ds5_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 = 5
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 5, con valores
atípicos.
Valores del documento csv.

         X0      X1     X2      X3     X4     X5     X6      X7      X8  \
0     8.219   9.589  6.849   5.479  1.370  6.849  2.740   6.849   1.370
1    11.675   6.091  5.584   2.030  0.000  6.091  6.091   3.553   1.523
2     9.091   6.061  4.762   6.494  2.165  6.494  3.463   6.494   3.896
3     4.895   4.196  2.797   1.399  7.692  1.399  2.797   6.993   0.699
4     7.143   2.381  9.524   4.762  2.381  4.762  0.000  11.905   4.762
..      …       …      …       …      …      …      …       …       …
995   4.950   3.960  3.960   3.960  4.950  5.941  1.980   2.970   1.980
996   5.085  10.169  0.000   0.000  5.085  3.390  3.390   1.695  10.169
997   3.205   9.615  4.487   8.974  1.923  9.615  1.923   7.051   0.641
998   7.463   4.975  4.478  10.945  1.493  4.478  3.980   8.955   2.985
999   6.364   1.818  2.727  10.000  3.636  8.182  4.545   6.364   4.545
```

```
         X9  …     X11    X12    X13     X14     X15    X16    X17    X18  \
0     5.479  …   6.849  2.740  2.740   5.479   8.219  9.589  0.000  4.110
1     4.061  …   5.076  4.569  1.015   7.107  12.690  6.599  1.015  4.061
2     5.628  …   9.091  1.732  5.195   2.597   5.195  3.896  0.000  3.896
3    11.189  …   0.699  2.098  9.091   4.895   2.797  2.797  3.497  4.196
4     4.762  …   4.762  2.381  4.762   7.143   4.762  4.762  4.762  2.381
..      … …     …    …    …     …     …    …    …    …
995   5.941  …   7.921  2.970  7.921   2.970   8.911  4.950  1.980  3.960
996   5.085  …   3.390  3.390  1.695  11.864   1.695  6.780  1.695  1.695
997   1.282  …  17.949  3.205  1.923   1.282   8.333  6.410  0.000  0.641
998   2.985  …   3.483  1.990  1.990   6.468   5.970  5.970  0.498  1.990
999   8.182  …   9.091  1.818  4.545   5.455   5.455  4.545  0.000  1.818

       X19         X20
0    1.370   efectores
1    5.076   efectores
2    4.762   efectores
3    9.091   efectores
4    4.762   efectores
..     …         …
995  6.931   efectores
996  8.475   efectores
997  3.846   efectores
998  7.463   efectores
999  1.818   efectores

[1000 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.926485     6.078096     4.309551     5.237439     2.455349
std       2.741119     2.582396     1.912150     2.325129     2.021688
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       5.083250     4.405000     3.125000     3.806750     1.127000
50%       6.667000     5.906500     4.170000     5.231000     2.033000
75%       8.333000     7.328000     5.288750     6.581000     3.262750
max      21.429000    20.000000    19.922000    30.392000    16.279000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.211196     3.851081     5.608517     2.317225     5.696186
std       2.897361     1.983701     2.884177     1.411649     2.325161
```

min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	4.357000	2.618750	3.892250	1.389000	4.193250
50%	6.029000	3.568500	5.225000	2.205500	5.546000
75%	7.630000	4.779000	6.964250	3.052250	6.973250
max	24.848000	16.327000	32.653000	10.169000	21.569000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	9.108540	5.771032	2.803935	4.397185	4.617031
std	3.050543	2.847592	1.425381	2.115062	2.785137
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	7.158000	3.893000	1.858250	2.954500	3.086750
50%	9.025500	5.393500	2.622500	4.200000	4.327000
75%	10.800000	7.203250	3.450000	5.619750	5.591000
max	26.471000	23.200000	9.804000	12.195000	31.481000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	7.785851	5.653203	1.277516	3.248736	6.645748
std	2.939797	2.347050	1.093329	1.815440	2.383340
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	5.907500	4.179750	0.519500	2.068500	5.100750
50%	7.438000	5.476500	1.097500	3.056500	6.515000
75%	9.427500	6.733750	1.826000	4.216000	8.000000
max	18.750000	20.468000	8.738000	19.697000	20.588000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8 \
0	5.769	5.769	2.885	3.846	3.846	8.654	4.808	8.654	0.962
1	2.985	13.433	1.493	5.970	1.493	2.985	0.000	7.463	0.000
2	4.290	1.650	3.960	2.310	0.000	1.650	1.320	6.931	2.310
3	6.723	10.084	3.361	5.042	0.000	7.563	3.361	3.361	1.681
4	4.603	5.439	7.113	5.439	3.347	5.021	5.021	5.439	3.347
..	…	…	…	…	…	…	…	…	…
995	6.122	6.122	2.041	0.000	8.163	2.041	10.204	6.122	0.000
996	11.664	2.058	4.117	2.058	1.372	4.631	3.431	8.576	2.573
997	3.971	7.942	3.971	2.527	1.083	3.610	3.249	2.166	3.610
998	5.438	7.162	4.111	8.488	1.194	12.467	4.642	2.520	3.448
999	5.201	5.795	4.755	4.606	3.863	5.498	3.120	2.972	1.932

	X9	…	X11	X12	X13	X14	X15	X16	X17	X18 \
0	1.923	…	11.538	0.962	1.923	4.808	4.808	2.885	1.923	1.923

```
1     5.970  …   2.985  4.478   2.985  5.970  16.418   8.955  2.985  0.000
2     6.601  …   2.640  6.601  11.881  4.290   9.901   3.300  3.300  4.290
3     4.202  …   4.202  3.361   2.521  0.840  11.765   6.723  4.202  1.681
4     6.276  …   6.695  1.255   7.113  5.858   5.439   2.929  2.092  2.092
..      …  …     …    …     …    …     …    …    …    …
995  10.204  …  16.327  2.041   4.082  0.000   6.122  12.245  0.000  0.000
996   6.175  …   4.288  3.774   6.003  5.146   6.690   6.690  3.259  3.945
997   7.581  …   5.054  2.527   6.137  4.332  10.108   7.942  1.805  3.971
998   3.581  …   8.223  2.255   2.653  2.785   8.090   5.570  0.133  0.663
999   7.429  …   5.944  2.377   6.686  5.052   8.618   5.349  0.743  4.309

        X19            X20
0     8.654   no_efectores
1     7.463   no_efectores
2     7.261   no_efectores
3    10.924   no_efectores
4     5.021   no_efectores
..      …            …
995   2.041   no_efectores
996   6.175   no_efectores
997   9.025   no_efectores
998   6.499   no_efectores
999   5.349   no_efectores

