.659  4.878
7    2.649   3.974  3.311   4.636  0.662  10.596  3.311   5.960  1.325  4.636
..     …       …      …       …      …       …      …       …      …      …
994  5.202   3.468  5.202   7.514  2.312   5.780  2.312   5.780  2.890  6.936
995  5.115   8.951  3.581   2.813  3.069   2.558  2.046   4.092  1.023  6.905
997  6.109   3.859  5.145   6.431  1.286   8.039  5.466   3.859  4.502  6.109
998  4.878   3.659  3.659   6.098  3.659   4.878  3.659   6.098  2.439  1.220
999  4.587  12.844  4.587   7.339  0.917   4.587  0.000   3.670  2.752  5.505

     …     X11    X12    X13    X14     X15    X16    X17    X18     X19  \
0    …  10.303  3.333  5.758  6.970   4.848  4.848  0.303  3.030   5.455
1    …   5.584  2.538  5.584  3.046   5.584  5.584  0.000  1.523   9.137
2    …   7.258  2.419  4.839  4.839   8.871  3.226  0.000  3.226   4.032
6    …   4.573  2.439  3.963  7.622  10.061  5.793  2.134  2.134   3.659
7    …  11.258  1.987  3.311  5.960  10.596  4.636  0.662  1.325   7.285
..   …      …      …      …      …       …      …      …      …       …
994  …   3.468  1.734  4.624  2.890  10.983  6.358  0.578  2.890   6.936
```

```
995  …   3.836  3.581  9.463  1.790   7.928  5.627  2.046  3.836  10.230
997  …   3.537  2.894  2.251  4.823  10.611  6.109  0.000  1.608   5.788
998  …  13.415  2.439  8.537  6.098   8.537  4.878  2.439  2.439   2.439
999  …   2.752  2.752  1.835  4.587   8.257  6.422  1.835  3.670  11.927

          X20
0    efectores
1    efectores
2    efectores
6    efectores
7    efectores
..         …
994  efectores
995  efectores
997  efectores
998  efectores
999  efectores

[828 rows x 21 columns]
```

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

```
                X0          X1          X2          X3          X4          X5  \
count  828.000000  828.000000  828.000000  828.000000  828.000000  828.000000
mean     6.799083    6.068024    4.397657    5.454164    2.247661    6.407594
std      2.305112    2.292023    1.613698    1.972041    1.563496    2.555603
min      1.200000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.263000    4.493000    3.349000    4.098000    1.168000    4.734000
50%      6.728000    5.726500    4.328500    5.409500    1.953500    6.125000
75%      8.174500    7.350750    5.389000    6.647000    2.950250    7.896000
max     14.815000   14.504000    9.302000   11.913000    8.271000   15.035000

                X6          X7          X8          X9         X10         X11  \
count  828.000000  828.000000  828.000000  828.000000  828.000000  828.000000
mean     3.689293    5.474331    2.433494    5.783091    9.235746    5.987267
std      1.638622    2.271173    1.179031    2.017295    2.491476    2.470719
min      0.000000    0.000000    0.000000    0.000000    2.062000    0.000000
25%      2.597000    3.876000    1.645750    4.344750    7.580500    4.325500
50%      3.561500    5.279500    2.349500    5.655500    9.240500    5.749500
75%      4.643500    6.861750    3.146250    7.143000   10.843000    7.266250
max      9.412000   15.260000    6.542000   12.360000   17.365000   14.286000

               X12         X13         X14         X15         X16         X17  \
count  828.000000  828.000000  828.000000  828.000000  828.000000  828.000000
mean     2.789651    4.406564    4.580083    7.685255    5.485769    1.188815
```

std	1.202218	1.854780	2.131980	2.623095	1.878884	0.878384
min	0.000000	0.000000	0.000000	1.242000	0.000000	0.000000
25%	1.922750	3.147000	3.179500	5.882000	4.316000	0.571000
50%	2.637500	4.281000	4.370500	7.435000	5.424500	1.096500
75%	3.509000	5.529250	5.722750	9.218250	6.403250	1.706000
max	7.143000	10.345000	13.074000	16.208000	12.069000	4.348000

	X18	X19
count	828.000000	828.000000
mean	3.207242	6.679187
std	1.543963	2.155833
min	0.000000	0.000000
25%	2.247000	5.341250
50%	3.075500	6.557000
75%	4.082000	7.975500
max	8.955000	13.699000

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	7.595	3.376	4.219	5.063	0.000	7.595	7.595	4.219	0.844	5.485	
1	7.407	2.469	4.012	3.086	2.469	3.704	2.160	3.086	1.852	7.716	
2	6.908	9.211	3.618	7.895	0.000	14.474	6.250	4.605	1.316	4.276	
4	8.772	2.924	5.848	5.263	1.170	9.357	3.509	3.509	3.509	8.187	
5	3.182	5.227	7.273	4.091	2.045	7.727	4.545	3.864	2.500	11.591	
..	…	…	…	…	…	…	…	…	…	…	
994	8.788	5.326	5.326	4.927	1.465	4.527	4.927	2.530	2.397	5.593	
995	7.973	4.784	2.733	2.506	1.822	2.506	2.050	6.378	1.139	8.428	
996	7.384	6.228	4.804	5.961	3.648	6.673	5.249	8.007	1.601	5.783	
998	0.948	1.896	11.374	2.370	2.370	5.687	5.213	5.213	2.370	6.635	
999	6.625	5.678	5.047	5.363	0.946	4.101	3.785	4.416	1.577	7.886	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	6.329	5.063	5.063	5.907	8.017	4.641	0.422	2.532	6.751	
1	…	4.321	4.321	8.333	4.012	7.407	7.407	1.235	6.173	8.951	
2	…	8.224	3.289	2.961	2.632	4.934	5.592	0.658	2.632	4.276	
4	…	9.942	4.094	1.754	4.094	5.263	4.678	0.585	1.754	5.848	
5	…	7.500	1.818	4.091	2.955	5.227	4.545	1.818	6.364	5.227	
..	…	…	…	…	…	…	…	…	…	…	
994	…	4.394	2.130	4.794	5.193	9.321	4.927	1.731	2.929	6.258	
995	…	4.100	2.733	9.112	5.011	8.200	5.923	0.456	3.872	7.289	
996	…	5.249	1.779	5.249	3.381	5.338	6.139	1.423	2.936	6.495	
998	…	9.479	0.948	9.005	2.370	11.848	7.109	0.474	6.635	2.370	

```
999  …  6.309  2.524  5.678  4.101   6.309  5.994  1.262  4.732  8.202

              X20
0     no_efectores
1     no_efectores
2     no_efectores
4     no_efectores
5     no_efectores
..             …
994   no_efectores
995   no_efectores
996   no_efectores
998   no_efectores
999   no_efectores

[818 rows x 21 columns]
```

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

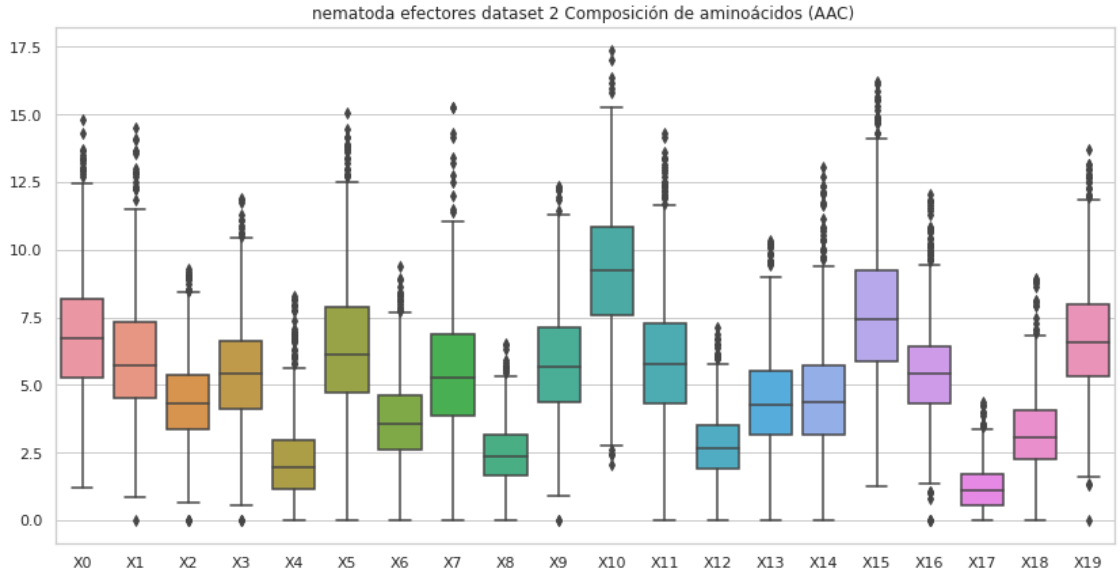
	X0	X1	X2	X3	X4	X5	\
count	818.000000	818.000000	818.000000	818.000000	818.000000	818.000000	
mean	5.952101	5.477578	5.212833	5.113730	2.077300	6.564998	
std	2.168405	2.139155	2.016208	1.880834	1.411257	2.719663	
min	0.000000	0.625000	0.000000	0.000000	0.000000	0.565000	
25%	4.371500	4.101000	3.911250	3.881500	1.145500	4.682000	
50%	5.783500	5.325500	4.904000	5.128000	1.852000	6.319500	
75%	7.432500	6.624750	6.309750	6.258250	2.614250	8.160250	
max	13.971000	12.941000	12.515000	11.538000	8.187000	15.661000	

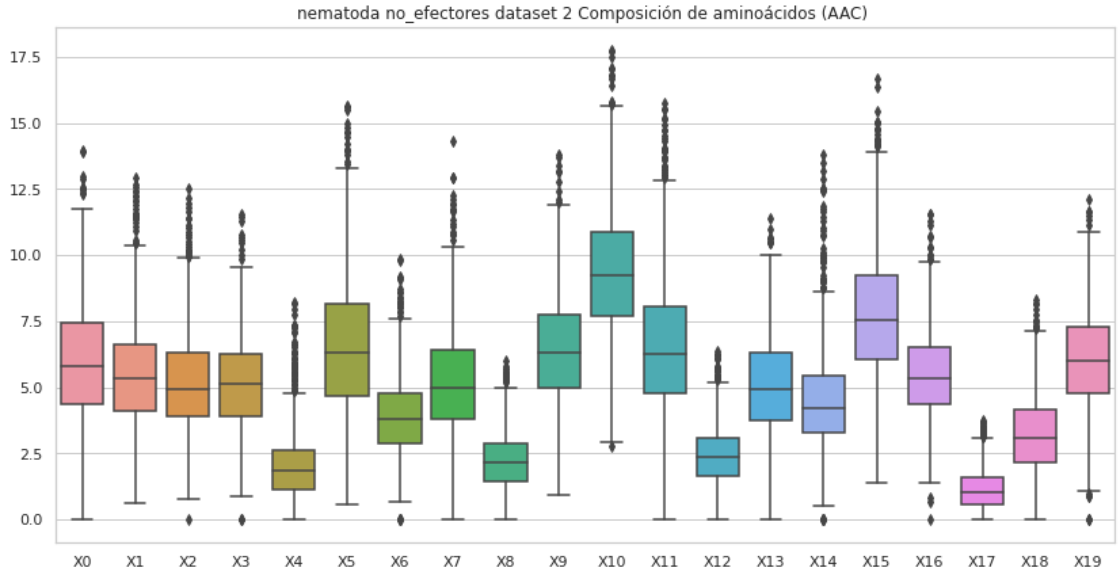
	X6	X7	X8	X9	X10	X11	\
count	818.000000	818.000000	818.000000	818.000000	818.000000	818.000000	
mean	3.946075	5.207076	2.212351	6.432579	9.402630	6.637444	
std	1.612074	2.072291	1.125988	2.173029	2.560245	2.743068	
min	0.000000	0.000000	0.000000	0.909000	2.768000	0.000000	
25%	2.877250	3.803000	1.449250	4.950000	7.704750	4.789000	
50%	3.807000	5.000000	2.174000	6.290000	9.233000	6.253500	
75%	4.792500	6.430750	2.878750	7.744750	10.883750	8.044500	
max	9.862000	14.286000	5.991000	13.821000	17.742000	15.723000	

	X12	X13	X14	X15	X16	X17	\
count	818.000000	818.000000	818.000000	818.000000	818.000000	818.000000	
mean	2.470075	5.049941	4.557356	7.824642	5.450196	1.148064	
std	1.094456	1.927998	2.083754	2.523408	1.753303	0.809050	
min	0.000000	0.000000	0.000000	1.389000	0.000000	0.000000	
25%	1.667000	3.724000	3.306750	6.040250	4.348000	0.585750	

50%	2.345500	4.935500	4.211000	7.535000	5.355500	1.031500
75%	3.066000	6.292750	5.455000	9.233500	6.512250	1.602250
max	6.364000	11.364000	13.776000	16.667000	11.547000	3.731000

	X18	X19
count	818.000000	818.000000
mean	3.232290	6.030729
std	1.569781	1.951784
min	0.000000	0.000000
25%	2.174750	4.762000
50%	3.099500	5.992500
75%	4.168500	7.293500
max	8.333000	12.121000





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	0.023863	0.003768	0.032654	0.020095	0.023863	0.047725	0.012559	
1	0.057918	0.026731	0.044552	0.044552	0.049008	0.040097	0.013366	
2	0.027834	0.009278	0.040205	0.015464	0.018556	0.006185	0.009278	
3	0.031483	0.001968	0.011806	0.039353	0.015741	0.017709	0.043288	
4	0.012633	0.063163	0.034108	0.029055	0.017686	0.045477	0.016422	
..	…	…	…	…	…	…	…	
995	0.024342	0.014605	0.013388	0.012171	0.045032	0.019473	0.004868	
996	0.057652	0.069183	0.000000	0.046122	0.034591	0.000000	0.023061	
997	0.031954	0.006727	0.033636	0.042045	0.011773	0.020182	0.023545	
998	0.071735	0.053802	0.089669	0.071735	0.125537	0.089669	0.035868	
999	0.041452	0.008290	0.066324	0.041452	0.016581	0.033162	0.024871	

	X7	X8	X9	…	X74	X75	X76	\
0	0.021351	0.042702	0.026375	…	0.005645	0.031443	-0.008278	
1	0.040097	0.049008	0.124747	…	-0.020004	-0.027943	0.005226	
2	0.030927	0.027834	0.046391	…	-0.005224	0.010450	0.002302	
3	0.019677	0.043288	0.017709	…	-0.017486	-0.010488	0.020431	
4	0.022739	0.025265	0.027792	…	-0.001965	-0.008444	0.044886	
..	…	…	…	…	…	…	…	
995	0.032861	0.018256	0.054769	…	-0.017398	-0.003537	0.025597	
996	0.034591	0.115304	0.092244	…	0.051978	0.075514	-0.027625	
997	0.031954	0.018500	0.060545	…	-0.024434	0.001155	0.007032	
998	0.017934	0.197273	0.125537	…	-0.039660	0.089055	-0.083908	
999	0.049743	0.024871	0.082904	…	-0.012081	0.053036	-0.016933	

	X77	X78	X79	X80	X81	X82	X83
0	-0.008938	0.004082	0.004575	0.024643	0.021343	0.019835	efectores
1	-0.013961	-0.006100	-0.004929	0.023606	0.040371	0.003511	efectores
2	0.006384	0.020209	0.018332	0.013658	0.014575	0.001365	efectores
3	-0.007918	0.021904	0.002707	-0.013399	0.026097	0.005039	efectores
4	0.000541	0.008916	0.016945	-0.017119	0.009605	0.007281	efectores
..	…	…	…	…	…	…	…
995	-0.020148	-0.021530	-0.001831	-0.004834	-0.012363	-0.009448	efectores
996	0.063478	0.103994	-0.101395	-0.025988	0.007839	0.015041	efectores
997	-0.020823	-0.003895	0.032947	-0.007918	0.000828	0.017497	efectores

```
998 -0.076355 -0.178578  0.000945  0.092784 -0.074532 -0.003151  efectores
999  0.079015  0.060765  0.059460  0.044247 -0.029724  0.024278  efectores

[1000 rows x 84 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.037603	0.013397	0.030934	0.036162	0.026711	
std	0.040289	0.015137	0.033754	0.036953	0.036259	
min	-0.061729	-0.015432	-0.030864	-0.030864	-0.108025	
25%	0.022495	0.004582	0.016478	0.019117	0.011763	
50%	0.032852	0.009577	0.026454	0.030766	0.020527	
75%	0.044389	0.016742	0.039077	0.044040	0.031749	
max	1.066978	0.140626	0.853582	0.853582	0.853582	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.030967	0.014764	0.033327	0.034727	0.052422	…	
std	0.025055	0.018436	0.032574	0.053509	0.042325	…	
min	-0.030864	-0.046296	-0.061729	-0.092593	-0.138889	…	
25%	0.017698	0.006273	0.016785	0.017534	0.029851	…	
50%	0.026169	0.011406	0.028036	0.027394	0.044362	…	
75%	0.038585	0.018982	0.039701	0.041705	0.065005	…	
max	0.426791	0.426791	0.640187	1.493769	0.640187	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.009953	0.001091	0.006653	0.009054	0.000452	
std	0.037383	0.035064	0.032869	0.044619	0.055929	
min	-0.781767	-0.262925	-0.283349	-1.086521	-1.300038	
25%	-0.001284	-0.011249	-0.004267	0.000888	-0.010865	
50%	0.011673	0.002454	0.006647	0.010758	0.002786	
75%	0.023271	0.015327	0.019575	0.022258	0.017734	
max	0.223624	0.255096	0.433914	0.197894	0.318366	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.005796	0.010200	0.001353	0.005188	0.011777
std	0.035949	0.034000	0.044940	0.031727	0.045919
min	-0.639007	-0.565071	-0.402143	-0.265808	-0.265206
25%	-0.005001	0.000160	-0.011291	-0.005159	0.000077
50%	0.006882	0.011813	0.002909	0.006861	0.010996
75%	0.019806	0.023525	0.015326	0.019459	0.022231

```
max       0.207504     0.212516     0.768988     0.286601     1.163490

[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	0.033482	0.000000	0.022321	0.033482	0.022321	0.018601	0.003720	
1	0.027127	0.009043	0.011303	0.013564	0.030518	0.011303	0.006782	
2	0.013223	0.000000	0.015112	0.027705	0.005667	0.008815	0.002519	
3	0.014965	0.000000	0.014965	0.000000	0.044896	0.029931	0.014965	
4	0.052406	0.006987	0.031443	0.055899	0.010481	0.020962	0.020962	
..	…	…	…	…	…	…	…	
995	0.026054	0.005955	0.008188	0.008188	0.029776	0.020843	0.003722	
996	0.036849	0.018203	0.029746	0.033297	0.026194	0.039957	0.007991	
997	0.003519	0.000000	0.003519	0.009550	0.000503	0.001005	0.001005	
998	0.006666	0.016666	0.016666	0.039998	0.063331	0.036665	0.016666	
999	0.035730	0.005104	0.028924	0.022119	0.030626	0.023820	0.008507	

	X7	X8	X9	…	X74	X75	X76	\
0	0.024181	0.027901	0.040922	…	-0.010669	0.000553	0.000904	
1	0.028258	0.015824	0.036170	…	0.012335	0.011147	-0.002152	
2	0.008186	0.015741	0.011963	…	-0.002419	0.017769	-0.003800	
3	0.089793	0.059862	0.074827	…	-0.049249	-0.009210	-0.006560	
4	0.048912	0.059393	0.059393	…	-0.022879	-0.021534	-0.008666	
..	…	…	…	…	…	…	…	
995	0.027543	0.013399	0.042431	…	0.013521	0.006475	0.010544	
996	0.028858	0.026194	0.033297	…	-0.006936	0.003505	0.006289	
997	0.002513	0.007540	0.004524	…	0.024932	0.023157	0.001444	
998	0.046665	0.066664	0.039998	…	-0.006401	-0.006740	-0.013715	
999	0.042536	0.034029	0.051043	…	-0.010132	-0.011198	-0.009684	

	X77	X78	X79	X80	X81	X82	X83
0	-0.006639	-0.009693	0.008575	-0.035002	-0.005602	0.001355	no_efectores
1	0.015168	0.012593	-0.003173	0.002607	-0.002711	-0.005599	no_efectores
2	0.009983	0.027588	-0.000697	0.006634	0.033341	-0.001778	no_efectores
3	-0.030628	-0.044082	-0.000554	0.078406	0.101580	0.075697	no_efectores
4	0.005845	0.018819	-0.003283	0.011572	0.048797	0.007849	no_efectores
..	…	…	…	…	…	…	
995	0.006640	0.001535	0.019780	0.016415	-0.000066	0.020291	no_efectores
996	-0.006242	0.014057	0.021518	-0.006587	0.008919	0.005525	no_efectores
997	0.017621	0.017137	0.000427	0.018414	0.016096	0.001274	no_efectores
998	-0.032413	-0.027042	-0.005683	0.061150	0.042544	-0.011273	no_efectores

```
999  0.012402  0.004062  0.024512  0.009448 -0.001867  0.001444  no_efectores

[1000 rows x 84 columns]
```

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

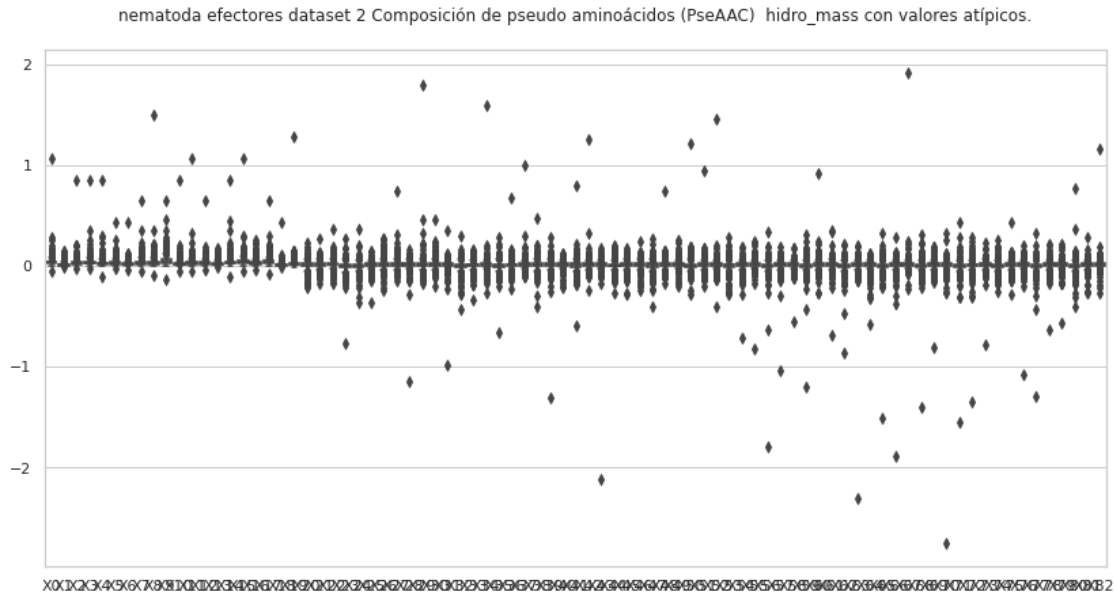
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.030594	0.013064	0.027314	0.033985	0.028162	
std	0.025349	0.023619	0.021672	0.026630	0.022631	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018310	0.004343	0.013540	0.018196	0.013879	
50%	0.027222	0.008777	0.022655	0.029721	0.023477	
75%	0.037569	0.015899	0.035563	0.043778	0.035634	
max	0.607059	0.607059	0.242824	0.485647	0.242824	

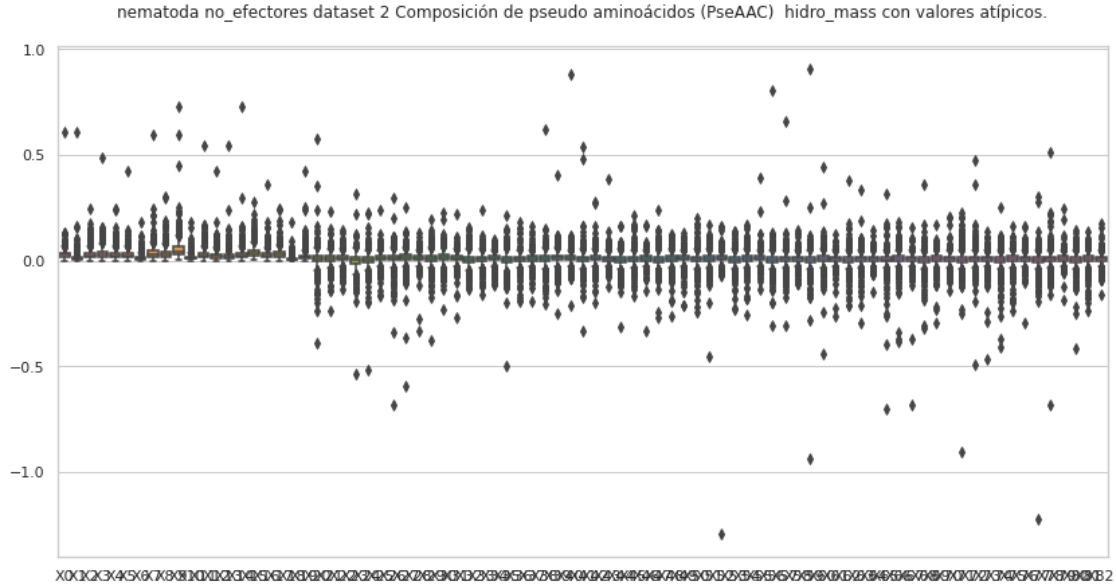
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.027895	0.012563	0.035366	0.035016	0.051704	…	
std	0.021379	0.012959	0.030694	0.026778	0.044479	…	
min	0.000000	0.000000	0.000000	0.000000	0.000806	…	
25%	0.015557	0.004947	0.018506	0.018371	0.029134	…	
50%	0.023905	0.010007	0.028917	0.029961	0.042786	…	
75%	0.035886	0.016563	0.044868	0.043376	0.063806	…	
max	0.424942	0.182118	0.595819	0.300332	0.728471	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.007292	0.000428	0.006267	0.008200	0.001075	
std	0.029607	0.036161	0.028941	0.024006	0.050291	
min	-0.465245	-0.412831	-0.238560	-0.294284	-1.225319	
25%	-0.000805	-0.008068	-0.002570	-0.001416	-0.007789	
50%	0.008851	0.003713	0.008225	0.008107	0.003711	
75%	0.018163	0.014374	0.019353	0.018192	0.014758	
max	0.134364	0.248788	0.163849	0.161626	0.299847	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007104	0.009061	0.002461	0.007347	0.008296
std	0.039230	0.022383	0.031916	0.028471	0.022350
min	-0.685083	-0.187117	-0.414808	-0.236770	-0.163198
25%	-0.002436	-0.000241	-0.007514	-0.002262	-0.001169
50%	0.007256	0.009028	0.004490	0.007425	0.007703
75%	0.018718	0.019292	0.015800	0.020538	0.018417
max	0.509182	0.247594	0.144149	0.137643	0.172524

[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	0.023863	0.003768	0.032654	0.020095	0.023863	0.047725	0.012559	
1	0.057918	0.026731	0.044552	0.044552	0.049008	0.040097	0.013366	
2	0.027834	0.009278	0.040205	0.015464	0.018556	0.006185	0.009278	
3	0.031483	0.001968	0.011806	0.039353	0.015741	0.017709	0.043288	
6	0.024326	0.018245	0.044091	0.030408	0.019765	0.034969	0.018245	
..	…	…	…	…	…	…	…	
993	0.016606	0.007866	0.019228	0.028843	0.013984	0.018354	0.007866	
994	0.026068	0.011586	0.037653	0.028964	0.023171	0.028964	0.014482	
995	0.024342	0.014605	0.013388	0.012171	0.045032	0.019473	0.004868	
997	0.031954	0.006727	0.033636	0.042045	0.011773	0.020182	0.023545	
999	0.041452	0.008290	0.066324	0.041452	0.016581	0.033162	0.024871	

	X7	X8	X9	…	X74	X75	X76	\
0	0.021351	0.042702	0.026375	…	0.005645	0.031443	-0.008278	
1	0.040097	0.049008	0.124747	…	-0.020004	-0.027943	0.005226	
2	0.030927	0.027834	0.046391	…	-0.005224	0.010450	0.002302	
3	0.019677	0.043288	0.017709	…	-0.017486	-0.010488	0.020431	
6	0.024326	0.022806	0.048652	…	-0.006163	0.008103	0.011754	
..	…	…	…	…	…	…	…	
993	0.009614	0.013984	0.018354	…	0.004661	0.022149	-0.003038	
994	0.034757	0.017378	0.060825	…	-0.054324	-0.027097	-0.006221	
995	0.032861	0.018256	0.054769	…	-0.017398	-0.003537	0.025597	
997	0.031954	0.018500	0.060545	…	-0.024434	0.001155	0.007032	
999	0.049743	0.024871	0.082904	…	-0.012081	0.053036	-0.016933	

	X77	X78	X79	X80	X81	X82	X83
0	-0.008938	0.004082	0.004575	0.024643	0.021343	0.019835	efectores
1	-0.013961	-0.006100	-0.004929	0.023606	0.040371	0.003511	efectores
2	0.006384	0.020209	0.018332	0.013658	0.014575	0.001365	efectores
3	-0.007918	0.021904	0.002707	-0.013399	0.026097	0.005039	efectores
6	-0.016769	-0.009855	0.006660	0.018095	0.015346	0.028076	efectores
..	…	…	…	…	…	…	
993	-0.002241	0.020106	0.013903	0.019869	0.025639	0.020869	efectores
994	-0.010720	-0.006498	0.016953	0.029581	0.020729	0.033352	efectores
995	-0.020148	-0.021530	-0.001831	-0.004834	-0.012363	-0.009448	efectores
997	-0.020823	-0.003895	0.032947	-0.007918	0.000828	0.017497	efectores
999	0.079015	0.060765	0.059460	0.044247	-0.029724	0.024278	efectores

[890 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	890.000000	890.000000	890.000000	890.000000	890.000000	890.000000	
mean	0.033619	0.011371	0.027354	0.031493	0.021871	0.027252	
std	0.016844	0.009857	0.015758	0.017855	0.014994	0.014841	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021914	0.004523	0.016192	0.018612	0.011027	0.016863	
50%	0.031778	0.009014	0.025067	0.029184	0.019312	0.024843	
75%	0.042440	0.015443	0.036346	0.040642	0.029356	0.035224	
max	0.120491	0.057644	0.092415	0.105766	0.092307	0.100476	

	X6	X7	X8	X9	…	X73	\
count	890.000000	890.000000	890.000000	890.000000	…	890.000000	
mean	0.012388	0.028746	0.029095	0.045448	…	0.011783	
std	0.008745	0.017400	0.017157	0.024072	…	0.019574	
min	0.000000	0.000000	0.000000	0.000944	…	-0.060261	
25%	0.006132	0.016070	0.016811	0.029078	…	0.000050	
50%	0.010827	0.026639	0.026507	0.042592	…	0.011830	
75%	0.017087	0.037345	0.038439	0.059906	…	0.022923	
max	0.051116	0.120081	0.123868	0.152667	…	0.085504	

	X74	X75	X76	X77	X78	X79	\
count	890.000000	890.000000	890.000000	890.000000	890.000000	890.000000	
mean	0.002056	0.007593	0.011653	0.003612	0.007682	0.011888	
std	0.022776	0.020603	0.018299	0.024217	0.021210	0.018708	
min	-0.094628	-0.076395	-0.083947	-0.136553	-0.097961	-0.086144	
25%	-0.009841	-0.002862	0.002545	-0.008927	-0.002891	0.001344	
50%	0.002702	0.007015	0.011248	0.003521	0.007247	0.011978	
75%	0.014615	0.019094	0.021879	0.017105	0.019224	0.022643	
max	0.094964	0.102187	0.074781	0.102959	0.091736	0.073519	

	X80	X81	X82
count	890.000000	890.000000	890.000000
mean	0.002185	0.006514	0.011955
std	0.023535	0.020862	0.019118
min	-0.107601	-0.075272	-0.074734
25%	-0.009452	-0.004123	0.001416
50%	0.002892	0.006867	0.011376
75%	0.014493	0.018957	0.021822
max	0.090808	0.094912	0.116640

[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    0.033482  0.000000  0.022321  0.033482  0.022321  0.018601  0.003720
1    0.027127  0.009043  0.011303  0.013564  0.030518  0.011303  0.006782
2    0.013223  0.000000  0.015112  0.027705  0.005667  0.008815  0.002519
4    0.052406  0.006987  0.031443  0.055899  0.010481  0.020962  0.020962
5    0.029397  0.018898  0.037796  0.071393  0.037796  0.035696  0.023098
..        …         …         …         …         …         …         …
995  0.026054  0.005955  0.008188  0.008188  0.029776  0.020843  0.003722
996  0.036849  0.018203  0.029746  0.033297  0.026194  0.039957  0.007991
997  0.003519  0.000000  0.003519  0.009550  0.000503  0.001005  0.001005
998  0.006666  0.016666  0.016666  0.039998  0.063331  0.036665  0.016666
999  0.035730  0.005104  0.028924  0.022119  0.030626  0.023820  0.008507

           X7        X8        X9  …       X74       X75       X76  \
0    0.024181  0.027901  0.040922  … -0.010669  0.000553  0.000904
1    0.028258  0.015824  0.036170  …  0.012335  0.011147 -0.002152
2    0.008186  0.015741  0.011963  … -0.002419  0.017769 -0.003800
4    0.048912  0.059393  0.059393  … -0.022879 -0.021534 -0.008666
5    0.107090  0.069293  0.077692  …  0.046993  0.042494 -0.016252
..        …         …         …  …         …         …         …
995  0.027543  0.013399  0.042431  …  0.013521  0.006475  0.010544
996  0.028858  0.026194  0.033297  … -0.006936  0.003505  0.006289
997  0.002513  0.007540  0.004524  …  0.024932  0.023157  0.001444
998  0.046665  0.066664  0.039998  … -0.006401 -0.006740 -0.013715
999  0.042536  0.034029  0.051043  … -0.010132 -0.011198 -0.009684

           X77       X78       X79       X80       X81       X82           X83
0   -0.006639 -0.009693  0.008575 -0.035002 -0.005602  0.001355  no_efectores
1    0.015168  0.012593 -0.003173  0.002607 -0.002711 -0.005599  no_efectores
2    0.009983  0.027588 -0.000697  0.006634  0.033341 -0.001778  no_efectores
4    0.005845  0.018819 -0.003283  0.011572  0.048797  0.007849  no_efectores
5   -0.038009 -0.079554 -0.014825 -0.040644 -0.025834  0.004558  no_efectores
..        …         …         …         …         …         …           …
995  0.006640  0.001535  0.019780  0.016415 -0.000066  0.020291  no_efectores
996 -0.006242  0.014057  0.021518 -0.006587  0.008919  0.005525  no_efectores
997  0.017621  0.017137  0.000427  0.018414  0.016096  0.001274  no_efectores
998 -0.032413 -0.027042 -0.005683  0.061150  0.042544 -0.011273  no_efectores
999  0.012402  0.004062  0.024512  0.009448 -0.001867  0.001444  no_efectores

[858 rows x 84 columns]
```

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

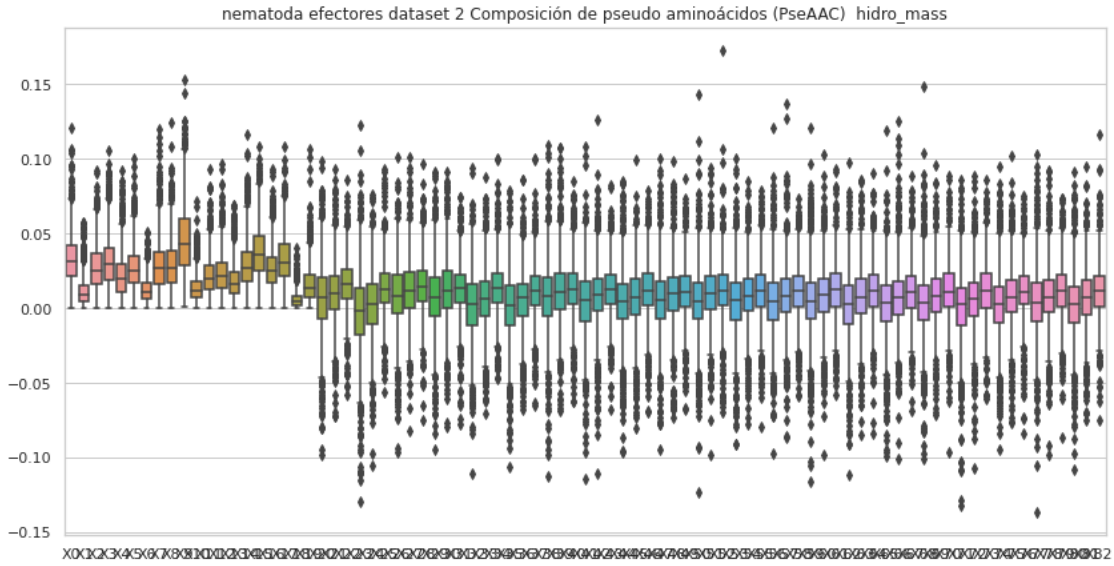
	X0	X1	X2	X3	X4	X5	\
count	858.000000	858.000000	858.000000	858.000000	858.000000	858.000000	
mean	0.027092	0.010369	0.023386	0.029686	0.023719	0.024545	
std	0.013093	0.009271	0.013894	0.016579	0.014452	0.013098	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.017706	0.004162	0.012671	0.017151	0.012980	0.014620	
50%	0.026188	0.008072	0.021009	0.027390	0.021580	0.022266	
75%	0.034555	0.013441	0.032550	0.039932	0.031361	0.032995	
max	0.081192	0.067804	0.090779	0.094060	0.087686	0.078878	

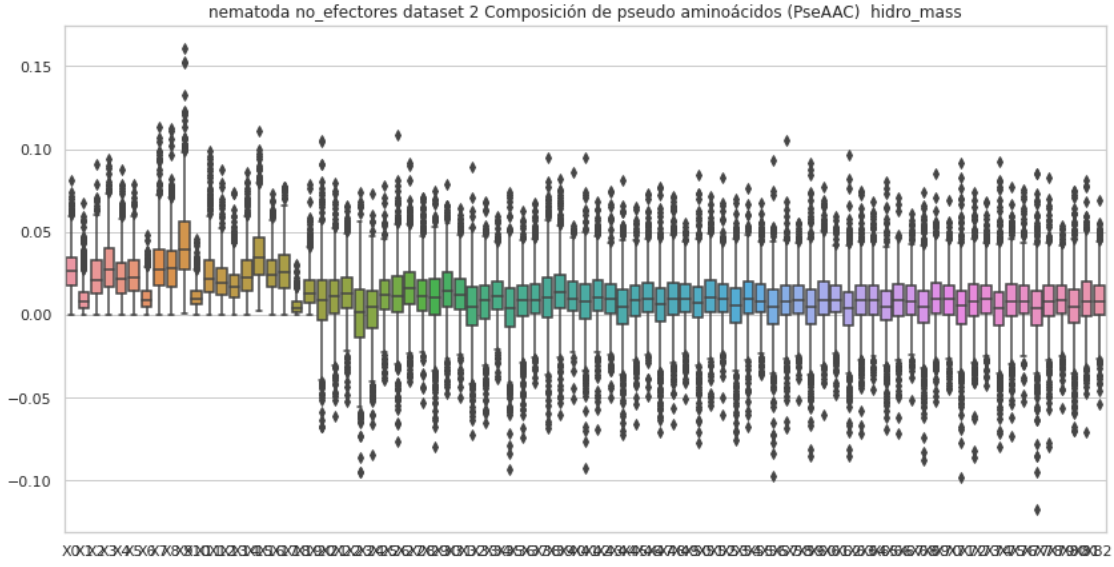
	X6	X7	X8	X9	…	X73	\
count	858.000000	858.000000	858.000000	858.000000	…	858.000000	
mean	0.010502	0.030214	0.029829	0.043299	…	0.009964	
std	0.007590	0.017927	0.017331	0.023600	…	0.015170	
min	0.000000	0.000000	0.000000	0.000806	…	-0.042763	
25%	0.004817	0.017746	0.017177	0.027572	…	0.000669	
50%	0.009136	0.027129	0.027749	0.039505	…	0.009257	
75%	0.014464	0.039637	0.038594	0.056079	…	0.017966	
max	0.048039	0.113456	0.112247	0.160721	…	0.068513	

	X74	X75	X76	X77	X78	X79	\
count	858.000000	858.000000	858.000000	858.000000	858.000000	858.000000	
mean	0.003338	0.008284	0.009291	0.003644	0.008131	0.008907	
std	0.019390	0.017268	0.015160	0.019890	0.016887	0.014632	
min	-0.080005	-0.073373	-0.052166	-0.117414	-0.079554	-0.056686	
25%	-0.006055	-0.000877	0.000410	-0.006104	-0.000520	0.000411	
50%	0.004198	0.008418	0.008341	0.004346	0.007763	0.008900	
75%	0.013711	0.018593	0.017489	0.014139	0.017448	0.017446	
max	0.092568	0.077217	0.075447	0.085507	0.083121	0.053611	

	X80	X81	X82
count	858.000000	858.000000	858.000000
mean	0.004540	0.009241	0.009196
std	0.018901	0.017235	0.015551
min	-0.070027	-0.070986	-0.054011
25%	-0.005132	-0.000953	0.000128
50%	0.005081	0.008033	0.008117
75%	0.015112	0.019896	0.018049
max	0.075061	0.080892	0.068819