[1000 rows x 21 columns]
```

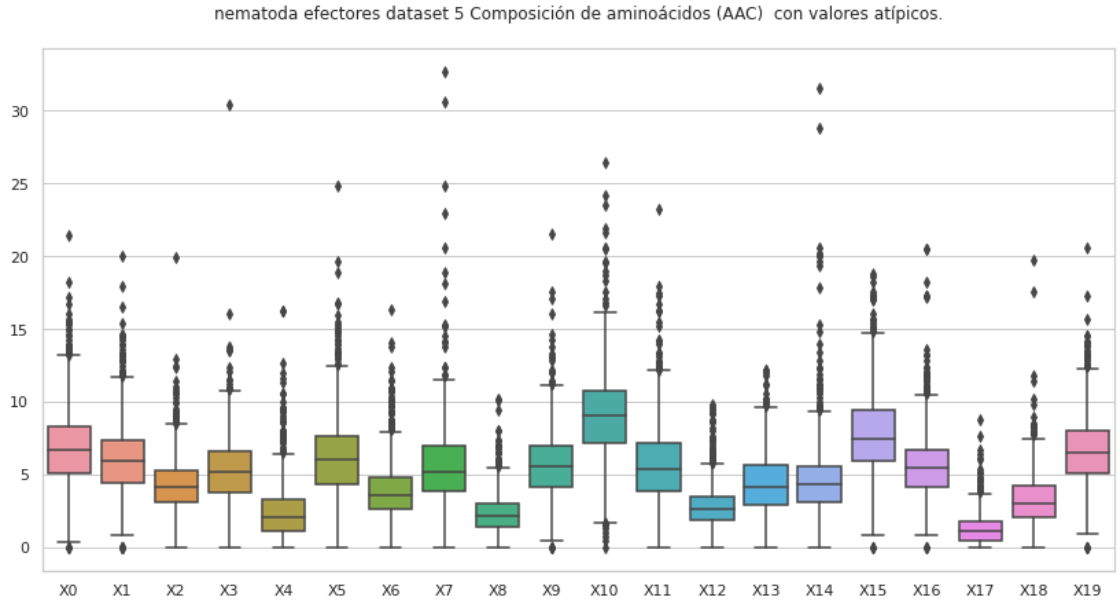
Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

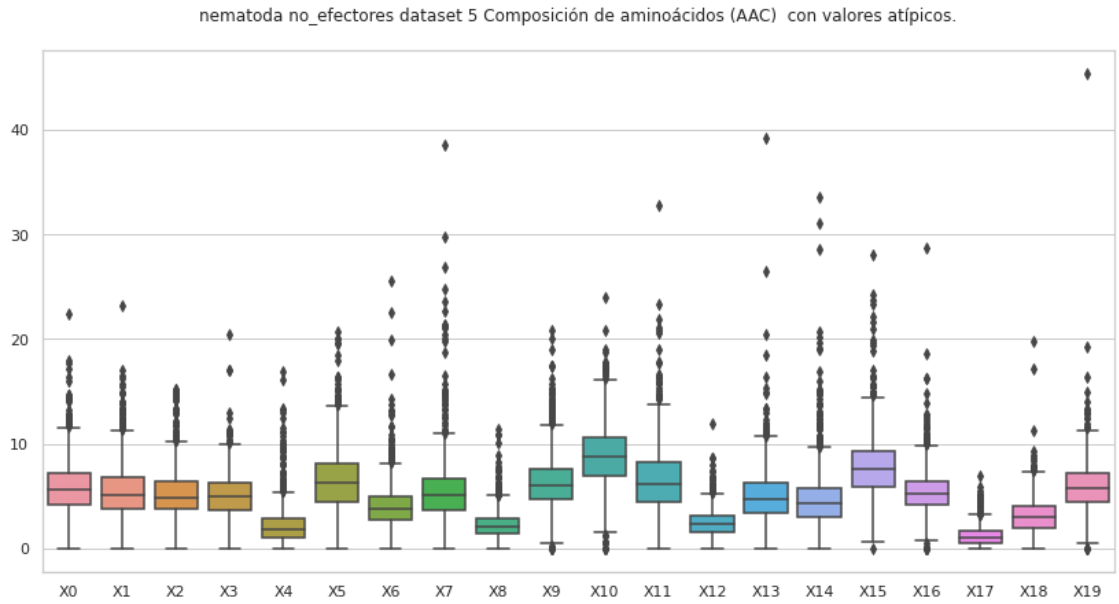
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      5.884422     5.526923     5.221171     4.975759     2.257216
std       2.616780     2.697398     2.368524     2.172565     2.002338
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.252500     3.782500     3.763500     3.725500     1.111000
50%       5.679000     5.170000     4.889500     4.935500     1.784000
75%       7.201250     6.789500     6.372500     6.236500     2.859000
max      22.388000    23.158000    15.244000    20.513000    16.923000

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      6.505224     4.042008     5.594080     2.271434     6.315016
std       3.050989     2.256453     3.292343     1.406210     2.785188
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       4.490000     2.778000     3.736250     1.447000     4.670750
50%       6.240000     3.764000     5.079500     2.083000     5.978000
```

75%	8.175250	4.921000	6.650500	2.925500	7.581250
max	20.677000	25.597000	38.587000	11.429000	20.896000

	X10	X11	X12	X13	X14 \
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	8.933062	6.646626	2.508524	5.046851	4.767627
std	3.084054	3.326252	1.258423	2.759280	3.084252
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	6.996500	4.493000	1.626000	3.364500	3.062000
50%	8.808000	6.153500	2.318000	4.683000	4.276000
75%	10.687000	8.312000	3.109000	6.330500	5.726250
max	24.051000	32.787000	11.940000	39.241000	33.621000

	X15	X16	X17	X18	X19
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	7.877886	5.373075	1.203662	3.135118	5.914359
std	3.067234	2.272993	0.976107	1.848376	2.587249
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	5.892500	4.133000	0.561250	1.926000	4.427250
50%	7.580500	5.222500	1.034000	2.963000	5.732500
75%	9.365000	6.460250	1.655500	4.118500	7.209250
max	28.077000	28.660000	6.977000	19.737000	45.360000





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 5, sin valores
atípicos.
Valores del documento csv.

         X0     X1     X2      X3     X4     X5     X6     X7     X8      X9  \
0     8.219  9.589  6.849   5.479  1.370  6.849  2.740  6.849  1.370   5.479
1    11.675  6.091  5.584   2.030  0.000  6.091  6.091  3.553  1.523   4.061
2     9.091  6.061  4.762   6.494  2.165  6.494  3.463  6.494  3.896   5.628
3     4.895  4.196  2.797   1.399  7.692  1.399  2.797  6.993  0.699  11.189
5    11.184  5.921  3.947   6.579  0.658  5.263  3.947  5.263  1.974   9.211
..      …      …      …       …      …      …      …      …      …
992  10.526  6.140  4.386   3.509  3.509  5.263  4.386  2.632  0.877   6.140
993   4.375  5.000  3.750   6.250  7.500  7.500  4.375  8.125  1.875   8.125
995   4.950  3.960  3.960   3.960  4.950  5.941  1.980  2.970  1.980   5.941
998   7.463  4.975  4.478  10.945  1.493  4.478  3.980  8.955  2.985   2.985
999   6.364  1.818  2.727  10.000  3.636  8.182  4.545  6.364  4.545   8.182

     …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
0    …   6.849  2.740  2.740  5.479   8.219  9.589  0.000  4.110  1.370
1    …   5.076  4.569  1.015  7.107  12.690  6.599  1.015  4.061  5.076
2    …   9.091  1.732  5.195  2.597   5.195  3.896  0.000  3.896  4.762
3    …   0.699  2.098  9.091  4.895   2.797  2.797  3.497  4.196  9.091
5    …   5.263  1.974  3.289  9.211   4.605  3.947  1.974  1.974  4.605
..   …     …      …      …      …       …      …      …      …
992  …  10.526  1.754  4.386  5.263   7.895  3.509  1.754  5.263  5.263
```