[8 rows x 83 columns]





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    0.034968  0.005521  0.047851  0.029447  0.034968  0.069937  0.018404
1    0.046954  0.021671  0.036118  0.036118  0.039730  0.032507  0.010836
2    0.055956  0.018652  0.080825  0.031086  0.037304  0.012435  0.018652
3    0.103257  0.006454  0.038721  0.129071  0.051628  0.058082  0.141978
4    0.014791  0.073957  0.039937  0.034020  0.020708  0.053249  0.019229
..        …         …         …         …         …         …         …
995  0.046901  0.028140  0.025795  0.023450  0.086766  0.037520  0.009380
996  0.070385  0.084463  0.000000  0.056308  0.042231  0.000000  0.028154
997  0.034727  0.007311  0.036554  0.045693  0.012794  0.021933  0.025588
998  0.063560  0.047670  0.079450  0.063560  0.111231  0.079450  0.031780
999  0.048354  0.009671  0.077367  0.048354  0.019342  0.038684  0.029013

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.031288  0.062575  0.038649  …  0.017667  0.036538  0.021905
1    0.032507  0.039730  0.101131  … -0.007599 -0.019442  0.001068
2    0.062173  0.055956  0.093259  …  0.035339 -0.005915  0.002631
3    0.064536  0.141978  0.058082  …  0.047059 -0.054695 -0.015617
4    0.026624  0.029583  0.032541  …  0.037199  0.025153  0.009696
..        …         …        …  …        …         …
995  0.063316  0.035175  0.105526  …  0.009670 -0.026962  0.028653
996  0.042231  0.140771  0.112617  …  0.061275  0.072912  0.040898
997  0.034727  0.020105  0.065798  …  0.009581  0.020842  0.017380
998  0.015890  0.174791  0.111231  … -0.027370  0.019837 -0.026457
999  0.058025  0.029013  0.096709  … -0.038043 -0.053905  0.070920

          X35       X36       X37       X38       X39       X40        X41
0    0.001887  0.007706  0.009464 -0.012130  0.006704  0.029066  efectores
1    0.036282  0.047961  0.025069  0.004237 -0.003996  0.002846  efectores
2   -0.000783  0.019987  0.031218  0.004627  0.036853  0.002744  efectores
3    0.053288  0.025255 -0.033588  0.067009  0.008879  0.016528  efectores
4    0.049491  0.015604  0.007728  0.052556  0.019841  0.008526  efectores
..        …         …         …         …         …         …
995 -0.012449 -0.028983  0.018046  0.049319 -0.003528 -0.018205  efectores
996  0.042316  0.047058  0.021088 -0.033726 -0.123790  0.018363  efectores
997  0.022263  0.024037  0.021522  0.007642  0.035806  0.019015  efectores
998 -0.056711 -0.113095  0.062852 -0.074346  0.000837 -0.002792  efectores
999  0.003652  0.002995 -0.011266 -0.019753  0.069360  0.028320  efectores

[1000 rows x 42 columns]
```

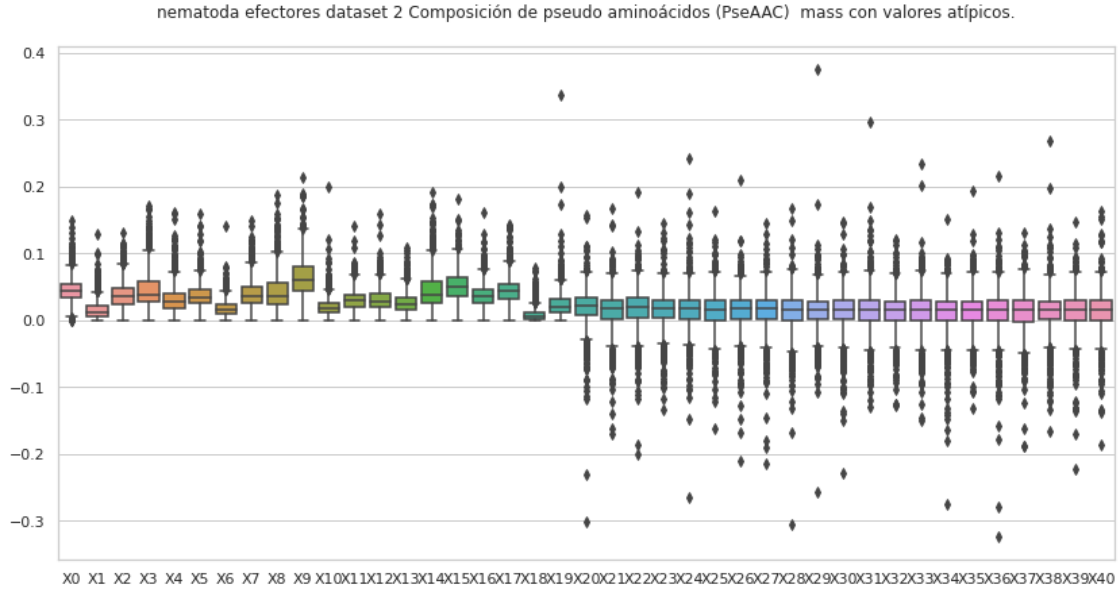
Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

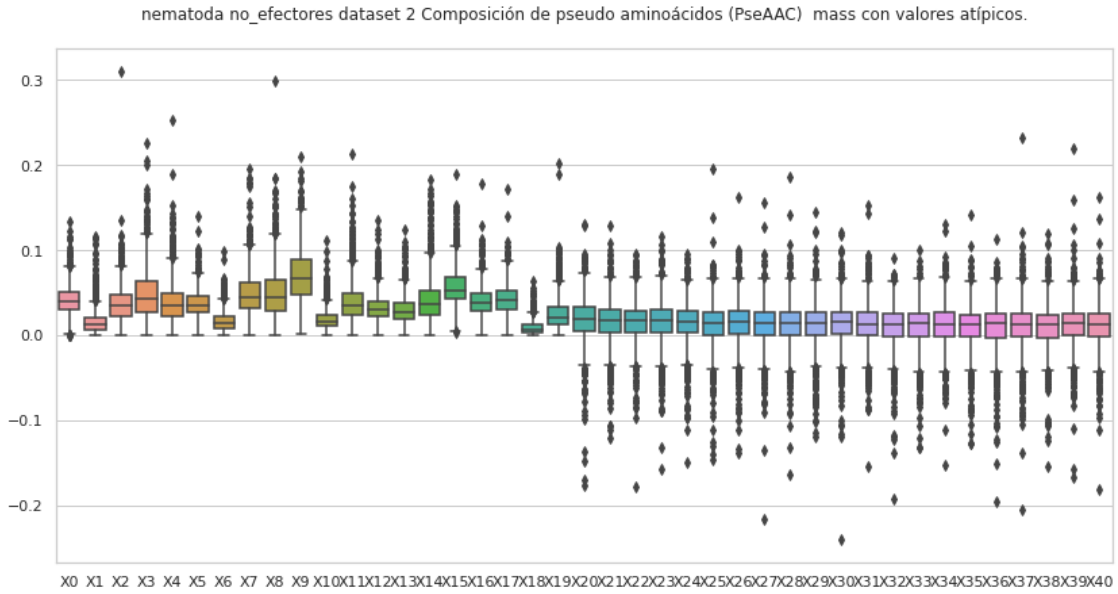
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.045646     0.016252     0.038065     0.046094     0.031580
std       0.018732     0.014763     0.020867     0.028110     0.021441
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034552     0.006719     0.023918     0.027613     0.017437
50%       0.044153     0.012455     0.035636     0.039214     0.027440
75%       0.054212     0.021422     0.048522     0.058923     0.039772
max       0.148908     0.129143     0.131238     0.171689     0.162390

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037188     0.017713     0.040422     0.043320     0.063992  …
std       0.017808     0.012926     0.021719     0.028093     0.029990  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.026006     0.009168     0.026084     0.024554     0.043605  …
```

```
50%       0.034215     0.015320     0.037124     0.037210     0.061469  …
75%       0.045627     0.023596     0.050608     0.055710     0.081211  …
max       0.159326     0.141978     0.148716     0.188343     0.213307  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014563     0.012737     0.014141     0.012064     0.013441
std       0.031618     0.027712     0.030256     0.031444     0.027762
min      -0.129734    -0.127147    -0.150510    -0.274729    -0.130591
25%       0.000013     0.000013    -0.000109    -0.000714    -0.000525
50%       0.015906     0.015820     0.016957     0.015214     0.015404
75%       0.030099     0.027986     0.029632     0.028634     0.029154
max       0.297526     0.120313     0.234816     0.150826     0.194377

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012462     0.012237     0.012925     0.013012     0.013344
std       0.033614     0.030888     0.031291     0.030350     0.030372
min      -0.323349    -0.187122    -0.164790    -0.221586    -0.186263
25%       0.000165    -0.001954     0.001195     0.000410     0.000119
50%       0.016162     0.016531     0.015104     0.016658     0.015413
75%       0.029554     0.029694     0.029075     0.029579     0.029379
max       0.217046     0.130729     0.269012     0.148128     0.164208

[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    0.052604  0.000000  0.035069  0.052604  0.035069  0.029225  0.005845
1    0.054233  0.018078  0.022597  0.027116  0.061012  0.022597  0.013558
2    0.069769  0.000000  0.079736  0.146183  0.029901  0.046513  0.013289
3    0.012474  0.000000  0.012474  0.000000  0.037421  0.024947  0.012474
4    0.062859  0.008381  0.037716  0.067050  0.012572  0.025144  0.025144
..        …         …         …         …         …         …         …
995  0.039218  0.008964  0.012326  0.012326  0.044820  0.031374  0.005603
996  0.046414  0.022927  0.037466  0.041940  0.032993  0.050328  0.010066
997  0.036931  0.000000  0.036931  0.100240  0.005276  0.010552  0.010552
998  0.010426  0.026064  0.026064  0.062554  0.099044  0.057341  0.026064
999  0.055525  0.007932  0.044949  0.034373  0.047593  0.037017  0.013220

          X7        X8        X9  …       X32       X33       X34  \
0    0.037992  0.043837  0.064294  …  0.006791  0.005767  0.010587
```

```
1    0.056493  0.031636  0.072311  …  0.007988  0.017462  0.034596
2    0.043190  0.083058  0.063124  … -0.001751 -0.006532 -0.007954
3    0.074842  0.049894  0.062368  … -0.006217  0.039692  0.007893
4    0.058669  0.071240  0.071240  …  0.041299  0.014669 -0.013841
..        …         …         …  …         …         …
995  0.041459  0.020169  0.063869  …  0.024335  0.029802  0.020947
996  0.036348  0.032993  0.041940  …  0.006312  0.016651  0.012115
997  0.026379  0.079137  0.047482  …  0.020791  0.023968 -0.008436
998  0.072980  0.104257  0.062554  … -0.002386  0.002360 -0.013045
999  0.066101  0.052881  0.079322  …  0.019636 -0.002085 -0.001153

          X35       X36       X37       X38       X39       X40           X41
0    0.012913  0.022654  0.000591  0.001420  0.013472  0.002129  no_efectores
1   -0.004824 -0.012705 -0.004698 -0.004303 -0.006343 -0.011193  no_efectores
2    0.014380  0.071315 -0.042175 -0.020053 -0.003677 -0.009383  no_efectores
3    0.032548  0.020587 -0.009453 -0.005468 -0.000462  0.063092  no_efectores
4    0.010196  0.009503  0.022691 -0.010394 -0.003938  0.009414  no_efectores
..        …         …         …         …         …         …           …
995  0.034460  0.023676  0.026549  0.015871  0.029773  0.030543  no_efectores
996  0.021664  0.013528  0.001051  0.007922  0.027104  0.006959  no_efectores
997  0.017480  0.044055  0.020806  0.015157  0.004487  0.013371  no_efectores
998 -0.052866 -0.023452 -0.044406 -0.021449 -0.008888 -0.017631  no_efectores
999  0.022108  0.018167  0.022240 -0.015049  0.038092  0.002244  no_efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.041543     0.016642     0.036873     0.048762     0.038477
std       0.017563     0.015653     0.021632     0.030264     0.024404
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.030626     0.007113     0.022994     0.027143     0.021989
50%       0.040138     0.013068     0.034796     0.043306     0.034626
75%       0.050890     0.020512     0.046923     0.064274     0.049404
max       0.133659     0.116200     0.309740     0.225957     0.252477

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.036938     0.016416     0.048038     0.050027     0.069013  …
std       0.015636     0.011546     0.026321     0.031035     0.031031  …
min       0.000000     0.000000     0.000000     0.000000     0.000964  …
25%       0.026609     0.008509     0.031216     0.028867     0.047151  …
50%       0.035641     0.014700     0.044595     0.043945     0.066026  …
```

```
75%       0.045375      0.022307      0.061311      0.064934      0.088380  …
max       0.140486      0.098090      0.196031      0.299456      0.208982  …

                X31           X32           X33           X34           X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012668     0.010295     0.010740     0.011909     0.009978
std       0.024718     0.025628     0.024957     0.025467     0.025491
min      -0.153838    -0.192333    -0.132338    -0.151853    -0.127015
25%       0.000702    -0.001322    -0.001545    -0.001230    -0.002047
50%       0.013247     0.012617     0.013622     0.013044     0.012847
75%       0.026550     0.025010     0.025693     0.026693     0.024385
max       0.152398     0.089910     0.100854     0.130196     0.141924

                X36           X37           X38           X39           X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009930     0.010967     0.010741     0.011208     0.010346
std       0.027182     0.028409     0.025793     0.026463     0.026744
min      -0.194998    -0.204426    -0.153696    -0.167341    -0.181414
25%      -0.002610    -0.001523    -0.002383    -0.000464    -0.001628
50%       0.013777     0.013516     0.012939     0.013802     0.012605
75%       0.024879     0.025637     0.024368     0.025131     0.025611
max       0.112856     0.232096     0.119389     0.219741     0.161977

[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	0.034968	0.005521	0.047851	0.029447	0.034968	0.069937	0.018404	
1	0.046954	0.021671	0.036118	0.036118	0.039730	0.032507	0.010836	
2	0.055956	0.018652	0.080825	0.031086	0.037304	0.012435	0.018652	
6	0.028593	0.021445	0.051825	0.035742	0.023232	0.041103	0.021445	
7	0.015291	0.003823	0.026759	0.061164	0.019114	0.034405	0.007646	
..	…	…	…	…	…	…	…	
992	0.031485	0.018891	0.018891	0.072415	0.028336	0.022039	0.009445	
993	0.040899	0.019373	0.047357	0.071036	0.034442	0.045205	0.019373	
994	0.029496	0.013110	0.042606	0.032774	0.026219	0.032774	0.016387	
995	0.046901	0.028140	0.025795	0.023450	0.086766	0.037520	0.009380	
997	0.034727	0.007311	0.036554	0.045693	0.012794	0.021933	0.025588	

	X7	X8	X9	…	X32	X33	X34	\
0	0.031288	0.062575	0.038649	…	0.017667	0.036538	0.021905	
1	0.032507	0.039730	0.101131	…	-0.007599	-0.019442	0.001068	
2	0.062173	0.055956	0.093259	…	0.035339	-0.005915	0.002631	
6	0.028593	0.026806	0.057187	…	0.010399	0.051040	-0.006440	
7	0.026759	0.064987	0.068810	…	0.012271	0.048196	0.048655	
..	…	…	…	…	…	…	…	
992	0.053524	0.066118	0.069267	…	-0.003991	-0.003269	-0.011530	
993	0.023679	0.034442	0.045205	…	0.024975	-0.011352	-0.000522	
994	0.039329	0.019664	0.068825	…	0.010913	-0.000137	0.058049	
995	0.063316	0.035175	0.105526	…	0.009670	-0.026962	0.028653	
997	0.034727	0.020105	0.065798	…	0.009581	0.020842	0.017380	

	X35	X36	X37	X38	X39	X40	X41
0	0.001887	0.007706	0.009464	-0.012130	0.006704	0.029066	efectores
1	0.036282	0.047961	0.025069	0.004237	-0.003996	0.002846	efectores
2	-0.000783	0.019987	0.031218	0.004627	0.036853	0.002744	efectores
6	0.041351	0.030654	0.015241	0.013816	0.007828	0.033001	efectores
7	0.027197	0.034987	0.016496	0.004079	0.015201	0.031129	efectores
..	…	…	…	…	…	…	
992	0.044144	0.021759	0.020352	-0.006093	0.015322	0.030938	efectores
993	0.041746	0.019671	-0.008913	-0.007482	0.034241	0.051398	efectores
994	0.019151	-0.003251	0.026962	-0.007040	0.019183	0.037739	efectores
995	-0.012449	-0.028983	0.018046	0.049319	-0.003528	-0.018205	efectores
997	0.022263	0.024037	0.021522	0.007642	0.035806	0.019015	efectores

[803 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	803.000000	803.000000	803.000000	803.000000	803.000000	803.000000	
mean	0.043611	0.014383	0.035608	0.041499	0.028101	0.034992	
std	0.015514	0.011110	0.017652	0.023241	0.016051	0.013526	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.034399	0.006548	0.023491	0.026312	0.016698	0.026016	
50%	0.043325	0.012005	0.033746	0.037163	0.026173	0.033607	
75%	0.051160	0.019352	0.044862	0.053060	0.037327	0.042927	
max	0.099873	0.060262	0.100042	0.128117	0.088114	0.087495	

	X6	X7	X8	X9	…	X31	\
count	803.000000	803.000000	803.000000	803.000000	…	803.000000	
mean	0.015498	0.037448	0.038747	0.059608	…	0.015852	
std	0.009163	0.018617	0.021617	0.025493	…	0.021731	
min	0.000000	0.000000	0.000000	0.001234	…	-0.073039	
25%	0.008960	0.024929	0.023868	0.041891	…	0.004296	
50%	0.014265	0.035264	0.035175	0.059082	…	0.016833	
75%	0.021256	0.047513	0.050942	0.075709	…	0.029247	
max	0.053398	0.104507	0.118178	0.145837	…	0.080665	

	X32	X33	X34	X35	X36	X37	\
count	803.000000	803.000000	803.000000	803.000000	803.000000	803.000000	
mean	0.016374	0.016656	0.015509	0.015526	0.015950	0.015409	
std	0.019928	0.021201	0.021621	0.021112	0.021343	0.022718	
min	-0.056736	-0.066005	-0.068520	-0.067424	-0.088024	-0.074647	
25%	0.004716	0.004827	0.003111	0.002605	0.003911	0.002797	
50%	0.017690	0.018132	0.016664	0.016811	0.017422	0.017841	
75%	0.028475	0.028999	0.028517	0.029083	0.029453	0.029727	

```
max      0.079483    0.095031    0.093520    0.087847    0.096560    0.101558

              X38         X39         X40
count  803.000000  803.000000  803.000000
mean     0.015574    0.016375    0.016153
std      0.020693    0.021079    0.020661
min     -0.080164   -0.065416   -0.067455
25%      0.004811    0.004172    0.004894
50%      0.016615    0.017824    0.016543
75%      0.028809    0.029471    0.029319
max      0.095651    0.081153    0.088628

[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    0.052604  0.000000  0.035069  0.052604  0.035069  0.029225  0.005845
1    0.054233  0.018078  0.022597  0.027116  0.061012  0.022597  0.013558
4    0.062859  0.008381  0.037716  0.067050  0.012572  0.025144  0.025144
5    0.029510  0.018971  0.037941  0.071667  0.037941  0.035833  0.023186
6    0.055108  0.023618  0.083974  0.086598  0.060357  0.049860  0.031490
..        …         …         …         …         …         …         …
993  0.066178  0.012255  0.029412  0.026961  0.056374  0.044119  0.019608
994  0.063511  0.010585  0.035605  0.032718  0.034642  0.018283  0.017321
995  0.039218  0.008964  0.012326  0.012326  0.044820  0.031374  0.005603
996  0.046414  0.022927  0.037466  0.041940  0.032993  0.050328  0.010066
999  0.055525  0.007932  0.044949  0.034373  0.047593  0.037017  0.013220

           X7        X8        X9  …       X32       X33       X34  \
0    0.037992  0.043837  0.064294  …  0.006791  0.005767  0.010587
1    0.056493  0.031636  0.072311  …  0.007988  0.017462  0.034596
4    0.058669  0.071240  0.071240  …  0.041299  0.014669 -0.013841
5    0.107500  0.069559  0.077990  … -0.000436 -0.019411  0.014834
6    0.039363  0.099719  0.060357  …  0.043912  0.011981  0.004420
..        …         …         …  …       …         …         …
993  0.049021  0.017157  0.078433  … -0.012557 -0.027277 -0.003723
994  0.040416  0.031755  0.090455  …  0.023913  0.013384 -0.000118
995  0.041459  0.020169  0.063869  …  0.024335  0.029802  0.020947
996  0.036348  0.032993  0.041940  …  0.006312  0.016651  0.012115
999  0.066101  0.052881  0.079322  …  0.019636 -0.002085 -0.001153

          X35       X36       X37       X38       X39       X40           X41
0    0.012913  0.022654  0.000591  0.001420  0.013472  0.002129  no_efectores
```

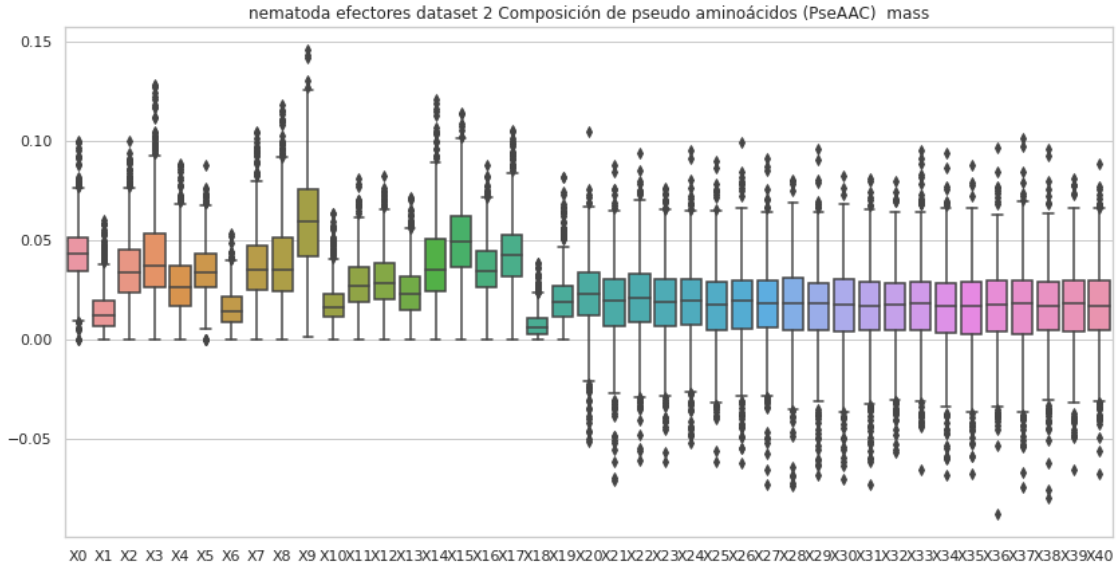
```
1   -0.004824 -0.012705 -0.004698 -0.004303 -0.006343 -0.011193  no_efectores
4    0.010196  0.009503  0.022691 -0.010394 -0.003938  0.009414  no_efectores
5    0.000753 -0.001944  0.006971 -0.016315 -0.014881  0.004576  no_efectores
6   -0.007113 -0.013685 -0.002133 -0.031581  0.036043  0.001077  no_efectores
..        …        …        …        …        …         …
993 -0.001284  0.037202  0.012814  0.036543  0.019004  0.019047  no_efectores
994  0.012953  0.009527  0.012447  0.008901  0.017740  0.014356  no_efectores
995  0.034460  0.023676  0.026549  0.015871  0.029773  0.030543  no_efectores
996  0.021664  0.013528  0.001051  0.007922  0.027104  0.006959  no_efectores
999  0.022108  0.018167  0.022240 -0.015049  0.038092  0.002244  no_efectores

[806 rows x 42 columns]
```

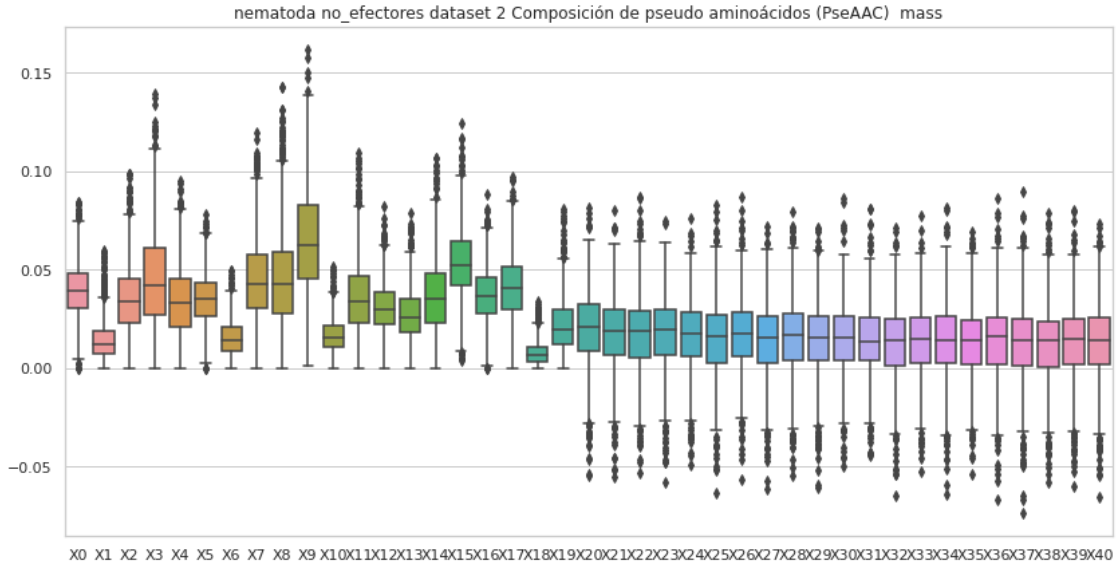
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  806.000000  806.000000  806.000000  806.000000  806.000000  806.000000
mean     0.039880    0.014464    0.035211    0.045800    0.034947    0.035414
std      0.014345    0.010680    0.017180    0.025283    0.018278    0.012587
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.030708    0.007251    0.023227    0.026999    0.021308    0.026646
50%      0.039612    0.012347    0.034192    0.041746    0.033073    0.034974
75%      0.048479    0.018784    0.045330    0.060961    0.045390    0.043420
max      0.084098    0.060058    0.098776    0.139212    0.095096    0.077897

                X6          X7          X8          X9  …         X31  \
count  806.000000  806.000000  806.000000  806.000000  …  806.000000
mean     0.015479    0.045269    0.046329    0.064905  …    0.014073
std      0.009123    0.021993    0.026025    0.027524  …    0.018323
min      0.000000    0.000000    0.000000    0.000964  …   -0.044213
25%      0.008584    0.030575    0.027844    0.045649  …    0.004033
50%      0.014536    0.043039    0.042522    0.062499  …    0.013555
75%      0.021224    0.057642    0.059135    0.082854  …    0.025834
max      0.049675    0.119433    0.142965    0.161669  …    0.080981

               X32         X33         X34         X35         X36         X37  \
count  806.000000  806.000000  806.000000  806.000000  806.000000  806.000000
mean     0.013143    0.013797    0.013698    0.013147    0.013711    0.013121
std      0.018948    0.018005    0.019803    0.018238    0.019491    0.020190
min     -0.064701   -0.052649   -0.064194   -0.053466   -0.066823   -0.073502
25%      0.001621    0.002960    0.002421    0.002251    0.001663    0.001349
50%      0.014275    0.014602    0.014070    0.014460    0.016141    0.013941
75%      0.025095    0.025943    0.026540    0.024597    0.025535    0.025354
max      0.071106    0.077518    0.081365    0.069184    0.086368    0.089524
```

	X38	X39	X40
count	806.000000	806.000000	806.000000
mean	0.012509	0.013192	0.013224
std	0.019491	0.019120	0.020157
min	-0.057552	-0.059727	-0.065545
25%	0.000879	0.002168	0.001803
50%	0.014083	0.015038	0.013861
75%	0.023879	0.025051	0.025868
max	0.078561	0.079993	0.073004

[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	0.032597	0.005147	0.044607	0.027450	0.032597	0.065194	0.017156	
1	0.089911	0.041497	0.069162	0.069162	0.076079	0.062246	0.020749	
2	0.031414	0.010471	0.045375	0.017452	0.020942	0.006981	0.010471	
3	0.030288	0.001893	0.011358	0.037860	0.015144	0.017037	0.041646	
4	0.019137	0.095687	0.051671	0.044016	0.026792	0.068895	0.024879	
..	…	…	…	…	…	…	…	
995	0.025439	0.015263	0.013991	0.012719	0.047061	0.020351	0.005088	
996	0.057677	0.069212	0.000000	0.046142	0.034606	0.000000	0.023071	
997	0.053002	0.011158	0.055792	0.069740	0.019527	0.033475	0.039054	
998	0.053456	0.040092	0.066820	0.053456	0.093548	0.066820	0.026728	
999	0.039614	0.007923	0.063383	0.039614	0.015846	0.031691	0.023768	

	X7	X8	X9	…	X53	X54	X55	\
0	0.029166	0.058332	0.036028	…	0.015041	-0.015121	0.018173	
1	0.062246	0.076079	0.193655	…	0.016344	0.041742	0.038650	
2	0.034904	0.031414	0.052356	…	0.028294	-0.004635	-0.017821	
3	0.018930	0.041646	0.017037	…	0.012452	0.031049	0.049915	
4	0.034447	0.038275	0.042102	…	0.041735	0.002837	0.023129	
..	…	…	…	…	…	…	…	
995	0.034342	0.019079	0.057237	…	0.009372	0.004757	-0.007176	
996	0.034606	0.115354	0.092283	…	0.081257	0.030577	0.046216	
997	0.053002	0.030686	0.100426	…	-0.022896	-0.001452	0.029888	
998	0.013364	0.147004	0.093548	…	-0.071292	-0.066734	-0.029161	
999	0.047537	0.023768	0.079228	…	0.019168	0.062749	0.081396	

	X56	X57	X58	X59	X60	X61	X62
0	0.007712	0.042952	-0.012209	0.005577	0.033664	0.029156	efectores
1	-0.031054	-0.043378	-0.021673	-0.009469	0.036645	0.062671	efectores
2	-0.005896	0.011794	0.007205	0.022808	0.015415	0.016449	efectores
3	-0.016823	-0.010090	-0.007617	0.021073	-0.012891	0.025107	efectores
4	-0.002977	-0.012792	0.000819	0.013507	-0.025935	0.014551	efectores
..	…	…	…	…	…	…	…
995	-0.018182	-0.003697	-0.021056	-0.022500	-0.005052	-0.012921	efectores
996	0.052000	0.075546	0.063505	0.104039	-0.026000	0.007843	efectores
997	-0.040529	0.001916	-0.034539	-0.006461	-0.013133	0.001373	efectores

```
998 -0.029554  0.066362 -0.056898 -0.133073  0.069141 -0.055540  efectores
999 -0.011545  0.050684  0.075512  0.058070  0.042285 -0.028406  efectores

[1000 rows x 63 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.051290	0.018122	0.040049	0.045938	0.033307	
std	0.032355	0.020059	0.024549	0.026314	0.025239	
min	-0.080007	-0.020002	-0.040003	-0.040003	-0.140012	
25%	0.029727	0.006514	0.023030	0.027857	0.017260	
50%	0.045240	0.013603	0.037384	0.043587	0.029235	
75%	0.066724	0.023442	0.052652	0.059323	0.044181	
max	0.291526	0.189864	0.273794	0.237330	0.199380	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.044003	0.018972	0.043206	0.043268	0.068331	…	
std	0.032308	0.015263	0.028888	0.027894	0.040321	…	
min	-0.040003	-0.060005	-0.080007	-0.120010	-0.180015	…	
25%	0.022107	0.008715	0.024490	0.024754	0.041855	…	
50%	0.036031	0.016104	0.038563	0.039298	0.062647	…	
75%	0.058522	0.025001	0.057143	0.056603	0.089884	…	
max	0.253617	0.133840	0.312908	0.273779	0.379728	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.001530	0.008879	-0.000946	0.006612	0.000931	
std	0.044251	0.034823	0.044506	0.039769	0.045057	
min	-0.301558	-0.201509	-0.285597	-0.367835	-0.354983	
25%	-0.015146	-0.005432	-0.019965	-0.008833	-0.015977	
50%	0.003653	0.010780	0.003805	0.008980	0.003606	
75%	0.020883	0.026181	0.018923	0.024364	0.019845	
max	0.308987	0.201475	0.295844	0.244221	0.357249	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007742	0.002125	0.008036	-0.000118	0.006395
std	0.035658	0.046769	0.036761	0.048091	0.038164
min	-0.247673	-0.395738	-0.224130	-0.563180	-0.364811
25%	-0.005863	-0.015971	-0.006798	-0.016734	-0.007841
50%	0.010247	0.004248	0.010230	0.003905	0.009827
75%	0.025331	0.023509	0.025068	0.020267	0.025109

```
max       0.178969     0.344608     0.279135     0.191482     0.160184

[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	0.041626	0.000000	0.027750	0.041626	0.027750	0.023125	0.004625	
1	0.031324	0.010441	0.013052	0.015662	0.035240	0.013052	0.007831	
2	0.013198	0.000000	0.015083	0.027653	0.005656	0.008799	0.002514	
3	0.019743	0.000000	0.019743	0.000000	0.059229	0.039486	0.019743	
4	0.068618	0.009149	0.041171	0.073193	0.013724	0.027447	0.027447	
..	…	…	…	…	…	…	…	
995	0.039330	0.008990	0.012361	0.012361	0.044949	0.031464	0.005619	
996	0.052262	0.025816	0.042187	0.047225	0.037150	0.056670	0.011334	
997	0.003605	0.000000	0.003605	0.009784	0.000515	0.001030	0.001030	
998	0.006266	0.015665	0.015665	0.037596	0.059528	0.034463	0.015665	
999	0.039884	0.005698	0.032287	0.024690	0.034186	0.026589	0.009496	

	X7	X8	X9	…	X53	X54	X55	\
0	0.030063	0.034688	0.050876	…	0.022245	0.025260	0.033361	
1	0.032630	0.018273	0.041766	…	0.013728	-0.010196	-0.008050	
2	0.008170	0.015712	0.011941	…	0.023075	0.013038	0.022434	
3	0.118458	0.078972	0.098715	…	-0.066696	-0.042146	-0.060039	
4	0.064044	0.077767	0.077767	…	0.069922	-0.050236	-0.012273	
..	…	…	…	…	…	…	…	
995	0.041578	0.020227	0.064052	…	-0.016302	0.003059	-0.009424	
996	0.040928	0.037150	0.047225	…	0.010383	-0.016252	0.011089	
997	0.002575	0.007724	0.004635	…	0.028297	0.024489	0.022757	
998	0.043863	0.062661	0.037596	…	0.012049	-0.004081	0.013863	
999	0.047481	0.037984	0.056977	…	0.015687	0.058632	0.044052	

	X56	X57	X58	X59	X60	X61	X62
0	-0.013264	0.000687	-0.008254	-0.012050	-0.043515	-0.006965	no_efectores
1	0.014243	0.012871	0.017515	0.014541	0.003011	-0.003131	no_efectores
2	-0.002414	0.017736	0.009964	0.027536	0.006622	0.033278	no_efectores
3	-0.064970	-0.012150	-0.040405	-0.058155	0.103436	0.134008	no_efectores
4	-0.029957	-0.028196	0.007654	0.024641	0.015151	0.063894	no_efectores
..	…	…	…	…	…	…	…
995	0.020411	0.009775	0.010023	0.002316	0.024780	-0.000100	no_efectores
996	-0.009838	0.004972	-0.008853	0.019937	-0.009342	0.012650	no_efectores
997	0.025543	0.023723	0.018052	0.017557	0.018864	0.016491	no_efectores
998	-0.006017	-0.006336	-0.030467	-0.025419	0.057478	0.039989	no_efectores

```
999 -0.011310 -0.012500  0.013844  0.004534  0.010546 -0.002084  no_efectores

[1000 rows x 63 columns]
```

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

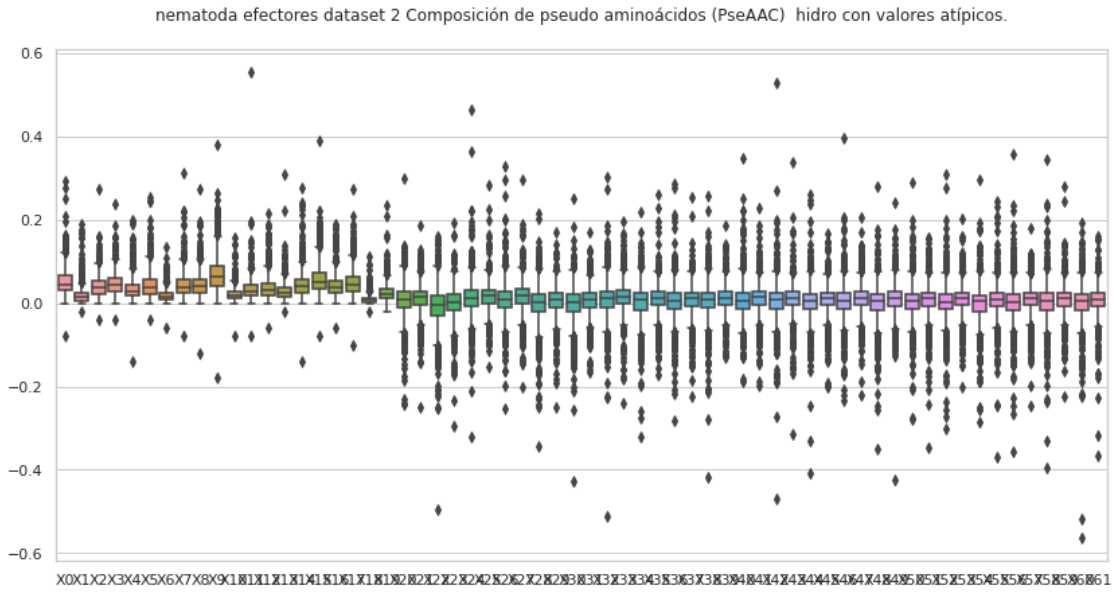
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.041052	0.016468	0.034685	0.042242	0.035030	
std	0.028083	0.019250	0.023901	0.026121	0.023889	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.022227	0.005849	0.017129	0.024690	0.018406	
50%	0.034634	0.011675	0.030644	0.038752	0.030907	
75%	0.052972	0.020271	0.047630	0.055443	0.046217	
max	0.273017	0.234514	0.230205	0.272060	0.211516	

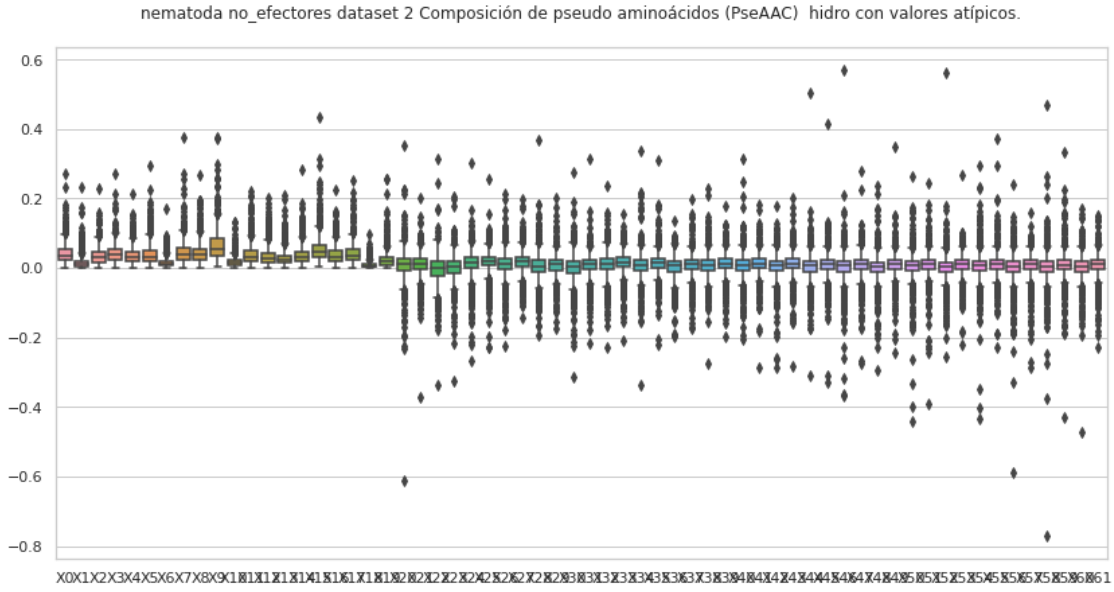
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.038467	0.015857	0.044367	0.043779	0.064654	…	
std	0.029479	0.014089	0.031493	0.029237	0.041604	…	
min	0.000000	0.000000	0.000000	0.000000	0.002380	…	
25%	0.018644	0.006321	0.024131	0.024024	0.036942	…	
50%	0.030956	0.013071	0.038935	0.038388	0.056127	…	
75%	0.050724	0.020908	0.057759	0.055836	0.084301	…	
max	0.296122	0.169213	0.374520	0.265993	0.374520	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.002259	0.008836	0.004104	0.009767	0.000628	
std	0.039761	0.030805	0.043178	0.035634	0.042837	
min	-0.256943	-0.219811	-0.431557	-0.228718	-0.586456	
25%	-0.010292	-0.002159	-0.009248	-0.003224	-0.010678	
50%	0.005923	0.012034	0.006736	0.010962	0.005277	
75%	0.017347	0.024418	0.019768	0.023675	0.018647	
max	0.563627	0.269306	0.294051	0.370464	0.239357	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007436	0.001967	0.008392	0.003086	0.008629
std	0.034197	0.049092	0.037516	0.037204	0.033606
min	-0.287984	-0.770211	-0.430629	-0.473937	-0.230011
25%	-0.003364	-0.010311	-0.003378	-0.009686	-0.003063
50%	0.011011	0.005156	0.009388	0.005735	0.010020
75%	0.023528	0.018727	0.023590	0.019507	0.024341
max	0.157638	0.470547	0.334039	0.172486	0.151819