```
993  …   5.625  4.375  6.250  5.625   4.375  1.250  0.000  3.125  9.375
995  …   7.921  2.970  7.921  2.970   8.911  4.950  1.980  3.960  6.931
998  …   3.483  1.990  1.990  6.468   5.970  5.970  0.498  1.990  7.463
999  …   9.091  1.818  4.545  5.455   5.455  4.545  0.000  1.818  1.818

          X20
0    efectores
1    efectores
2    efectores
3    efectores
5    efectores
..         …
992  efectores
993  efectores
995  efectores
998  efectores
999  efectores

[822 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  822.000000  822.000000  822.000000  822.000000  822.000000  822.000000
mean     7.004095    6.026370    4.326073    5.331926    2.372567    6.343006
std      2.472111    2.258554    1.623201    1.949455    1.574706    2.523001
min      0.699000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.275250    4.533750    3.267500    4.124000    1.255000    4.800750
50%      6.801500    5.933500    4.257500    5.353000    2.080000    6.214000
75%      8.333000    7.224750    5.233000    6.593000    3.254250    7.649500
max     15.000000   13.636000   10.000000   12.069000    8.059000   14.894000

                X6          X7          X8          X9         X10         X11  \
count  822.000000  822.000000  822.000000  822.000000  822.000000  822.000000
mean     3.791609    5.428953    2.340365    5.709288    9.164481    5.847466
std      1.559901    2.116610    1.200688    2.038464    2.630299    2.480508
min      0.000000    0.000000    0.000000    0.000000    0.769000    0.000000
25%      2.712500    3.967250    1.523000    4.377750    7.426750    4.188500
50%      3.643500    5.199000    2.247000    5.621000    9.130000    5.503500
75%      4.678000    6.748000    3.060000    6.961000   10.817750    7.286250
max      9.639000   14.173000    6.522000   12.181000   17.582000   14.167000

               X12         X13         X14         X15         X16         X17  \
count  822.000000  822.000000  822.000000  822.000000  822.000000  822.000000
mean     2.731796    4.460372    4.510749    7.772900    5.604024    1.249560
```

std	1.201618	1.906532	1.980467	2.660335	1.982111	0.913449
min	0.000000	0.000000	0.000000	0.000000	0.833000	0.000000
25%	1.899000	3.074000	3.193250	6.061000	4.324250	0.581000
50%	2.627500	4.292000	4.374000	7.462500	5.518500	1.119500
75%	3.371000	5.631750	5.584250	9.331250	6.627500	1.818000
max	6.977000	10.619000	12.791000	16.080000	12.405000	4.225000

	X18	X19
count	822.000000	822.000000
mean	3.271982	6.712318
std	1.539927	2.063785
min	0.000000	1.010000
25%	2.217500	5.436750
50%	3.116000	6.641500
75%	4.222000	7.899500
max	8.511000	13.750000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
0	5.769	5.769	2.885	3.846	3.846	8.654	4.808	8.654	0.962	1.923	
1	2.985	13.433	1.493	5.970	1.493	2.985	0.000	7.463	0.000	5.970	
4	4.603	5.439	7.113	5.439	3.347	5.021	5.021	5.439	3.347	6.276	
5	4.396	8.242	5.495	7.692	0.549	9.890	1.099	3.297	1.648	6.044	
6	5.085	4.661	4.661	3.814	2.542	5.932	4.661	2.119	1.695	6.780	
..	…	…	…	…	…	…	…	…	…		
992	2.212	4.425	6.195	5.752	6.637	7.080	4.867	6.637	2.212	5.310	
996	11.664	2.058	4.117	2.058	1.372	4.631	3.431	8.576	2.573	6.175	
997	3.971	7.942	3.971	2.527	1.083	3.610	3.249	2.166	3.610	7.581	
998	5.438	7.162	4.111	8.488	1.194	12.467	4.642	2.520	3.448	3.581	
999	5.201	5.795	4.755	4.606	3.863	5.498	3.120	2.972	1.932	7.429	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
0	…	11.538	0.962	1.923	4.808	4.808	2.885	1.923	1.923	8.654	
1	…	2.985	4.478	2.985	5.970	16.418	8.955	2.985	0.000	7.463	
4	…	6.695	1.255	7.113	5.858	5.439	2.929	2.092	2.092	5.021	
5	…	6.593	2.198	5.495	7.692	7.143	4.945	1.099	0.549	8.242	
6	…	4.661	2.119	6.356	0.847	8.475	5.932	2.119	5.932	11.017	
..	…	…	…	…	…	…	…	…	…		
992	…	6.637	0.885	8.850	4.867	5.752	3.097	0.885	3.540	2.655	
996	…	4.288	3.774	6.003	5.146	6.690	6.690	3.259	3.945	6.175	
997	…	5.054	2.527	6.137	4.332	10.108	7.942	1.805	3.971	9.025	
998	…	8.223	2.255	2.653	2.785	8.090	5.570	0.133	0.663	6.499	

```
999  …   5.944  2.377  6.686  5.052   8.618  5.349  0.743  4.309   5.349

              X20
0     no_efectores
1     no_efectores
4     no_efectores
5     no_efectores
6     no_efectores
..             …
992   no_efectores
996   no_efectores
997   no_efectores
998   no_efectores
999   no_efectores

[817 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

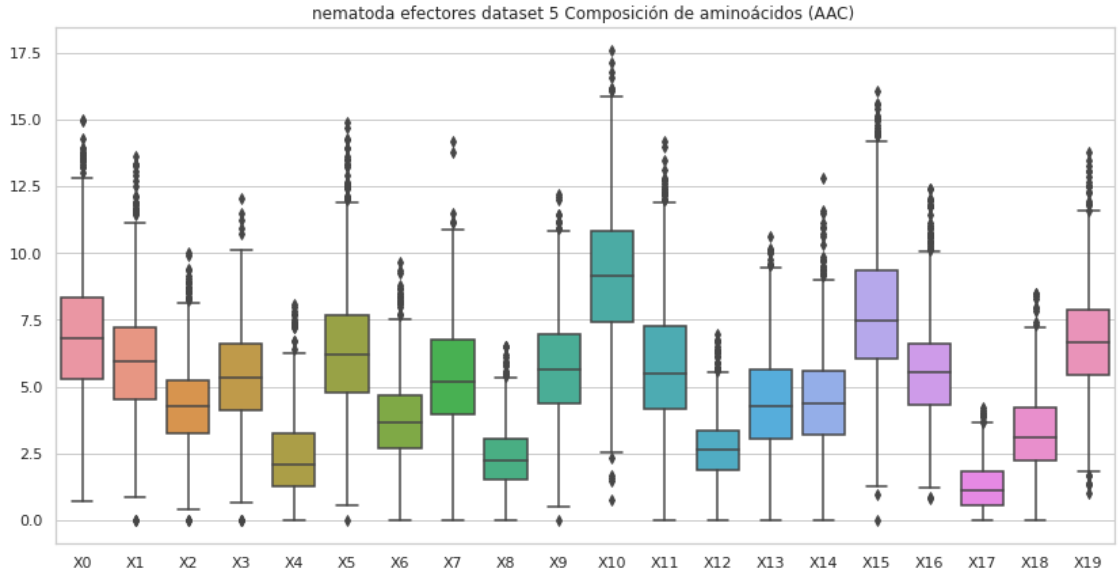
	X0	X1	X2	X3	X4	X5	\
count	817.000000	817.00000	817.000000	817.000000	817.000000	817.000000	
mean	5.876832	5.54739	5.158692	5.103823	2.058667	6.607410	
std	2.177999	2.24328	1.970040	1.821456	1.367331	2.620825	
min	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	
25%	4.462000	4.00400	3.902000	3.971000	1.159000	4.806000	
50%	5.697000	5.25100	4.902000	5.085000	1.791000	6.424000	
75%	7.173000	6.71500	6.286000	6.250000	2.740000	8.130000	
max	12.903000	13.43300	12.185000	11.111000	8.046000	15.294000	

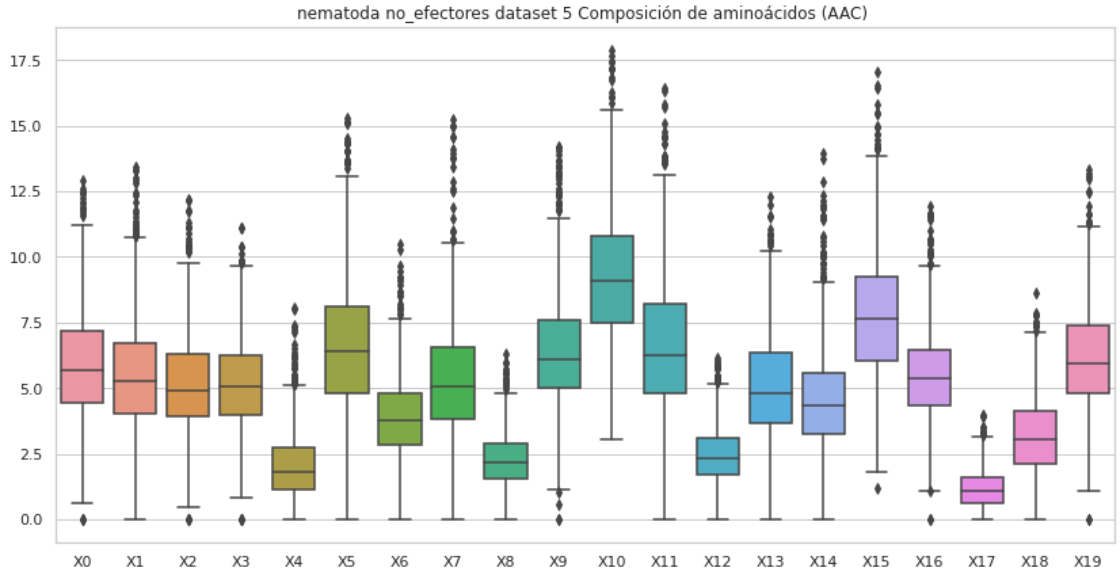
	X6	X7	X8	X9	X10	X11	\
count	817.000000	817.000000	817.000000	817.000000	817.000000	817.000000	
mean	3.904528	5.360337	2.269403	6.374517	9.281115	6.596305	
std	1.594344	2.223181	1.130016	2.312151	2.579369	2.696123	
min	0.000000	0.000000	0.000000	0.000000	3.030000	0.000000	
25%	2.857000	3.834000	1.572000	5.000000	7.479000	4.789000	
50%	3.756000	5.055000	2.174000	6.089000	9.091000	6.250000	
75%	4.781000	6.545000	2.904000	7.583000	10.773000	8.223000	
max	10.472000	15.217000	6.322000	14.182000	17.865000	16.438000	

	X12	X13	X14	X15	X16	X17	\
count	817.000000	817.000000	817.000000	817.000000	817.000000	817.000000	
mean	2.495771	5.054556	4.571723	7.852166	5.456933	1.158524	
std	1.085004	2.041903	2.107166	2.506387	1.794483	0.791242	
min	0.000000	0.000000	0.000000	1.176000	0.000000	0.000000	
25%	1.729000	3.672000	3.267000	6.061000	4.317000	0.599000	

50%	2.343000	4.819000	4.327000	7.629000	5.349000	1.064000
75%	3.101000	6.330000	5.592000	9.259000	6.468000	1.626000
max	6.154000	12.308000	13.947000	17.063000	11.940000	4.000000

	X18	X19
count	817.000000	817.000000
mean	3.185783	6.085552
std	1.579238	2.025476
min	0.000000	0.000000
25%	2.092000	4.792000
50%	3.050000	5.946000
75%	4.110000	7.381000
max	8.605000	13.333000





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 5, con valores atípicos.
Valores del documento csv.

          X0        X1        X2        X3        X4        X5        X6  \
0    0.027899  0.004650  0.018599  0.023249  0.009300  0.023249  0.004650
1    0.019051  0.000000  0.003313  0.009939  0.001657  0.005798  0.002485
2    0.056395  0.013427  0.040282  0.040282  0.032225  0.040282  0.024169
3    0.007599  0.011942  0.002171  0.002171  0.014113  0.010856  0.001086
4    0.067795  0.022598  0.045197  0.045197  0.045197  0.112992  0.045197
..        …         …         …         …         …         …         …
995  0.034298  0.034298  0.027439  0.041158  0.054878  0.020579  0.013719
996  1.341346  1.341346  0.000000  0.894231  0.447115  0.447115  2.682693
997  0.003871  0.002323  0.010839  0.011614  0.002323  0.008517  0.000774
998  0.025377  0.005075  0.037219  0.015226  0.006767  0.030452  0.010151
999  0.034715  0.019837  0.054551  0.044633  0.024796  0.034715  0.024796

          X7        X8        X9  …       X74       X75       X76  \
0    0.018599  0.023249  0.013949  …  0.029724  0.020103  0.014667
1    0.006626  0.008283  0.009939  …  0.020061  0.018223  0.025399
2    0.034911  0.056395  0.056395  … -0.037810  0.014358  0.017723
3    0.017370  0.001086  0.026055  …  0.020337  0.012712  0.000892
4    0.045197  0.045197  0.067795  …  0.015630  0.023208 -0.075167
..        …         …         …  …  …       …         …         …
995  0.041158  0.054878  0.075457  …  0.042289  0.056003  0.008943
996  1.341346  0.894231  4.024039  … -1.872802 -0.882339 -0.272339
997  0.001548  0.021679  0.009291  …  0.011184  0.028965  0.001163
998  0.010151  0.011842  0.038911  …  0.011218  0.010129  0.024436
999  0.044633  0.049592  0.049592  …  0.054240  0.040891  0.013750