[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	0.032597	0.005147	0.044607	0.027450	0.032597	0.065194	0.017156	
2	0.031414	0.010471	0.045375	0.017452	0.020942	0.006981	0.010471	
3	0.030288	0.001893	0.011358	0.037860	0.015144	0.017037	0.041646	
6	0.037545	0.028159	0.068050	0.046931	0.030505	0.053971	0.028159	
7	0.017434	0.004358	0.030509	0.069735	0.021792	0.039226	0.008717	
..	…	…	…	…	…	…	…	
993	0.019161	0.009076	0.022187	0.033280	0.016136	0.021178	0.009076	
994	0.042227	0.018768	0.060995	0.046919	0.037535	0.046919	0.023459	
995	0.025439	0.015263	0.013991	0.012719	0.047061	0.020351	0.005088	
997	0.053002	0.011158	0.055792	0.069740	0.019527	0.033475	0.039054	
999	0.039614	0.007923	0.063383	0.039614	0.015846	0.031691	0.023768	

	X7	X8	X9	…	X53	X54	X55	\
0	0.029166	0.058332	0.036028	…	0.015041	-0.015121	0.018173	
2	0.034904	0.031414	0.052356	…	0.028294	-0.004635	-0.017821	
3	0.018930	0.041646	0.017037	…	0.012452	0.031049	0.049915	
6	0.037545	0.035198	0.075090	…	0.046271	-0.001208	0.024143	
7	0.030509	0.074094	0.078452	…	0.053911	0.013845	0.049274	
..	…	…	…	…	…	…	…	
993	0.011093	0.016136	0.021178	…	0.026659	0.029323	0.031590	
994	0.056303	0.028151	0.098530	…	-0.009489	0.032684	-0.006578	
995	0.034342	0.019079	0.057237	…	0.009372	0.004757	-0.007176	
997	0.053002	0.030686	0.100426	…	-0.022896	-0.001452	0.029888	
999	0.047537	0.023768	0.079228	…	0.019168	0.062749	0.081396	

	X56	X57	X58	X59	X60	X61	X62
0	0.007712	0.042952	-0.012209	0.005577	0.033664	0.029156	efectores
2	-0.005896	0.011794	0.007205	0.022808	0.015415	0.016449	efectores
3	-0.016823	-0.010090	-0.007617	0.021073	-0.012891	0.025107	efectores
6	-0.009513	0.012506	-0.025881	-0.015210	0.027928	0.023685	efectores
7	-0.015917	0.000940	-0.011305	0.032641	-0.024746	0.015459	efectores
..	…	…	…	…	…	…	…
993	0.005378	0.025557	-0.002586	0.023199	0.022925	0.029583	efectores
994	-0.088000	-0.043895	-0.017366	-0.010526	0.047918	0.033579	efectores
995	-0.018182	-0.003697	-0.021056	-0.022500	-0.005052	-0.012921	efectores
997	-0.040529	0.001916	-0.034539	-0.006461	-0.013133	0.001373	efectores
999	-0.011545	0.050684	0.075512	0.058070	0.042285	-0.028406	efectores

[806 rows x 63 columns]

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

	X0	X1	X2	X3	X4	X5	\
count	806.000000	806.000000	806.000000	806.000000	806.000000	806.000000	
mean	0.045844	0.015196	0.036683	0.041659	0.028923	0.037784	
std	0.025085	0.012740	0.019583	0.020453	0.017291	0.024069	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.027885	0.006223	0.022266	0.026738	0.016315	0.020360	
50%	0.042207	0.012320	0.035743	0.040409	0.026952	0.032275	
75%	0.059963	0.021080	0.049274	0.054887	0.039163	0.050579	
max	0.146969	0.075675	0.104143	0.114090	0.090633	0.128748	

	X6	X7	X8	X9	…	X52	\
count	806.000000	806.000000	806.000000	806.000000	…	806.000000	
mean	0.016460	0.038455	0.038393	0.060958	…	0.004576	
std	0.011263	0.020955	0.020323	0.030126	…	0.029444	
min	0.000000	0.000000	0.000000	0.001090	…	-0.122997	
25%	0.007983	0.022706	0.023287	0.040048	…	-0.010082	
50%	0.014438	0.035688	0.036002	0.058501	…	0.005546	
75%	0.022486	0.051340	0.051949	0.081896	…	0.020726	
max	0.063754	0.124980	0.111730	0.178697	…	0.116559	

	X53	X54	X55	X56	X57	X58	\
count	806.000000	806.000000	806.000000	806.000000	806.000000	806.000000	
mean	0.010742	0.002035	0.008256	0.003330	0.010934	0.004186	
std	0.024825	0.029319	0.025356	0.027785	0.024574	0.029183	
min	-0.094967	-0.131205	-0.095303	-0.095808	-0.092531	-0.115004	
25%	-0.002735	-0.014906	-0.006012	-0.011882	-0.002642	-0.011573	
50%	0.011089	0.005368	0.009785	0.004872	0.011379	0.005101	
75%	0.025941	0.017972	0.023809	0.019243	0.025113	0.021464	
max	0.104359	0.114211	0.122827	0.097255	0.109836	0.122443	

	X59	X60	X61
count	806.000000	806.000000	806.000000
mean	0.010320	0.002873	0.008735
std	0.025043	0.029780	0.025648
min	-0.088103	-0.138182	-0.098655
25%	-0.003705	-0.012839	-0.005565
50%	0.010750	0.004139	0.010067
75%	0.024532	0.018314	0.024279
max	0.095917	0.118610	0.098976

[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    0.041626  0.000000  0.027750  0.041626  0.027750  0.023125  0.004625
1    0.031324  0.010441  0.013052  0.015662  0.035240  0.013052  0.007831
2    0.013198  0.000000  0.015083  0.027653  0.005656  0.008799  0.002514
4    0.068618  0.009149  0.041171  0.073193  0.013724  0.027447  0.027447
5    0.031687  0.020370  0.040741  0.076955  0.040741  0.038477  0.024897
..        …         …         …         …         …         …         …
995  0.039330  0.008990  0.012361  0.012361  0.044949  0.031464  0.005619
996  0.052262  0.025816  0.042187  0.047225  0.037150  0.056670  0.011334
997  0.003605  0.000000  0.003605  0.009784  0.000515  0.001030  0.001030
998  0.006266  0.015665  0.015665  0.037596  0.059528  0.034463  0.015665
999  0.039884  0.005698  0.032287  0.024690  0.034186  0.026589  0.009496

           X7        X8        X9  …       X53       X54       X55  \
0    0.030063  0.034688  0.050876  …  0.022245  0.025260  0.033361
1    0.032630  0.018273  0.041766  …  0.013728 -0.010196 -0.008050
2    0.008170  0.015712  0.011941  …  0.023075  0.013038  0.022434
4    0.064044  0.077767  0.077767  …  0.069922 -0.050236 -0.012273
5    0.115432  0.074691  0.083745  … -0.002146  0.052141  0.017404
..        …         …         …  …         …         …         …
995  0.041578  0.020227  0.064052  … -0.016302  0.003059 -0.009424
996  0.040928  0.037150  0.047225  …  0.010383 -0.016252  0.011089
997  0.002575  0.007724  0.004635  …  0.028297  0.024489  0.022757
998  0.043863  0.062661  0.037596  …  0.012049 -0.004081  0.013863
999  0.047481  0.037984  0.056977  …  0.015687  0.058632  0.044052

          X56       X57       X58       X59       X60       X61           X62
0   -0.013264  0.000687 -0.008254 -0.012050 -0.043515 -0.006965  no_efectores
1    0.014243  0.012871  0.017515  0.014541  0.003011 -0.003131  no_efectores
2   -0.002414  0.017736  0.009964  0.027536  0.006622  0.033278  no_efectores
4   -0.029957 -0.028196  0.007654  0.024641  0.015151  0.063894  no_efectores
5    0.050653  0.045804 -0.040970 -0.085752 -0.043811 -0.027846  no_efectores
..        …         …         …         …         …         …           …
995  0.020411  0.009775  0.010023  0.002316  0.024780 -0.000100  no_efectores
996 -0.009838  0.004972 -0.008853  0.019937 -0.009342  0.012650  no_efectores
997  0.025543  0.023723  0.018052  0.017557  0.018864  0.016491  no_efectores
998 -0.006017 -0.006336 -0.030467 -0.025419  0.057478  0.039989  no_efectores
999 -0.011310 -0.012500  0.013844  0.004534  0.010546 -0.002084  no_efectores

[820 rows x 63 columns]
```

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

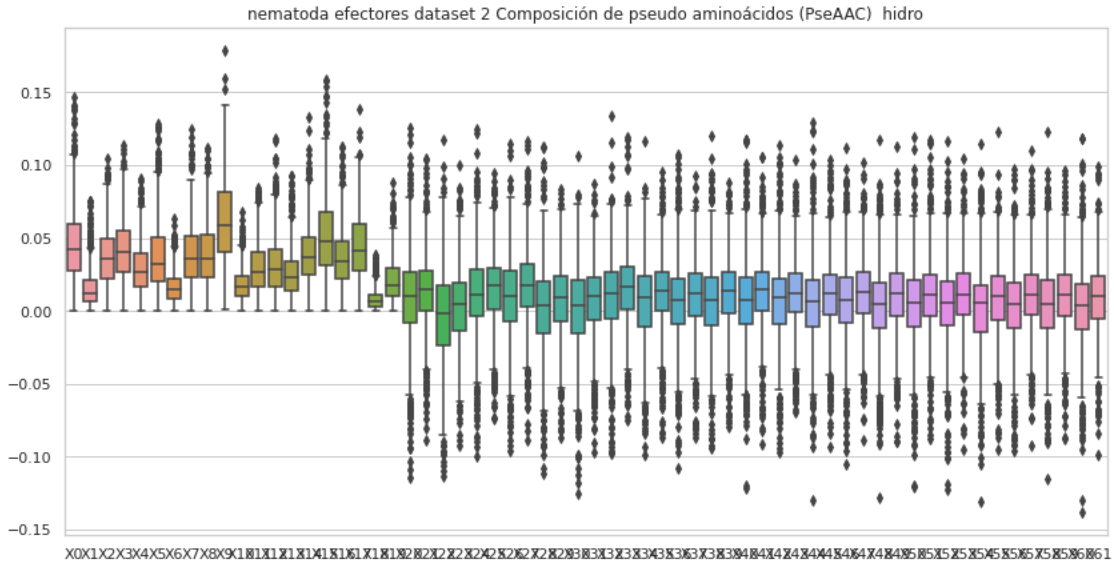
	X0	X1	X2	X3	X4	X5	\
count	820.000000	820.000000	820.000000	820.000000	820.000000	820.000000	
mean	0.035894	0.013033	0.030810	0.038015	0.030461	0.032352	
std	0.020369	0.011233	0.018684	0.019537	0.017575	0.020915	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.020862	0.005312	0.016095	0.023916	0.017487	0.017140	
50%	0.032201	0.010117	0.027519	0.036303	0.028026	0.026539	
75%	0.047050	0.017643	0.042962	0.051313	0.040569	0.044247	
max	0.124441	0.070295	0.094132	0.097465	0.098191	0.125197	

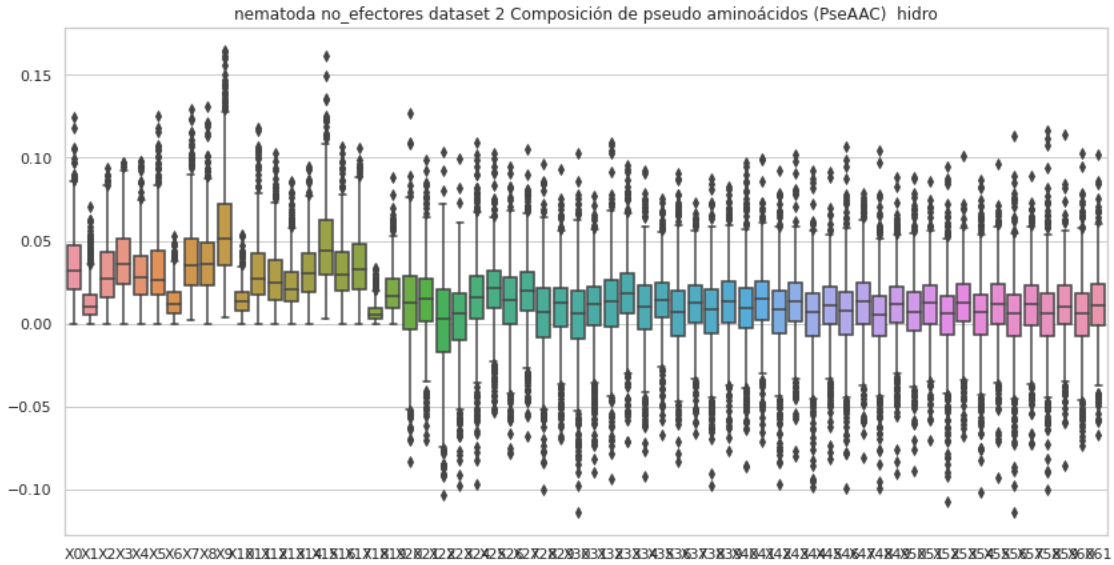
	X6	X7	X8	X9	…	X52	\
count	820.000000	820.000000	820.000000	820.000000	…	820.000000	
mean	0.013426	0.038872	0.038147	0.056006	…	0.004248	
std	0.009485	0.021808	0.020509	0.029565	…	0.023753	
min	0.000000	0.002500	0.000000	0.003643	…	-0.106844	
25%	0.005951	0.023144	0.022797	0.035093	…	-0.006866	
50%	0.011826	0.035288	0.036023	0.050962	…	0.006371	
75%	0.018894	0.051074	0.049114	0.072415	…	0.016820	
max	0.052714	0.129145	0.130921	0.164632	…	0.094890	

	X53	X54	X55	X56	X57	X58	\
count	820.000000	820.000000	820.000000	820.000000	820.000000	820.000000	
mean	0.011711	0.005438	0.010953	0.004159	0.010644	0.005103	
std	0.020178	0.023329	0.020715	0.023692	0.020997	0.024424	
min	-0.068009	-0.101536	-0.092369	-0.113220	-0.072782	-0.099635	
25%	0.001011	-0.006911	-0.000041	-0.007766	-0.000814	-0.007752	
50%	0.013028	0.007357	0.011884	0.006237	0.012260	0.005910	
75%	0.024299	0.017768	0.023526	0.017340	0.023117	0.017973	
max	0.100835	0.086785	0.096025	0.113245	0.088896	0.116743	

	X59	X60	X61
count	820.000000	820.000000	820.000000
mean	0.010638	0.005640	0.011391
std	0.020175	0.023583	0.020596
min	-0.085752	-0.073614	-0.067031
25%	-0.000376	-0.007512	-0.000872
50%	0.010401	0.006477	0.010897
75%	0.023407	0.018619	0.023748
max	0.114417	0.102432	0.101614

[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    0.083639 -0.005352  0.035217 -0.075515 -0.049660  0.002040  0.051325
1    0.076786 -0.015779 -0.019657  0.071822  0.041807 -0.014048 -0.041745
2    0.030428 -0.026574 -0.102359 -0.051958  0.007096  0.010145 -0.005425
3   -0.035800 -0.025878  0.111503  0.002468 -0.023529  0.052107 -0.061651
4   -0.088596 -0.097694  0.062204 -0.011290  0.006157 -0.001605 -0.037680
..        …         …         …         …         …         …         …
995  0.021342 -0.076007  0.080615  0.104933 -0.014077  0.098101  0.036762
996  0.013499 -0.020290 -0.060178 -0.014671 -0.067453 -0.113696  0.022916
997  0.000233 -0.013153  0.008006 -0.010721 -0.039223  0.050749 -0.013885
998 -0.095723  0.046219 -0.070288  0.002326 -0.148286 -0.052007 -0.105643
999  0.011385 -0.031704  0.026109 -0.177202 -0.007913 -0.085267  0.070085

           X7        X8        X9       X10       X11       X12        X13
0   -0.000714 -0.032872  0.046546 -0.015022 -0.045487 -0.014852  efectores
```

```
1   -0.011559  0.007187 -0.056380 -0.052853  0.037983 -0.064151  efectores
2   -0.045543 -0.053033  0.014090  0.027548  0.104552  0.062548  efectores
3   -0.123213 -0.091409  0.005354 -0.088815 -0.152576  0.083276  efectores
4    0.036548 -0.120872 -0.047274  0.080199 -0.077982  0.050726  efectores
..        …        …        …        …        …        …
995 -0.065346 -0.024897 -0.121400 -0.034164  0.081307  0.016672  efectores
996  0.007105 -0.072166 -0.101698 -0.071585  0.036963  0.113315  efectores
997  0.088244 -0.039619  0.009982 -0.026287  0.064800 -0.031305  efectores
998 -0.115396  0.076165  0.007793  0.040005  0.140017 -0.048874  efectores
999  0.063389 -0.130041  0.070803 -0.154984  0.011004 -0.079799  efectores

[1000 rows x 14 columns]
```

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.016487	0.004659	0.015174	0.005992	0.006699	
std	0.084053	0.071220	0.086237	0.069075	0.074446	
min	-1.260053	-0.302711	-0.360285	-0.293564	-0.335849	
25%	-0.020759	-0.032896	-0.026970	-0.026811	-0.033250	
50%	0.017334	0.002567	0.010657	0.006724	0.007402	
75%	0.056541	0.043087	0.050794	0.043054	0.044136	
max	0.331436	0.437983	0.660108	0.321100	0.611287	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.005285	0.004640	0.001927	0.005201	0.000758	
std	0.082888	0.073427	0.073150	0.082656	0.070704	
min	-0.721238	-0.292843	-0.782818	-0.385898	-0.366903	
25%	-0.037178	-0.033149	-0.036654	-0.038345	-0.036821	
50%	0.001097	0.003479	0.005276	0.003017	0.000869	
75%	0.040256	0.040196	0.042391	0.043578	0.037919	
max	0.563584	0.707573	0.342486	0.596790	0.462342	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.001482	0.008344	-0.000646
std	0.075886	0.083553	0.070102
min	-0.382596	-0.354719	-0.416622
25%	-0.037434	-0.035257	-0.037491
50%	0.002789	0.003412	0.002387
75%	0.042596	0.042862	0.037316
max	0.503795	0.506667	0.236848

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	-0.005499	0.076731	0.082069	0.042332	-0.025083	-0.009055	0.016039	
1	0.047362	-0.016310	0.088810	-0.012256	0.054297	0.061081	0.032326	
2	0.017583	0.020081	0.010700	-0.003859	-0.030420	-0.011289	-0.044492	
3	0.020465	0.056849	-0.130197	-0.148779	-0.042863	0.032723	0.024572	
4	-0.034586	0.037085	-0.053004	0.051479	0.036376	-0.015853	-0.026307	
..	…	…	…	…	…	…	…	
995	0.061555	0.044298	-0.030270	0.086378	0.014327	-0.031923	0.009457	
996	0.022294	0.018190	-0.000886	0.047173	0.036217	0.029687	0.041919	
997	0.154071	0.044966	0.073540	0.061560	-0.006831	-0.010841	-0.077733	
998	0.088668	-0.006023	-0.011232	0.028521	-0.013962	-0.026034	0.051717	
999	-0.032358	0.085311	-0.076869	0.033565	-0.095810	-0.010500	0.052686	