          X77       X78       X79       X80       X81       X82         X83
0   -0.010200 -0.020339 -0.004709  0.023236  0.037944  0.022577  efectores
1    0.006203  0.010530  0.019505  0.010735  0.006658  0.020965  efectores
2   -0.001161 -0.019031  0.008970  0.019129 -0.003418  0.003198  efectores
3    0.021824  0.018224  0.002189  0.013355  0.008544  0.007145  efectores
4    0.063708 -0.095869  0.106484  0.053208 -0.052168  0.029389  efectores
..        …         …         …         …         …         …         …
995  0.053761  0.002044  0.016202  0.087139  0.049992  0.036953  efectores
996 -3.287804  0.991441  0.420859 -3.805288 -1.052797  0.357655  efectores
997  0.004783  0.024534  0.004498  0.011659  0.027195  0.001361  efectores
```

```
998  0.011420  0.012248  0.040087 -0.012459  0.016460  0.023053  efectores
999 -0.020650  0.063863  0.009877 -0.000730 -0.021471  0.018012  efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.040435	0.016485	0.030853	0.036665	0.028159	
std	0.057135	0.047162	0.030757	0.038555	0.038004	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023694	0.004832	0.015858	0.019622	0.012318	
50%	0.033475	0.010359	0.026636	0.030669	0.021376	
75%	0.045514	0.019026	0.039735	0.045734	0.033433	
max	1.341346	1.341346	0.728651	0.894231	0.874381	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.031720	0.017514	0.035952	0.034676	0.059353	…	
std	0.030584	0.087618	0.056489	0.043741	0.144410	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.017160	0.005553	0.017473	0.017997	0.028940	…	
50%	0.027046	0.010483	0.028205	0.028235	0.045111	…	
75%	0.039121	0.018620	0.041399	0.041273	0.067057	…	
max	0.582921	2.682693	1.341346	0.894231	4.024039	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.012261	0.000778	0.006971	0.011870	-0.005340	
std	0.063154	0.090762	0.062944	0.034870	0.137029	
min	-0.189914	-1.872802	-0.882339	-0.272339	-3.287804	
25%	-0.000232	-0.010868	-0.003900	-0.000282	-0.011586	
50%	0.010847	0.005059	0.007999	0.010696	0.003229	
75%	0.021590	0.017761	0.020878	0.023515	0.015659	
max	1.746408	1.707968	1.403298	0.644221	0.323974	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.004658	0.009814	-0.003509	0.005095	0.009913
std	0.057523	0.041545	0.131813	0.053364	0.046933
min	-1.091751	-0.647420	-3.805288	-1.052797	-1.055566
25%	-0.006052	-0.000960	-0.013039	-0.005527	0.000075
50%	0.006823	0.010683	0.003868	0.007151	0.010920
75%	0.018067	0.022060	0.016690	0.020201	0.022747

```
max       0.991441     0.420859     0.444002     0.230821     0.370312

[8 rows x 83 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.015827  0.010551  0.010551  0.023740  0.005276  0.023740  0.002638
1    0.018255  0.009127  0.036509  0.018255  0.018255  0.045636  0.000000
2    0.010072  0.000000  0.005424  0.003874  0.027893  0.016271  0.005424
3    0.112818  0.000000  0.084614  0.126920  0.042307  0.056409  0.028205
4    0.045153  0.032839  0.053363  0.049258  0.069782  0.053363  0.032839
..        …         …         …         …         …         …         …
995  0.046576  0.062101  0.000000  0.015525  0.031051  0.046576  0.000000
996  0.036036  0.004240  0.006359  0.014309  0.018548  0.026497  0.007949
997  0.017023  0.004643  0.010833  0.015476  0.026309  0.009286  0.015476
998  0.013940  0.003060  0.021760  0.031960  0.006800  0.006460  0.008840
999  0.030673  0.022786  0.027167  0.032426  0.039437  0.017527  0.011393

           X7        X8        X9  …       X74       X75       X76  \
0    0.005276  0.031654  0.036929  …  0.011567  0.015001 -0.000051
1    0.036509  0.018255  0.036509  …  0.024911  0.049238  0.051454
2    0.015496  0.006198  0.036415  …  0.014260  0.007749  0.001498
3    0.070511  0.070511  0.141023  … -0.147580 -0.088579  0.099396
4    0.061572  0.065677  0.102621  … -0.014228  0.076198 -0.003342
..        …         …         …  …         …         …         …
995  0.077626  0.124202  0.046576  … -0.073175 -0.074915 -0.018189
996  0.019078  0.013249  0.022788  …  0.009338  0.001193  0.007893
997  0.032499  0.021666  0.040237  …  0.002705  0.004538 -0.014060
998  0.009180  0.021080  0.025840  …  0.002047  0.008376  0.004357
999  0.043818  0.035055  0.061346  … -0.010355  0.002109  0.029650

           X77       X78       X79       X80       X81       X82           X83
0    0.025468  0.017432  0.000395 -0.029943 -0.018292 -0.004743  no_efectores
1    0.140631  0.110830 -0.000976  0.098962  0.065634  0.054569  no_efectores
2    0.021416  0.007264  0.008909  0.026126  0.009990 -0.005290  no_efectores
3    0.020090  0.010049  0.039143  0.117229  0.171067  0.000365  no_efectores
4   -0.007451 -0.004733 -0.015012 -0.008630 -0.010203  0.017614  no_efectores
..        …         …         …         …         …         …           …
995  0.061638  0.067703  0.078151  0.025717  0.031087  0.022992  no_efectores
996  0.005676 -0.003396  0.012151  0.006363 -0.003111  0.024868  no_efectores
997 -0.021468 -0.004558 -0.019784 -0.012067 -0.018943 -0.005989  no_efectores
998  0.004675  0.020065  0.011035  0.010083  0.036677  0.008674  no_efectores
```

```
999 -0.001852 -0.006586 -0.003370  0.018966  0.027324  0.002432  no_efectores

[1000 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, con valores atípicos.
Estadísticas.

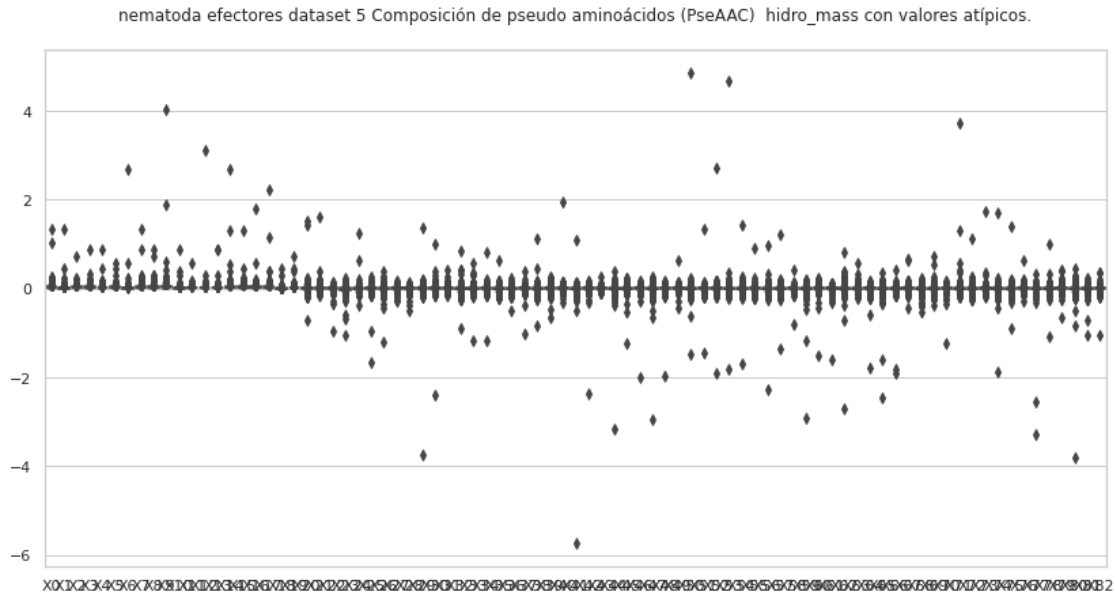
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.029724	0.012616	0.026292	0.034914	0.028079	
std	0.018716	0.015190	0.048360	0.028223	0.041990	
min	-0.062989	-0.020996	-1.245164	-0.311291	-0.933873	
25%	0.018218	0.004433	0.013314	0.017820	0.012730	
50%	0.026831	0.008609	0.023839	0.031735	0.023241	
75%	0.037609	0.016096	0.037332	0.045566	0.036851	
max	0.224062	0.213620	0.597499	0.224062	0.522812	

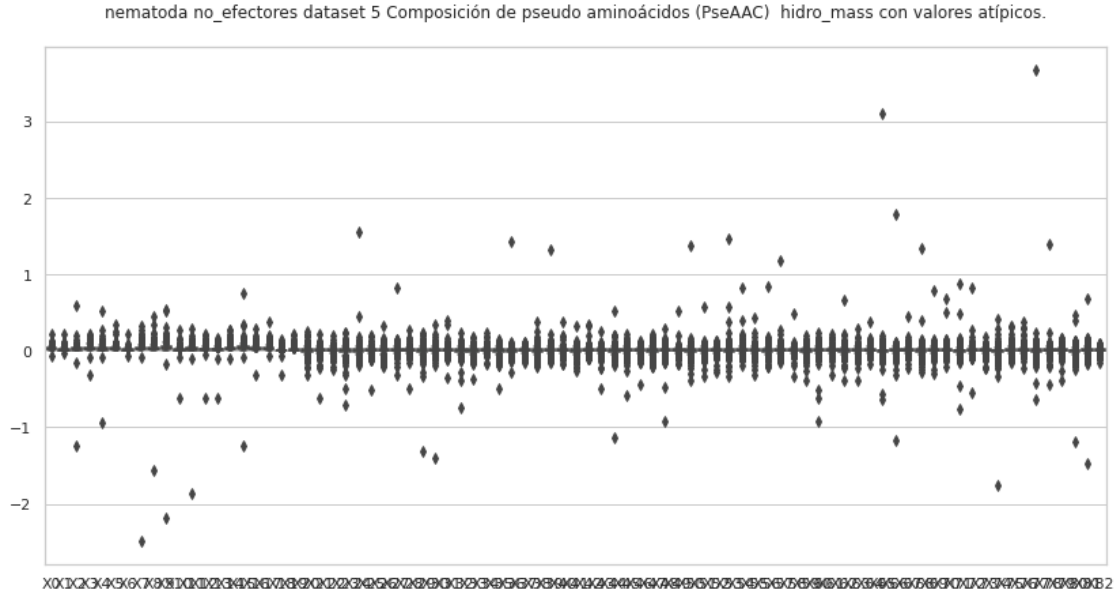
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.028445	0.012688	0.032900	0.034873	0.047585	…	
std	0.021640	0.013012	0.085049	0.059340	0.081175	…	
min	-0.000000	-0.062989	-2.490328	-1.556455	-2.179037	…	
25%	0.016058	0.005355	0.017291	0.018252	0.027212	…	
50%	0.024412	0.009805	0.029516	0.029669	0.041738	…	
75%	0.036824	0.016701	0.044780	0.045817	0.061965	…	
max	0.340794	0.224062	0.320430	0.448124	0.545270	…	

	X73	X74	X75	X76	X77	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.007677	-0.000171	0.008691	0.010734	0.004733	
std	0.025717	0.066803	0.030552	0.030219	0.122970	
min	-0.227129	-1.754576	-0.161036	-0.223598	-0.630184	
25%	-0.001888	-0.007946	-0.002486	0.000295	-0.007711	
50%	0.007996	0.004312	0.008493	0.009421	0.004684	
75%	0.018406	0.014665	0.019838	0.019207	0.016072	
max	0.179443	0.420393	0.329134	0.373696	3.677439	

	X78	X79	X80	X81	X82
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.007485	0.008495	-0.000004	0.006535	0.007410
std	0.053492	0.027518	0.053452	0.058008	0.022022
min	-0.444404	-0.394200	-1.190371	-1.471002	-0.157083
25%	-0.002568	-0.000421	-0.008998	-0.003022	-0.001366
50%	0.007404	0.008470	0.003742	0.008177	0.007601
75%	0.018729	0.019922	0.014568	0.019701	0.017949
max	1.395782	0.173296	0.462722	0.683887	0.094436

[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 5, sin valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.027899	0.004650	0.018599	0.023249	0.009300	0.023249	0.004650	
1	0.019051	0.000000	0.003313	0.009939	0.001657	0.005798	0.002485	
2	0.056395	0.013427	0.040282	0.040282	0.032225	0.040282	0.024169	
3	0.007599	0.011942	0.002171	0.002171	0.014113	0.010856	0.001086	
5	0.067353	0.003962	0.039619	0.031696	0.019810	0.031696	0.011886	
..	…	…	…	…	…	…	…	
994	0.043587	0.019372	0.024215	0.043587	0.048430	0.014529	0.072644	
995	0.034298	0.034298	0.027439	0.041158	0.054878	0.020579	0.013719	
997	0.003871	0.002323	0.010839	0.011614	0.002323	0.008517	0.000774	
998	0.025377	0.005075	0.037219	0.015226	0.006767	0.030452	0.010151	
999	0.034715	0.019837	0.054551	0.044633	0.024796	0.034715	0.024796	

	X7	X8	X9	…	X74	X75	X76	\
0	0.018599	0.023249	0.013949	…	0.029724	0.020103	0.014667	
1	0.006626	0.008283	0.009939	…	0.020061	0.018223	0.025399	
2	0.034911	0.056395	0.056395	…	-0.037810	0.014358	0.017723	
3	0.017370	0.001086	0.026055	…	0.020337	0.012712	0.000892	
5	0.055467	0.031696	0.055467	…	0.000497	-0.006285	0.050241	
..	…	…	…	…	…	…	…	
994	0.038744	0.014529	0.062959	…	-0.018344	-0.012170	0.034973	
995	0.041158	0.054878	0.075457	…	0.042289	0.056003	0.008943	
997	0.001548	0.021679	0.009291	…	0.011184	0.028965	0.001163	
998	0.010151	0.011842	0.038911	…	0.011218	0.010129	0.024436	
999	0.044633	0.049592	0.049592	…	0.054240	0.040891	0.013750	

	X77	X78	X79	X80	X81	X82	X83