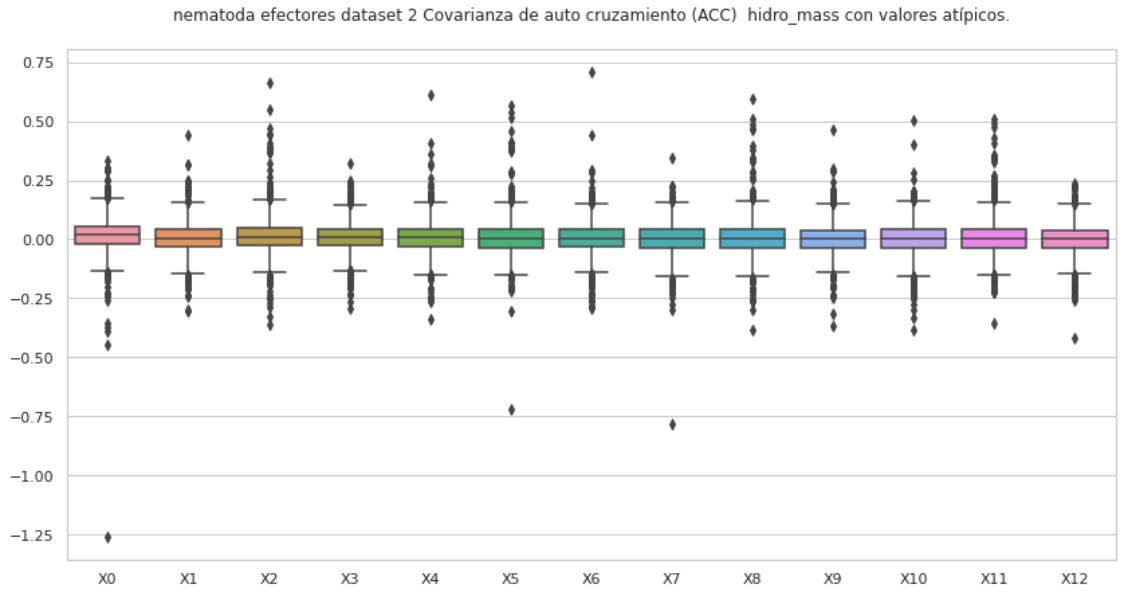
	X7	X8	X9	X10	X11	X12	X13
0	0.023109	0.028802	-0.041983	-0.044756	-0.021222	-0.021476	no_efectores
1	0.056652	0.155209	-0.040102	-0.015698	0.019768	0.003394	no_efectores
2	-0.026387	-0.013318	0.004201	-0.001084	0.027858	-0.007239	no_efectores
3	0.079469	-0.083367	-0.036984	-0.059086	-0.012986	-0.046602	no_efectores
4	-0.006043	0.020451	0.014907	0.042281	-0.020481	0.077228	no_efectores
..	…	…	…	…	…	…	
995	-0.006285	0.058376	-0.014329	0.046068	0.025824	0.026863	no_efectores
996	0.043219	0.020760	-0.000530	-0.003995	-0.031675	-0.029359	no_efectores
997	-0.086480	-0.037878	0.013159	0.005679	-0.011079	0.045636	no_efectores
998	-0.014522	-0.062910	-0.037692	0.003649	0.066375	-0.010581	no_efectores
999	-0.036550	0.011394	-0.054323	0.048895	-0.042793	0.026262	no_efectores

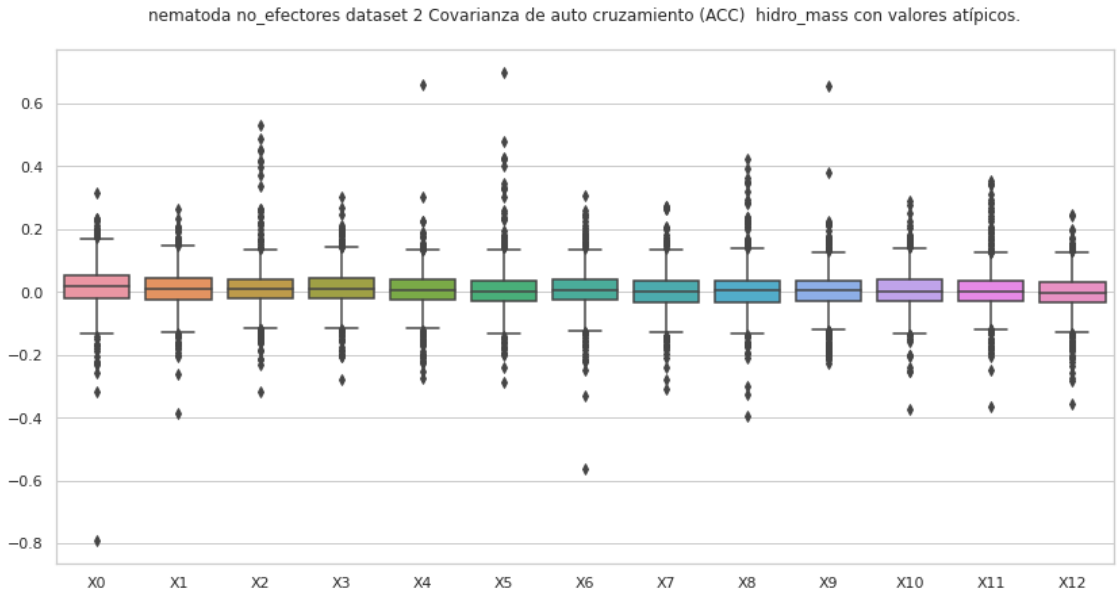
[1000 rows x 14 columns]

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

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.016331	0.009608	0.013093	0.012879	0.007002	
std	0.071328	0.061263	0.072036	0.060608	0.063184	
min	-0.791093	-0.388256	-0.315379	-0.278335	-0.275304	
25%	-0.021346	-0.024410	-0.021331	-0.019319	-0.025158	
50%	0.016799	0.012000	0.010348	0.011663	0.007424	
75%	0.054965	0.045019	0.041927	0.045171	0.038855	

```
max       0.316066     0.262523     0.531280     0.301710     0.658696

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006226     0.005860     0.002009     0.007046     0.004325
std       0.075050     0.064361     0.060462     0.070106     0.066063
min      -0.285741    -0.564501    -0.307445    -0.392712    -0.225464
25%      -0.030129    -0.025377    -0.031967    -0.031517    -0.027230
50%       0.001062     0.007349     0.002102     0.004259     0.005088
75%       0.038101     0.039334     0.036117     0.037396     0.035347
max       0.698981     0.305243     0.273823     0.423853     0.655517

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005102     0.005749    -0.001382
std       0.061155     0.066160     0.061773
min      -0.373780    -0.365769    -0.354705
25%      -0.029701    -0.027269    -0.031834
50%       0.003492     0.001701    -0.002034
75%       0.038513     0.033965     0.032423
max       0.290336     0.353955     0.247970
```





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  \
0    0.083639 -0.005352  0.035217 -0.075515 -0.049660  0.002040  0.051325
1    0.076786 -0.015779 -0.019657  0.071822  0.041807 -0.014048 -0.041745
2    0.030428 -0.026574 -0.102359 -0.051958  0.007096  0.010145 -0.005425
3   -0.035800 -0.025878  0.111503  0.002468 -0.023529  0.052107 -0.061651
4   -0.088596 -0.097694  0.062204 -0.011290  0.006157 -0.001605 -0.037680
..        …         …         …         …         …         …         …
995  0.021342 -0.076007  0.080615  0.104933 -0.014077  0.098101  0.036762
996  0.013499 -0.020290 -0.060178 -0.014671 -0.067453 -0.113696  0.022916
997  0.000233 -0.013153  0.008006 -0.010721 -0.039223  0.050749 -0.013885
998 -0.095723  0.046219 -0.070288  0.002326 -0.148286 -0.052007 -0.105643
999  0.011385 -0.031704  0.026109 -0.177202 -0.007913 -0.085267  0.070085

           X7        X8        X9       X10       X11       X12        X13
0   -0.000714 -0.032872  0.046546 -0.015022 -0.045487 -0.014852  efectores
1   -0.011559  0.007187 -0.056380 -0.052853  0.037983 -0.064151  efectores
2   -0.045543 -0.053033  0.014090  0.027548  0.104552  0.062548  efectores
3   -0.123213 -0.091409  0.005354 -0.088815 -0.152576  0.083276  efectores
4    0.036548 -0.120872 -0.047274  0.080199 -0.077982  0.050726  efectores
..        …         …         …         …         …         …         …
995 -0.065346 -0.024897 -0.121400 -0.034164  0.081307  0.016672  efectores
```

```
996  0.007105 -0.072166 -0.101698 -0.071585  0.036963  0.113315  efectores
997  0.088244 -0.039619  0.009982 -0.026287  0.064800 -0.031305  efectores
998 -0.115396  0.076165  0.007793  0.040005  0.140017 -0.048874  efectores
999  0.063389 -0.130041  0.070803 -0.154984  0.011004 -0.079799  efectores

[900 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	900.000000	900.000000	900.000000	900.000000	900.000000	900.000000
mean	0.019054	0.004803	0.010884	0.006660	0.006741	0.000739
std	0.060270	0.060343	0.062920	0.059132	0.059738	0.058267
min	-0.222268	-0.204401	-0.196184	-0.186019	-0.208635	-0.212994
25%	-0.017222	-0.029946	-0.025721	-0.025459	-0.029276	-0.035313
50%	0.017437	0.003579	0.009597	0.006452	0.007773	0.000664
75%	0.053768	0.041723	0.045363	0.040994	0.041586	0.037588
max	0.246816	0.212917	0.264517	0.198484	0.223582	0.223075

	X6	X7	X8	X9	X10	X11
count	900.000000	900.000000	900.000000	900.000000	900.000000	900.000000
mean	0.005436	0.005531	0.000582	0.000782	0.004134	0.002073
std	0.058648	0.060497	0.060715	0.060426	0.062700	0.063989
min	-0.204100	-0.193717	-0.236035	-0.203957	-0.204677	-0.225909
25%	-0.030389	-0.032960	-0.035948	-0.035540	-0.033204	-0.034348
50%	0.004227	0.005738	0.001574	0.000962	0.003430	0.002616
75%	0.038796	0.042103	0.038255	0.035687	0.041876	0.038642
max	0.196971	0.220623	0.171001	0.199146	0.202497	0.221052

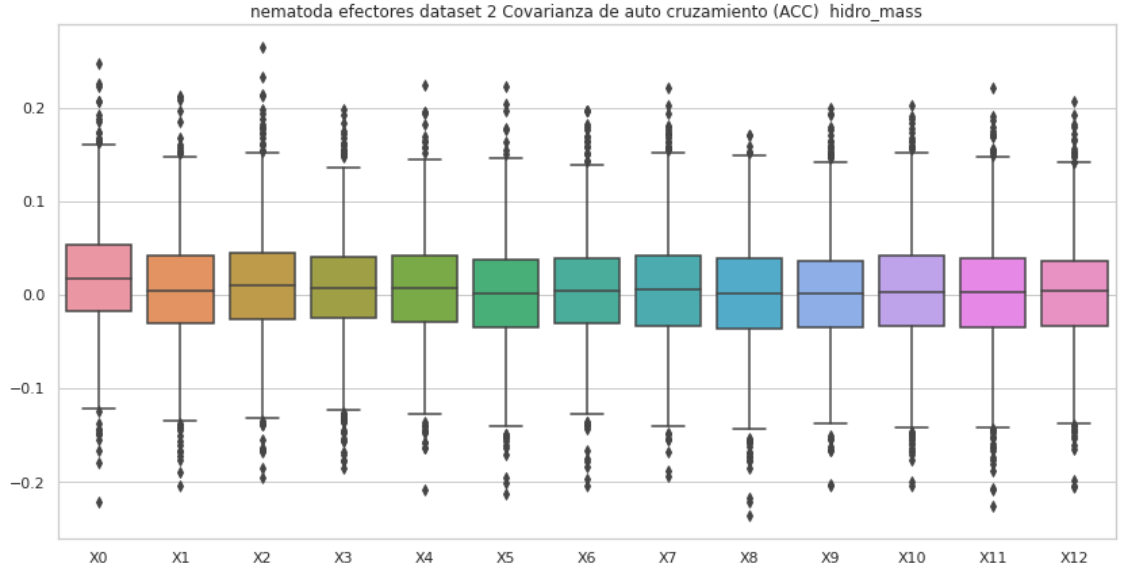
	X12
count	900.000000
mean	0.002058
std	0.059843
min	-0.205522
25%	-0.033540
50%	0.003495
75%	0.036439
max	0.206806

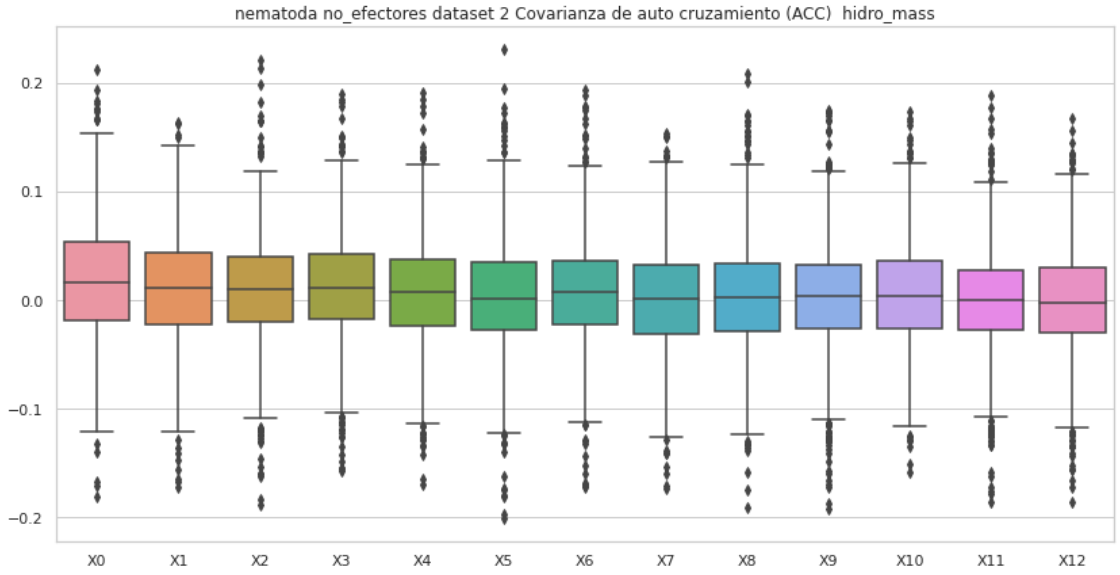
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 2, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.005499  0.076731  0.082069  0.042332 -0.025083 -0.009055  0.016039
1    0.047362 -0.016310  0.088810 -0.012256  0.054297  0.061081  0.032326
2    0.017583  0.020081  0.010700 -0.003859 -0.030420 -0.011289 -0.044492
3    0.020465  0.056849 -0.130197 -0.148779 -0.042863  0.032723  0.024572
4   -0.034586  0.037085 -0.053004  0.051479  0.036376 -0.015853 -0.026307
..        …         …         …         …         …         …
995  0.061555  0.044298 -0.030270  0.086378  0.014327 -0.031923  0.009457
996  0.022294  0.018190 -0.000886  0.047173  0.036217  0.029687  0.041919
997  0.154071  0.044966  0.073540  0.061560 -0.006831 -0.010841 -0.077733
998  0.088668 -0.006023 -0.011232  0.028521 -0.013962 -0.026034  0.051717
999 -0.032358  0.085311 -0.076869  0.033565 -0.095810 -0.010500  0.052686

           X7        X8        X9       X10       X11       X12           X13
0    0.023109  0.028802 -0.041983 -0.044756 -0.021222 -0.021476  no_efectores
1    0.056652  0.155209 -0.040102 -0.015698  0.019768  0.003394  no_efectores
2   -0.026387 -0.013318  0.004201 -0.001084  0.027858 -0.007239  no_efectores
3    0.079469 -0.083367 -0.036984 -0.059086 -0.012986 -0.046602  no_efectores
4   -0.006043  0.020451  0.014907  0.042281 -0.020481  0.077228  no_efectores
..        …         …         …         …         …         …           …
995 -0.006285  0.058376 -0.014329  0.046068  0.025824  0.026863  no_efectores
996  0.043219  0.020760 -0.000530 -0.003995 -0.031675 -0.029359  no_efectores
997 -0.086480 -0.037878  0.013159  0.005679 -0.011079  0.045636  no_efectores
998 -0.014522 -0.062910 -0.037692  0.003649  0.066375 -0.010581  no_efectores
999 -0.036550  0.011394 -0.054323  0.048895 -0.042793  0.026262  no_efectores

[897 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  897.000000  897.000000  897.000000  897.000000  897.000000  897.000000
mean     0.018054    0.009478    0.009581    0.012235    0.007921    0.002283
std      0.057209    0.052363    0.052380    0.051516    0.050062    0.054382
min     -0.181212   -0.172645   -0.187987   -0.157458   -0.170328   -0.200941
25%     -0.018554   -0.022453   -0.019699   -0.016822   -0.023573   -0.027882
50%      0.016883    0.011597    0.010325    0.011042    0.007736    0.000939
75%      0.053829    0.043946    0.039695    0.042642    0.037352    0.035274
max      0.212557    0.164062    0.221281    0.189621    0.191602    0.231029

                X6          X7          X8          X9         X10         X11  \
count  897.000000  897.000000  897.000000  897.000000  897.000000  897.000000
mean     0.006378    0.000703    0.004256    0.003080    0.004731    0.000697
std      0.050864    0.049948    0.052128    0.052788    0.050176    0.050284
```

min	-0.172087	-0.173785	-0.190901	-0.192326	-0.158385	-0.185856
25%	-0.022835	-0.030789	-0.028972	-0.025722	-0.026206	-0.027236
50%	0.007494	0.001525	0.003079	0.004382	0.003472	0.000574
75%	0.036899	0.033126	0.034382	0.032102	0.036056	0.027903
max	0.193953	0.153300	0.208419	0.174836	0.173814	0.188286

	X12
count	897.000000
mean	-0.001097
std	0.050784
min	-0.186520
25%	-0.029824
50%	-0.002598
75%	0.029845
max	0.167926





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    0.083639 -0.005352  0.035217 -0.075515 -0.049660  0.002040  0.051325
1    0.076786 -0.015779 -0.019657  0.071822  0.041807 -0.014048 -0.041745
2    0.030428 -0.026574 -0.102359 -0.051958  0.007096  0.010145 -0.005425
3   -0.035800 -0.025878  0.111503  0.002468 -0.023529  0.052107 -0.061651
4   -0.088596 -0.097694  0.062204 -0.011290  0.006157 -0.001605 -0.037680
..        …         …         …         …         …         …         …
995  0.021342 -0.076007  0.080615  0.104933 -0.014077  0.098101  0.036762
996  0.013499 -0.020290 -0.060178 -0.014671 -0.067453 -0.113696  0.022916
997  0.000233 -0.013153  0.008006 -0.010721 -0.039223  0.050749 -0.013885
998 -0.095723  0.046219 -0.070288  0.002326 -0.148286 -0.052007 -0.105643
999  0.011385 -0.031704  0.026109 -0.177202 -0.007913 -0.085267  0.070085

           X7        X8        X9       X10       X11       X12         X13
0   -0.000714 -0.032872  0.046546 -0.015022 -0.045487 -0.014852  efectores
1   -0.011559  0.007187 -0.056380 -0.052853  0.037983 -0.064151  efectores
2   -0.045543 -0.053033  0.014090  0.027548  0.104552  0.062548  efectores
3   -0.123213 -0.091409  0.005354 -0.088815 -0.152576  0.083276  efectores
4    0.036548 -0.120872 -0.047274  0.080199 -0.077982  0.050726  efectores
..        …         …         …         …         …         …         …
995 -0.065346 -0.024897 -0.121400 -0.034164  0.081307  0.016672  efectores
996  0.007105 -0.072166 -0.101698 -0.071585  0.036963  0.113315  efectores
997  0.088244 -0.039619  0.009982 -0.026287  0.064800 -0.031305  efectores
998 -0.115396  0.076165  0.007793  0.040005  0.140017 -0.048874  efectores
999  0.063389 -0.130041  0.070803 -0.154984  0.011004 -0.079799  efectores

[1000 rows x 14 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016487     0.004659     0.015174     0.005992     0.006699
```

```
std       0.084053     0.071220     0.086237     0.069075     0.074446
min      -1.260053    -0.302711    -0.360285    -0.293564    -0.335849
25%      -0.020759    -0.032896    -0.026970    -0.026811    -0.033250
50%       0.017334     0.002567     0.010657     0.006724     0.007402
75%       0.056541     0.043087     0.050794     0.043054     0.044136
max       0.331436     0.437983     0.660108     0.321100     0.611287

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.005285     0.004640     0.001927     0.005201     0.000758
std       0.082888     0.073427     0.073150     0.082656     0.070704
min      -0.721238    -0.292843    -0.782818    -0.385898    -0.366903
25%      -0.037178    -0.033149    -0.036654    -0.038345    -0.036821
50%       0.001097     0.003479     0.005276     0.003017     0.000869
75%       0.040256     0.040196     0.042391     0.043578     0.037919
max       0.563584     0.707573     0.342486     0.596790     0.462342

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.001482     0.008344    -0.000646
std       0.075886     0.083553     0.070102
min      -0.382596    -0.354719    -0.416622
25%      -0.037434    -0.035257    -0.037491
50%       0.002789     0.003412     0.002387
75%       0.042596     0.042862     0.037316
max       0.503795     0.506667     0.236848
```

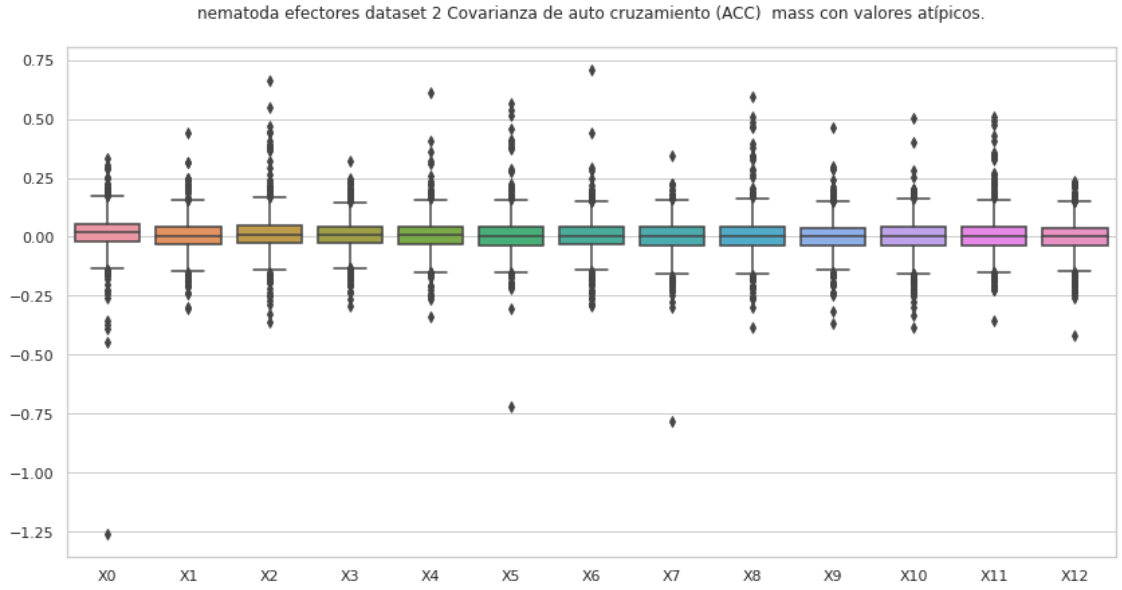
no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, con valores atípicos.
Valores del documento csv.

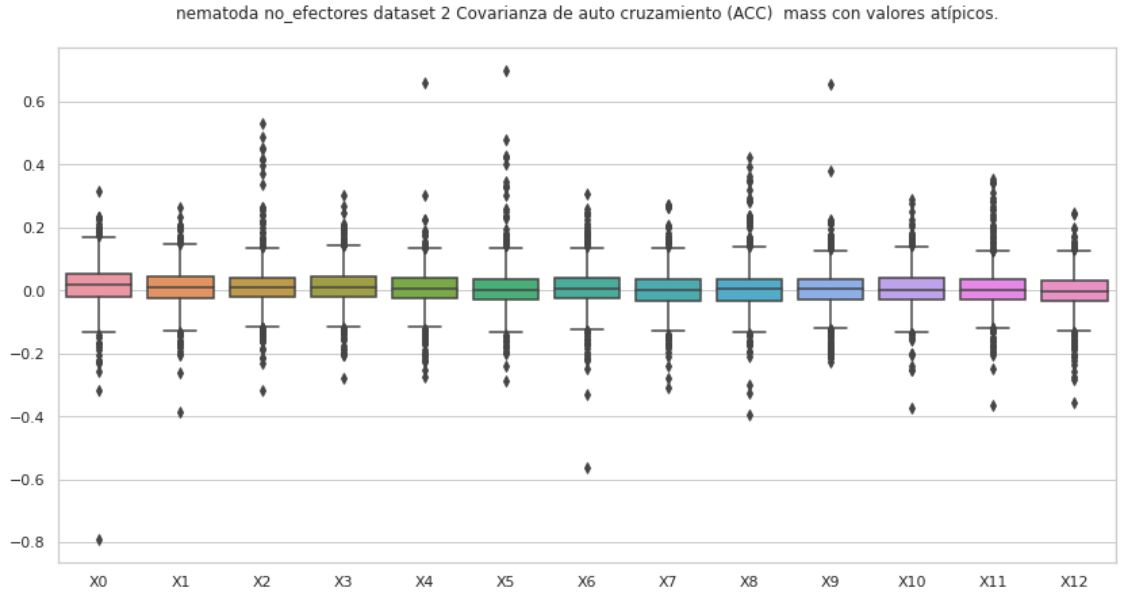
```
           X0        X1        X2        X3        X4        X5        X6  \
0   -0.005499  0.076731  0.082069  0.042332 -0.025083 -0.009055  0.016039
1    0.047362 -0.016310  0.088810 -0.012256  0.054297  0.061081  0.032326
2    0.017583  0.020081  0.010700 -0.003859 -0.030420 -0.011289 -0.044492
3    0.020465  0.056849 -0.130197 -0.148779 -0.042863  0.032723  0.024572
4   -0.034586  0.037085 -0.053004  0.051479  0.036376 -0.015853 -0.026307
..        …         …         …         …         …         …         …
995  0.061555  0.044298 -0.030270  0.086378  0.014327 -0.031923  0.009457
996  0.022294  0.018190 -0.000886  0.047173  0.036217  0.029687  0.041919
997  0.154071  0.044966  0.073540  0.061560 -0.006831 -0.010841 -0.077733
998  0.088668 -0.006023 -0.011232  0.028521 -0.013962 -0.026034  0.051717
999 -0.032358  0.085311 -0.076869  0.033565 -0.095810 -0.010500  0.052686

           X7        X8        X9       X10       X11       X12           X13
```

```
0    0.023109  0.028802 -0.041983 -0.044756 -0.021222 -0.021476  no_efectores
1    0.056652  0.155209 -0.040102 -0.015698  0.019768  0.003394  no_efectores
2   -0.026387 -0.013318  0.004201 -0.001084  0.027858 -0.007239  no_efectores
3    0.079469 -0.083367 -0.036984 -0.059086 -0.012986 -0.046602  no_efectores
4   -0.006043  0.020451  0.014907  0.042281 -0.020481  0.077228  no_efectores
..        …       …       …       …       …       …           …
995 -0.006285  0.058376 -0.014329  0.046068  0.025824  0.026863  no_efectores
996  0.043219  0.020760 -0.000530 -0.003995 -0.031675 -0.029359  no_efectores
997 -0.086480 -0.037878  0.013159  0.005679 -0.011079  0.045636  no_efectores
998 -0.014522 -0.062910 -0.037692  0.003649  0.066375 -0.010581  no_efectores
999 -0.036550  0.011394 -0.054323  0.048895 -0.042793  0.026262  no_efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.016331     0.009608     0.013093     0.012879     0.007002
std       0.071328     0.061263     0.072036     0.060608     0.063184
min      -0.791093    -0.388256    -0.315379    -0.278335    -0.275304
25%      -0.021346    -0.024410    -0.021331    -0.019319    -0.025158
50%       0.016799     0.012000     0.010348     0.011663     0.007424
75%       0.054965     0.045019     0.041927     0.045171     0.038855
max       0.316066     0.262523     0.531280     0.301710     0.658696

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.006226     0.005860     0.002009     0.007046     0.004325
std       0.075050     0.064361     0.060462     0.070106     0.066063
min      -0.285741    -0.564501    -0.307445    -0.392712    -0.225464
25%      -0.030129    -0.025377    -0.031967    -0.031517    -0.027230
50%       0.001062     0.007349     0.002102     0.004259     0.005088
75%       0.038101     0.039334     0.036117     0.037396     0.035347
max       0.698981     0.305243     0.273823     0.423853     0.655517

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.005102     0.005749    -0.001382
std       0.061155     0.066160     0.061773
min      -0.373780    -0.365769    -0.354705
25%      -0.029701    -0.027269    -0.031834
50%       0.003492     0.001701    -0.002034
75%       0.038513     0.033965     0.032423
max       0.290336     0.353955     0.247970
```


nematoda efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 2 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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	\
0	0.083639	-0.005352	0.035217	-0.075515	-0.049660	0.002040	0.051325	
1	0.076786	-0.015779	-0.019657	0.071822	0.041807	-0.014048	-0.041745	
2	0.030428	-0.026574	-0.102359	-0.051958	0.007096	0.010145	-0.005425	
3	-0.035800	-0.025878	0.111503	0.002468	-0.023529	0.052107	-0.061651	
4	-0.088596	-0.097694	0.062204	-0.011290	0.006157	-0.001605	-0.037680	
..	…	…	…	…	…	…	…	
995	0.021342	-0.076007	0.080615	0.104933	-0.014077	0.098101	0.036762	
996	0.013499	-0.020290	-0.060178	-0.014671	-0.067453	-0.113696	0.022916	
997	0.000233	-0.013153	0.008006	-0.010721	-0.039223	0.050749	-0.013885	
998	-0.095723	0.046219	-0.070288	0.002326	-0.148286	-0.052007	-0.105643	
999	0.011385	-0.031704	0.026109	-0.177202	-0.007913	-0.085267	0.070085	

	X7	X8	X9	X10	X11	X12	X13
0	-0.000714	-0.032872	0.046546	-0.015022	-0.045487	-0.014852	efectores
1	-0.011559	0.007187	-0.056380	-0.052853	0.037983	-0.064151	efectores
2	-0.045543	-0.053033	0.014090	0.027548	0.104552	0.062548	efectores
3	-0.123213	-0.091409	0.005354	-0.088815	-0.152576	0.083276	efectores
4	0.036548	-0.120872	-0.047274	0.080199	-0.077982	0.050726	efectores
..	…	…	…	…	…	…	
995	-0.065346	-0.024897	-0.121400	-0.034164	0.081307	0.016672	efectores
996	0.007105	-0.072166	-0.101698	-0.071585	0.036963	0.113315	efectores
997	0.088244	-0.039619	0.009982	-0.026287	0.064800	-0.031305	efectores
998	-0.115396	0.076165	0.007793	0.040005	0.140017	-0.048874	efectores
999	0.063389	-0.130041	0.070803	-0.154984	0.011004	-0.079799	efectores

```
[900 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	900.000000	900.000000	900.000000	900.000000	900.000000	900.000000	
mean	0.019054	0.004803	0.010884	0.006660	0.006741	0.000739	
std	0.060270	0.060343	0.062920	0.059132	0.059738	0.058267	
min	-0.222268	-0.204401	-0.196184	-0.186019	-0.208635	-0.212994	
25%	-0.017222	-0.029946	-0.025721	-0.025459	-0.029276	-0.035313	
50%	0.017437	0.003579	0.009597	0.006452	0.007773	0.000664	
75%	0.053768	0.041723	0.045363	0.040994	0.041586	0.037588	

```
max      0.246816    0.212917    0.264517    0.198484    0.223582    0.223075

               X6          X7          X8          X9         X10         X11  \
count  900.000000  900.000000  900.000000  900.000000  900.000000  900.000000
mean     0.005436    0.005531    0.000582    0.000782    0.004134    0.002073
std      0.058648    0.060497    0.060715    0.060426    0.062700    0.063989
min     -0.204100   -0.193717   -0.236035   -0.203957   -0.204677   -0.225909
25%     -0.030389   -0.032960   -0.035948   -0.035540   -0.033204   -0.034348
50%      0.004227    0.005738    0.001574    0.000962    0.003430    0.002616
75%      0.038796    0.042103    0.038255    0.035687    0.041876    0.038642
max      0.196971    0.220623    0.171001    0.199146    0.202497    0.221052

              X12
count  900.000000
mean     0.002058
std      0.059843
min     -0.205522
25%     -0.033540
50%      0.003495
75%      0.036439
max      0.206806
```

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   -0.005499  0.076731  0.082069  0.042332 -0.025083 -0.009055  0.016039
1    0.047362 -0.016310  0.088810 -0.012256  0.054297  0.061081  0.032326
2    0.017583  0.020081  0.010700 -0.003859 -0.030420 -0.011289 -0.044492
3    0.020465  0.056849 -0.130197 -0.148779 -0.042863  0.032723  0.024572
4   -0.034586  0.037085 -0.053004  0.051479  0.036376 -0.015853 -0.026307
..        …         …         …         …         …         …
995  0.061555  0.044298 -0.030270  0.086378  0.014327 -0.031923  0.009457
996  0.022294  0.018190 -0.000886  0.047173  0.036217  0.029687  0.041919
997  0.154071  0.044966  0.073540  0.061560 -0.006831 -0.010841 -0.077733
998  0.088668 -0.006023 -0.011232  0.028521 -0.013962 -0.026034  0.051717
999 -0.032358  0.085311 -0.076869  0.033565 -0.095810 -0.010500  0.052686

           X7        X8        X9       X10       X11       X12           X13
0    0.023109  0.028802 -0.041983 -0.044756 -0.021222 -0.021476  no_efectores
1    0.056652  0.155209 -0.040102 -0.015698  0.019768  0.003394  no_efectores
2   -0.026387 -0.013318  0.004201 -0.001084  0.027858 -0.007239  no_efectores
3    0.079469 -0.083367 -0.036984 -0.059086 -0.012986 -0.046602  no_efectores
4   -0.006043  0.020451  0.014907  0.042281 -0.020481  0.077228  no_efectores
..        …         …         …         …         …         …           …
```

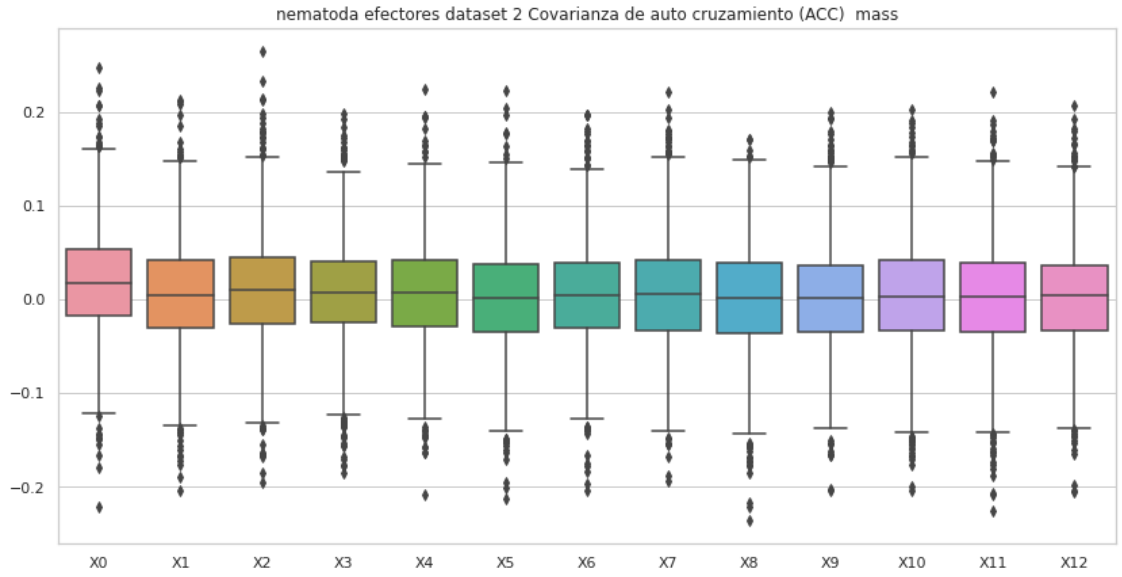
```
995 -0.006285  0.058376 -0.014329  0.046068  0.025824  0.026863  no_efectores
996  0.043219  0.020760 -0.000530 -0.003995 -0.031675 -0.029359  no_efectores
997 -0.086480 -0.037878  0.013159  0.005679 -0.011079  0.045636  no_efectores
998 -0.014522 -0.062910 -0.037692  0.003649  0.066375 -0.010581  no_efectores
999 -0.036550  0.011394 -0.054323  0.048895 -0.042793  0.026262  no_efectores

[897 rows x 14 columns]
```

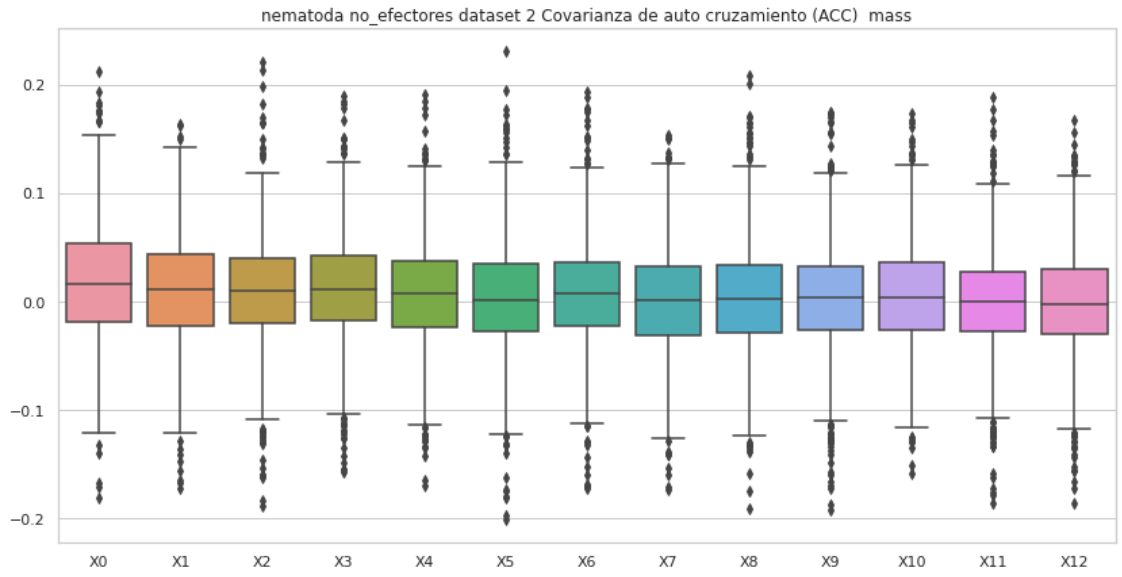
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	897.000000	897.000000	897.000000	897.000000	897.000000	897.000000	
mean	0.018054	0.009478	0.009581	0.012235	0.007921	0.002283	
std	0.057209	0.052363	0.052380	0.051516	0.050062	0.054382	
min	-0.181212	-0.172645	-0.187987	-0.157458	-0.170328	-0.200941	
25%	-0.018554	-0.022453	-0.019699	-0.016822	-0.023573	-0.027882	
50%	0.016883	0.011597	0.010325	0.011042	0.007736	0.000939	
75%	0.053829	0.043946	0.039695	0.042642	0.037352	0.035274	
max	0.212557	0.164062	0.221281	0.189621	0.191602	0.231029	

	X6	X7	X8	X9	X10	X11	\
count	897.000000	897.000000	897.000000	897.000000	897.000000	897.000000	
mean	0.006378	0.000703	0.004256	0.003080	0.004731	0.000697	
std	0.050864	0.049948	0.052128	0.052788	0.050176	0.050284	
min	-0.172087	-0.173785	-0.190901	-0.192326	-0.158385	-0.185856	
25%	-0.022835	-0.030789	-0.028972	-0.025722	-0.026206	-0.027236	
50%	0.007494	0.001525	0.003079	0.004382	0.003472	0.000574	
75%	0.036899	0.033126	0.034382	0.032102	0.036056	0.027903	
max	0.193953	0.153300	0.208419	0.174836	0.173814	0.188286	

	X12
count	897.000000
mean	-0.001097
std	0.050784
min	-0.186520
25%	-0.029824
50%	-0.002598
75%	0.029845
max	0.167926





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    0.005487 -0.065761 -0.041958  0.178190 -0.074579 -0.032554 -0.007919
1   -0.029845 -0.166679  0.032014  0.075351 -0.142143 -0.137507 -0.050132
2    0.045278  0.027006  0.019188  0.033605  0.154109  0.081685  0.160990
3    0.049300 -0.114221  0.228394  0.238401 -0.013490 -0.116776  0.074942
4   -0.028155 -0.006474 -0.048199  0.066140  0.059818 -0.058710 -0.007243
..        …       …       …       …       …       …       …
995  0.109731  0.010849  0.071020  0.113469  0.038145  0.154615 -0.005366
996  0.236047  0.165040 -0.040376 -0.168658 -0.190137 -0.063790  0.007101
997 -0.010838 -0.031124  0.007333  0.059656 -0.029610 -0.065264  0.045684
998  0.098222 -0.208374  0.210427 -0.011188 -0.131294  0.047619 -0.165651
999 -0.007118  0.002452 -0.116180 -0.110279  0.111882 -0.026476 -0.011044

           X7        X8        X9       X10       X11       X12        X13
0    0.042818  0.036866 -0.024790 -0.017977  0.000969  0.050111  efectores
1   -0.016305 -0.037514  0.121154  0.024898 -0.083630  0.061478  efectores
2    0.236606  0.096198  0.086292 -0.170299 -0.013370  0.088767  efectores
```