0	-0.010200	-0.020339	-0.004709	0.023236	0.037944	0.022577	efectores
1	0.006203	0.010530	0.019505	0.010735	0.006658	0.020965	efectores
2	-0.001161	-0.019031	0.008970	0.019129	-0.003418	0.003198	efectores
3	0.021824	0.018224	0.002189	0.013355	0.008544	0.007145	efectores
5	-0.011383	-0.011017	0.020543	0.012983	-0.002391	0.015874	efectores
..	…	…	…	…	…	…	
994	0.003049	-0.003051	-0.006320	-0.019233	-0.046159	0.030062	efectores
995	0.053761	0.002044	0.016202	0.087139	0.049992	0.036953	efectores
997	0.004783	0.024534	0.004498	0.011659	0.027195	0.001361	efectores
998	0.011420	0.012248	0.040087	-0.012459	0.016460	0.023053	efectores
999	-0.020650	0.063863	0.009877	-0.000730	-0.021471	0.018012	efectores

[914 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	914.000000	914.000000	914.000000	914.000000	914.000000	914.000000	
mean	0.034779	0.012846	0.027720	0.031945	0.023042	0.027978	
std	0.018109	0.011870	0.016429	0.018368	0.016180	0.015783	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023092	0.004784	0.015272	0.018666	0.011847	0.016652	
50%	0.032432	0.009778	0.025379	0.029525	0.019797	0.026166	
75%	0.042650	0.017628	0.037902	0.042962	0.030609	0.035970	
max	0.132438	0.099489	0.108705	0.108120	0.130424	0.121809	

	X6	X7	X8	X9	…	X73	\
count	914.000000	914.000000	914.000000	914.000000	…	914.000000	
mean	0.012194	0.029553	0.029673	0.046570	…	0.010622	
std	0.009891	0.017964	0.018222	0.025266	…	0.020239	
min	0.000000	0.000000	0.000000	0.000000	…	-0.086336	
25%	0.005323	0.016794	0.017323	0.028004	…	0.000703	
50%	0.009703	0.026963	0.026736	0.042420	…	0.010608	
75%	0.016692	0.039075	0.038333	0.060575	…	0.020693	
max	0.078254	0.129414	0.124436	0.164307	…	0.116095	

	X74	X75	X76	X77	X78	X79	\
count	914.000000	914.000000	914.000000	914.000000	914.000000	914.000000	
mean	0.003808	0.009041	0.012025	0.002254	0.006514	0.011143	
std	0.027814	0.023654	0.019962	0.027138	0.023643	0.021053	
min	-0.171915	-0.102065	-0.073855	-0.125726	-0.127607	-0.111034	
25%	-0.008360	-0.002271	0.000981	-0.010374	-0.003976	0.000496	
50%	0.005736	0.008479	0.010754	0.003678	0.007125	0.010911	
75%	0.017254	0.020670	0.022802	0.014892	0.017498	0.021441	
max	0.146413	0.131924	0.097831	0.117510	0.142096	0.107562	

	X80	X81	X82
count	914.000000	914.000000	914.000000
mean	0.003301	0.008710	0.011458
std	0.028723	0.023649	0.020913
min	-0.172597	-0.135730	-0.096708
25%	-0.009794	-0.003179	0.001072
50%	0.004582	0.007465	0.011032
75%	0.016585	0.019468	0.022134
max	0.152384	0.127319	0.117017

[8 rows x 83 columns]

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 5, sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.015827  0.010551  0.010551  0.023740  0.005276  0.023740  0.002638
2    0.010072  0.000000  0.005424  0.003874  0.027893  0.016271  0.005424
4    0.045153  0.032839  0.053363  0.049258  0.069782  0.053363  0.032839
5    0.025648  0.003206  0.044884  0.057708  0.032060  0.019236  0.009618
6    0.028326  0.014163  0.021244  0.033047  0.035407  0.011802  0.009442
..        …         …         …         …         …         …         …
994  0.032841  0.003649  0.007298  0.014596  0.014596  0.003649  0.007298
996  0.036036  0.004240  0.006359  0.014309  0.018548  0.026497  0.007949
997  0.017023  0.004643  0.010833  0.015476  0.026309  0.009286  0.015476
998  0.013940  0.003060  0.021760  0.031960  0.006800  0.006460  0.008840
999  0.030673  0.022786  0.027167  0.032426  0.039437  0.017527  0.011393

           X7        X8        X9  …       X74       X75       X76  \
0    0.005276  0.031654  0.036929  …  0.011567  0.015001 -0.000051
2    0.015496  0.006198  0.036415  …  0.014260  0.007749  0.001498
4    0.061572  0.065677  0.102621  … -0.014228  0.076198 -0.003342
5    0.035266  0.038472  0.044884  …  0.011231  0.046439  0.017936
6    0.037768  0.025965  0.059012  …  0.031552  0.016776  0.009617
..        …         …         …  …         …         …         …
994  0.014596  0.058383  0.032841  …  0.057400  0.049834  0.022388
996  0.019078  0.013249  0.022788  …  0.009338  0.001193  0.007893
997  0.032499  0.021666  0.040237  …  0.002705  0.004538 -0.014060
998  0.009180  0.021080  0.025840  …  0.002047  0.008376  0.004357
999  0.043818  0.035055  0.061346  … -0.010355  0.002109  0.029650

           X77       X78       X79       X80       X81       X82           X83