```
3    0.259265  0.102810 -0.147847  0.051641  0.099506 -0.004343  efectores
4   -0.014042  0.017436 -0.070943 -0.010336 -0.053594 -0.000475  efectores
..        …       …        …        …        …        …        …
995  0.037277  0.032327  0.048997  0.140814  0.107225 -0.072937  efectores
996  0.033595 -0.082945 -0.087496 -0.150374 -0.127837  0.003126  efectores
997  0.025975 -0.074362  0.057303  0.054134  0.002172 -0.009771  efectores
998 -0.070635  0.143460  0.073582 -0.180373 -0.154924  0.134362  efectores
999 -0.097273  0.037709  0.164849  0.031713  0.065001  0.080295  efectores

[1000 rows x 14 columns]
```

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

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014161    -0.016804     0.021964     0.024855    -0.005344
std       0.094258     0.093423     0.084306     0.090419     0.088527
min      -0.303317    -0.346397    -0.220723    -0.385116    -0.306711
25%      -0.035509    -0.069689    -0.029273    -0.027389    -0.057520
50%       0.008928    -0.021508     0.021374     0.020890    -0.005913
75%       0.059097     0.034999     0.068377     0.078834     0.044050
max       0.811952     0.665826     0.568786     0.619428     0.443600

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.004962     0.020443     0.010091     0.001384     0.011175
std       0.085368     0.084719     0.084353     0.084033     0.081418
min      -0.312704    -0.431753    -0.336159    -0.371803    -0.356187
25%      -0.054259    -0.026955    -0.039675    -0.049255    -0.036883
50%      -0.003757     0.017928     0.007804     0.000597     0.010232
75%       0.042984     0.068179     0.053013     0.052285     0.059308
max       0.405250     0.419183     0.442479     0.319801     0.471986

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.012562     0.005461     0.001725
std       0.088131     0.083456     0.086494
min      -0.419070    -0.441492    -0.396141
25%      -0.034581    -0.040325    -0.043445
50%       0.015057     0.007915     0.005411
75%       0.056592     0.050234     0.045189
max       0.674340     0.414265     0.382024
```

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	0.068748	0.056947	-0.042360	-0.018273	0.005737	0.081655	0.023611	
1	0.051789	0.032428	0.065307	-0.011303	-0.014856	0.016662	0.008354	
2	0.039582	0.063449	-0.010639	0.028605	0.130174	0.136114	0.049844	
3	-0.008659	-0.104624	-0.038670	0.009722	-0.120511	-0.064990	-0.026213	
4	0.136121	-0.120622	-0.161591	0.148650	-0.028389	-0.020394	0.038823	
..	…	…	…	…	…	…	…	
995	0.128755	0.035459	0.109912	0.065066	0.048928	0.033528	-0.007972	
996	-0.018770	-0.030946	0.079082	0.010967	-0.042402	0.005248	0.037244	
997	0.089502	-0.075489	0.005154	0.029694	0.039195	0.093126	0.102349	
998	-0.039679	0.031083	0.093381	0.045741	-0.174179	0.170890	-0.044283	
999	-0.034723	-0.128242	0.071076	0.110590	0.086740	-0.117433	0.005722	

	X7	X8	X9	X10	X11	X12	X13
0	0.082080	0.042146	0.061169	-0.012962	0.047095	0.058166	no_efectores
1	-0.015595	-0.006109	-0.010000	-0.037298	0.020351	0.012291	no_efectores
2	0.074022	0.036318	0.149772	-0.047811	-0.008311	0.031173	no_efectores
3	-0.006255	-0.119927	0.046382	0.204636	0.138684	-0.032102	no_efectores
4	-0.053321	0.035279	0.043157	-0.005550	-0.115273	-0.055995	no_efectores
..	…	…	…	…	…	…	…
995	0.013315	-0.040147	-0.053504	-0.037818	-0.051283	-0.035223	no_efectores
996	0.017546	0.046421	-0.020746	0.035457	0.015404	0.002755	no_efectores
997	-0.038711	-0.048634	0.029522	0.152447	0.153752	-0.037280	no_efectores
998	0.020534	-0.031291	-0.037835	0.159663	0.062752	0.042555	no_efectores
999	0.137757	0.039492	-0.053610	0.042271	-0.070214	0.057762	no_efectores

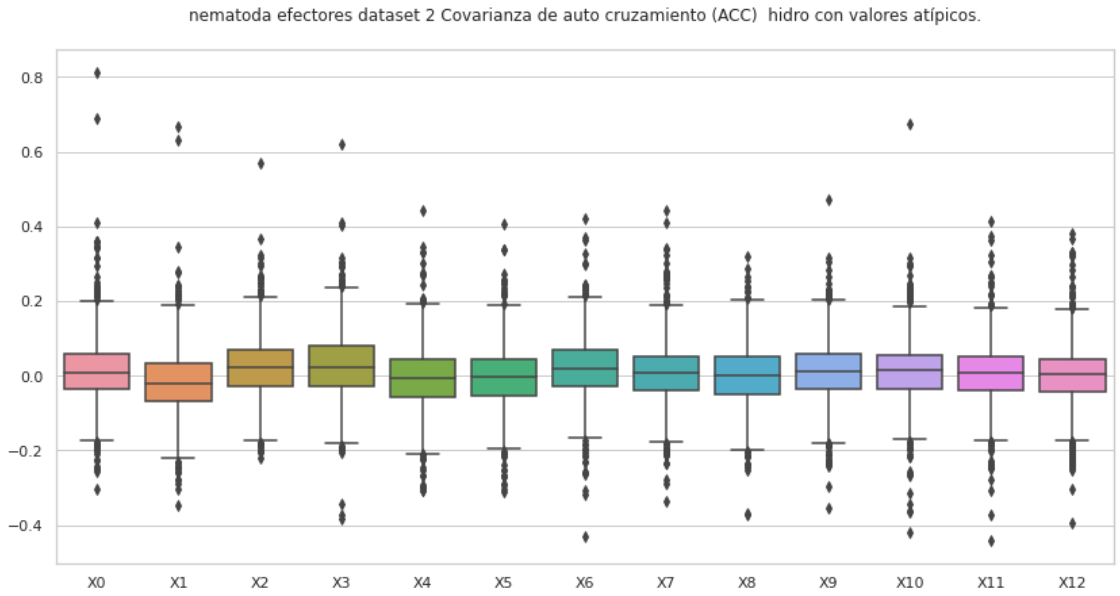
[1000 rows x 14 columns]

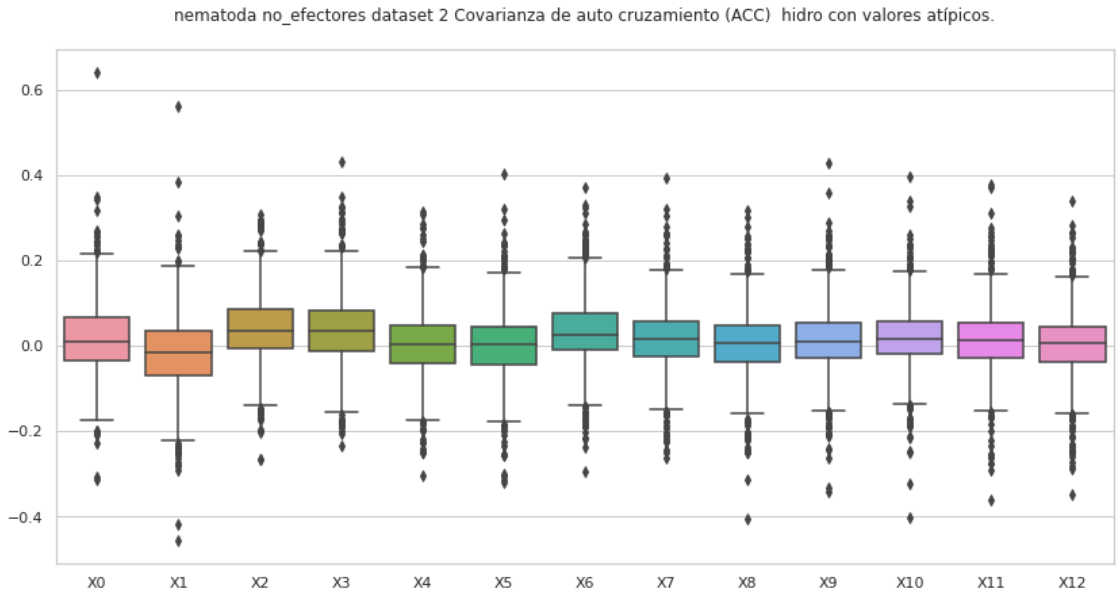
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 2, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.016874	-0.016579	0.039737	0.036028	0.004172	
std	0.085694	0.089535	0.079245	0.079299	0.078074	
min	-0.315257	-0.456637	-0.266640	-0.236093	-0.304816	
25%	-0.035549	-0.069751	-0.005451	-0.011958	-0.041185	
50%	0.010081	-0.016223	0.035237	0.033454	0.003163	
75%	0.065264	0.033232	0.085375	0.083351	0.048590	
max	0.639787	0.560943	0.306258	0.431869	0.314213	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.001187	0.033410	0.017194	0.006910	0.013615	
std	0.080845	0.081221	0.075647	0.075870	0.078233	
min	-0.319825	-0.294973	-0.263087	-0.406546	-0.341136	
25%	-0.044478	-0.011047	-0.026929	-0.036666	-0.028044	
50%	0.003667	0.026632	0.014139	0.007471	0.010019	
75%	0.045559	0.075525	0.055766	0.047207	0.055188	
max	0.403113	0.370331	0.392892	0.316771	0.427117	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.017378	0.012722	0.004064
std	0.074672	0.077196	0.075765
min	-0.404302	-0.360174	-0.349674
25%	-0.020807	-0.029992	-0.037291
50%	0.014450	0.013895	0.005695
75%	0.057178	0.052279	0.043607
max	0.396414	0.377848	0.338518





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    0.005487 -0.065761 -0.041958  0.178190 -0.074579 -0.032554 -0.007919
1   -0.029845 -0.166679  0.032014  0.075351 -0.142143 -0.137507 -0.050132
2    0.045278  0.027006  0.019188  0.033605  0.154109  0.081685  0.160990
3    0.049300 -0.114221  0.228394  0.238401 -0.013490 -0.116776  0.074942
4   -0.028155 -0.006474 -0.048199  0.066140  0.059818 -0.058710 -0.007243
..        …       …       …       …       …       …       …
995  0.109731  0.010849  0.071020  0.113469  0.038145  0.154615 -0.005366
996  0.236047  0.165040 -0.040376 -0.168658 -0.190137 -0.063790  0.007101
997 -0.010838 -0.031124  0.007333  0.059656 -0.029610 -0.065264  0.045684
998  0.098222 -0.208374  0.210427 -0.011188 -0.131294  0.047619 -0.165651
999 -0.007118  0.002452 -0.116180 -0.110279  0.111882 -0.026476 -0.011044

           X7        X8        X9       X10       X11       X12        X13
0    0.042818  0.036866 -0.024790 -0.017977  0.000969  0.050111  efectores
1   -0.016305 -0.037514  0.121154  0.024898 -0.083630  0.061478  efectores
2    0.236606  0.096198  0.086292 -0.170299 -0.013370  0.088767  efectores
3    0.259265  0.102810 -0.147847  0.051641  0.099506 -0.004343  efectores
4   -0.014042  0.017436 -0.070943 -0.010336 -0.053594 -0.000475  efectores
..        …       …       …       …       …       …       …
995  0.037277  0.032327  0.048997  0.140814  0.107225 -0.072937  efectores
```

```
996  0.033595 -0.082945 -0.087496 -0.150374 -0.127837  0.003126  efectores
997  0.025975 -0.074362  0.057303  0.054134  0.002172 -0.009771  efectores
998 -0.070635  0.143460  0.073582 -0.180373 -0.154924  0.134362  efectores
999 -0.097273  0.037709  0.164849  0.031713  0.065001  0.080295  efectores

[914 rows x 14 columns]
```

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

```
              X0          X1          X2          X3          X4          X5  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.010848   -0.018270    0.017108    0.021551   -0.005679   -0.005590
std      0.076527    0.080608    0.076610    0.077323    0.077393    0.072890
min     -0.255681   -0.259522   -0.220723   -0.199821   -0.266707   -0.252556
25%     -0.032018   -0.068147   -0.031975   -0.026705   -0.055766   -0.050390
50%      0.009055   -0.022318    0.019823    0.019662   -0.007092   -0.003704
75%      0.056230    0.032401    0.064628    0.073531    0.041069    0.038139
max      0.263605    0.242524    0.266278    0.290661    0.243343    0.247648

              X6          X7          X8          X9         X10         X11  \
count  914.000000  914.000000  914.000000  914.000000  914.000000  914.000000
mean     0.020900    0.007797   -0.001261    0.008518    0.012356    0.003518
std      0.072966    0.072658    0.075791    0.070826    0.073754    0.069518
min     -0.229957   -0.234289   -0.237475   -0.232506   -0.218615   -0.239633
25%     -0.024398   -0.038149   -0.048216   -0.035124   -0.031731   -0.038895
50%      0.018333    0.007572   -0.001320    0.009469    0.014542    0.006443
75%      0.063738    0.049591    0.046880    0.054416    0.053641    0.046589
max      0.242555    0.262321    0.223843    0.234159    0.246495    0.218209

             X12
count  914.000000
mean    -0.000480
std      0.075072
min     -0.252134
25%     -0.042148
50%      0.005161
75%      0.043410
max      0.241175
```

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

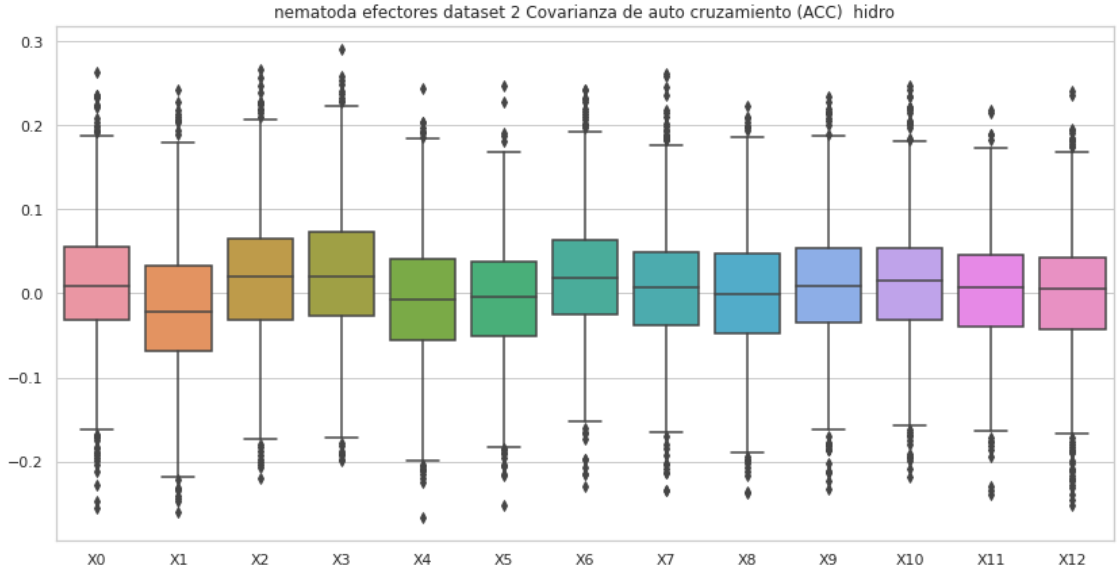
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.068748  0.056947 -0.042360 -0.018273  0.005737  0.081655  0.023611
1    0.051789  0.032428  0.065307 -0.011303 -0.014856  0.016662  0.008354
2    0.039582  0.063449 -0.010639  0.028605  0.130174  0.136114  0.049844
3   -0.008659 -0.104624 -0.038670  0.009722 -0.120511 -0.064990 -0.026213
4    0.136121 -0.120622 -0.161591  0.148650 -0.028389 -0.020394  0.038823
..        …         …         …         …         …         …
995  0.128755  0.035459  0.109912  0.065066  0.048928  0.033528 -0.007972
996 -0.018770 -0.030946  0.079082  0.010967 -0.042402  0.005248  0.037244
997  0.089502 -0.075489  0.005154  0.029694  0.039195  0.093126  0.102349
998 -0.039679  0.031083  0.093381  0.045741 -0.174179  0.170890 -0.044283
999 -0.034723 -0.128242  0.071076  0.110590  0.086740 -0.117433  0.005722

           X7        X8        X9       X10       X11       X12           X13
0    0.082080  0.042146  0.061169 -0.012962  0.047095  0.058166  no_efectores
1   -0.015595 -0.006109 -0.010000 -0.037298  0.020351  0.012291  no_efectores
2    0.074022  0.036318  0.149772 -0.047811 -0.008311  0.031173  no_efectores
3   -0.006255 -0.119927  0.046382  0.204636  0.138684 -0.032102  no_efectores
4   -0.053321  0.035279  0.043157 -0.005550 -0.115273 -0.055995  no_efectores
..        …         …         …         …         …         …
995  0.013315 -0.040147 -0.053504 -0.037818 -0.051283 -0.035223  no_efectores
996  0.017546  0.046421 -0.020746  0.035457  0.015404  0.002755  no_efectores
997 -0.038711 -0.048634  0.029522  0.152447  0.153752 -0.037280  no_efectores
998  0.020534 -0.031291 -0.037835  0.159663  0.062752  0.042555  no_efectores
999  0.137757  0.039492 -0.053610  0.042271 -0.070214  0.057762  no_efectores

[903 rows x 14 columns]
```

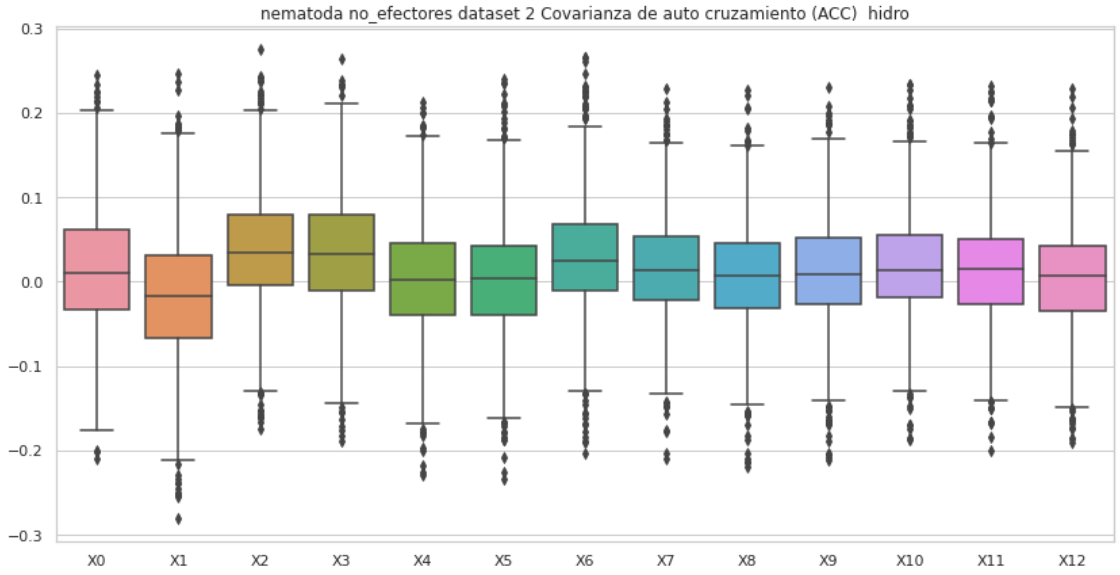
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 2, sin valores atípicos.
Estadísticas.

```
              X0          X1          X2          X3          X4          X5  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.015217   -0.017006    0.037011    0.032695    0.001999    0.001587
std      0.074372    0.078531    0.070688    0.068622    0.068587    0.070252
min     -0.209296   -0.280583   -0.174387   -0.189465   -0.229044   -0.234026
25%     -0.033951   -0.067049   -0.004748   -0.010869   -0.040145   -0.040120
50%      0.009996   -0.016799    0.033966    0.032292    0.002642    0.003290
75%      0.061206    0.031053    0.079015    0.079293    0.044822    0.042867
max      0.245226    0.245787    0.275169    0.264230    0.213202    0.240170

              X6          X7          X8          X9         X10         X11  \
count  903.000000  903.000000  903.000000  903.000000  903.000000  903.000000
mean     0.029384    0.017282    0.006531    0.010616    0.016317    0.013076
std      0.071900    0.064511    0.065449    0.065285    0.063396    0.063843
```

min	-0.203336	-0.209579	-0.219533	-0.211325	-0.186629	-0.200279
25%	-0.010960	-0.021276	-0.032114	-0.027513	-0.019410	-0.026276
50%	0.025326	0.013817	0.007526	0.009267	0.013895	0.014935
75%	0.068546	0.053502	0.045115	0.051595	0.055287	0.049920
max	0.266360	0.228573	0.226826	0.230581	0.233349	0.232225

	X12
count	903.000000
mean	0.005770
std	0.062423
min	-0.190830
25%	-0.034420
50%	0.006472
75%	0.042160
max	0.228535




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