0    0.025468  0.017432  0.000395 -0.029943 -0.018292 -0.004743  no_efectores
2    0.021416  0.007264  0.008909  0.026126  0.009990 -0.005290  no_efectores
4   -0.007451 -0.004733 -0.015012 -0.008630 -0.010203  0.017614  no_efectores
5   -0.004473 -0.005313  0.034740  0.026029  0.046687 -0.001784  no_efectores
6    0.020321  0.039154  0.012650 -0.018525 -0.031008 -0.009717  no_efectores
..        …         …         …         …         …         …           …
994 -0.005183 -0.010517 -0.001094  0.024768  0.026250  0.004364  no_efectores
996  0.005676 -0.003396  0.012151  0.006363 -0.003111  0.024868  no_efectores
997 -0.021468 -0.004558 -0.019784 -0.012067 -0.018943 -0.005989  no_efectores
998  0.004675  0.020065  0.011035  0.010083  0.036677  0.008674  no_efectores
999 -0.001852 -0.006586 -0.003370  0.018966  0.027324  0.002432  no_efectores

[880 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 5, sin valores atípicos.
Estadísticas.

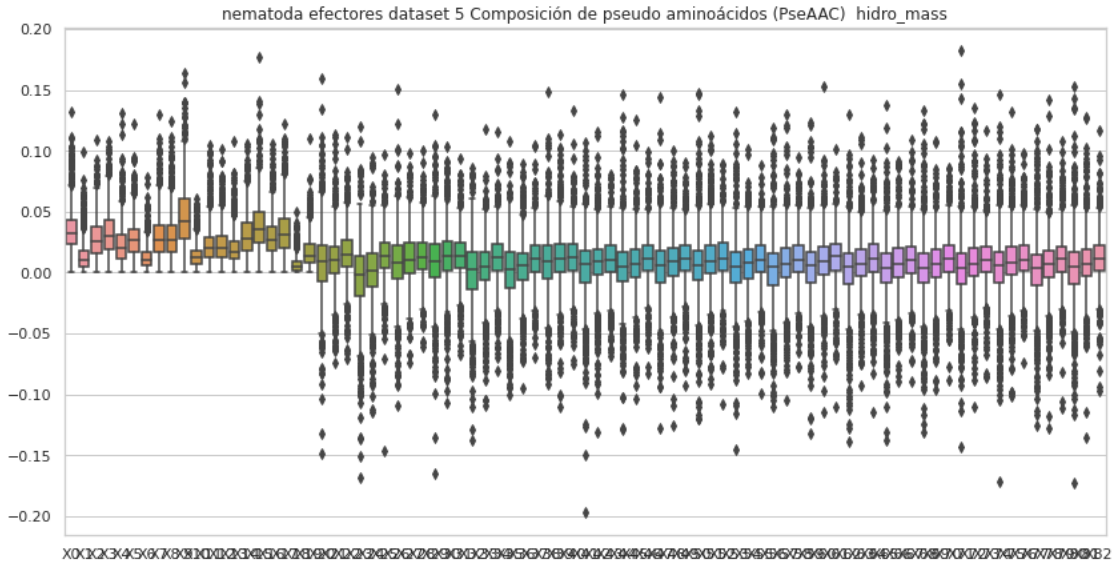
	X0	X1	X2	X3	X4	X5	\
count	880.000000	880.000000	880.000000	880.000000	880.000000	880.000000	
mean	0.027172	0.010232	0.024455	0.031533	0.024333	0.025373	
std	0.013607	0.008674	0.014935	0.018932	0.016437	0.013484	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.017781	0.004239	0.012427	0.016905	0.012036	0.015337	
50%	0.026006	0.007975	0.022279	0.030036	0.021606	0.023188	
75%	0.035319	0.014433	0.033701	0.042403	0.032725	0.033759	
max	0.085786	0.055285	0.078408	0.102239	0.113564	0.079406	

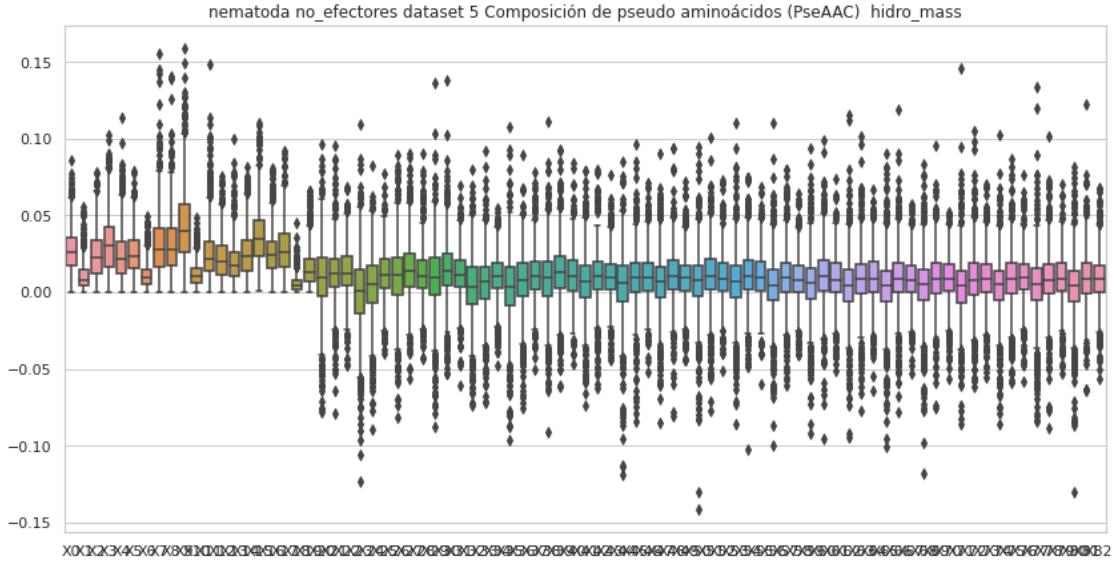
	X6	X7	X8	X9	…	X73	\
count	880.000000	880.000000	880.000000	880.000000	…	880.000000	
mean	0.011100	0.031005	0.031370	0.043672	…	0.009192	
std	0.008320	0.020556	0.019430	0.025199	…	0.016573	
min	0.000000	0.000000	0.000000	0.000000	…	-0.055815	
25%	0.005131	0.016399	0.017158	0.026375	…	-0.000327	
50%	0.009276	0.027458	0.027912	0.039570	…	0.008887	
75%	0.015128	0.041493	0.041743	0.057314	…	0.018169	
max	0.049888	0.155635	0.141021	0.158757	…	0.070383	

	X74	X75	X76	X77	X78	X79	\
count	880.000000	880.000000	880.000000	880.000000	880.000000	880.000000	
mean	0.003373	0.009081	0.009870	0.003826	0.007876	0.009411	
std	0.019978	0.018166	0.015491	0.021668	0.018676	0.015457	
min	-0.086018	-0.059194	-0.058255	-0.084451	-0.088091	-0.053122	
25%	-0.005523	-0.000498	0.001361	-0.005573	-0.001135	0.000722	
50%	0.004832	0.008841	0.009518	0.004804	0.007635	0.008522	
75%	0.014189	0.019045	0.018561	0.015507	0.018167	0.018886	
max	0.102175	0.087220	0.076447	0.133640	0.101280	0.069969	

	X80	X81	X82
count	880.000000	880.000000	880.000000
mean	0.002857	0.008839	0.009209
std	0.020824	0.017948	0.016344
min	-0.130190	-0.064567	-0.056180
25%	-0.006452	-0.001063	0.000233
50%	0.004348	0.008763	0.008536
75%	0.014056	0.019204	0.017792
max	0.081326	0.122153	0.067652

[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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.058835  0.009806  0.039223  0.049029  0.019612  0.049029  0.009806
1    0.032546  0.000000  0.005660  0.016980  0.002830  0.009905  0.004245
2    0.068885  0.016401  0.049203  0.049203  0.039363  0.049203  0.029522
3    0.027721  0.043561  0.007920  0.007920  0.051481  0.039601  0.003960
4    0.060968  0.020323  0.040645  0.040645  0.040645  0.101613  0.040645
..        …         …         …         …         …         …         …
995  0.037389  0.037389  0.029911  0.044867  0.059823  0.022433  0.014956
996  0.058909  0.058909  0.000000  0.039273  0.019636  0.019636  0.117818
997  0.021225  0.012735  0.059430  0.063675  0.012735  0.046695  0.004245
998  0.028331  0.005666  0.041552  0.016998  0.007555  0.033997  0.011332
999  0.041687  0.023821  0.065508  0.053598  0.029776  0.041687  0.029776

           X7        X8        X9  …       X32       X33       X34  \
```

```
0    0.039223  0.049029  0.029417  … -0.034988  0.030486  0.023527
1    0.011320  0.014150  0.016980  …  0.026083  0.026544  0.037673
2    0.042643  0.068885  0.068885  …  0.019186 -0.011132  0.003562
3    0.063362  0.003960  0.095042  …  0.044825 -0.003311  0.011174
4    0.040645  0.040645  0.060968  … -0.040554  0.104387 -0.026961
..        …         …         …  …         …         …
995  0.044867  0.059823  0.082256  …  0.010917  0.054794 -0.017944
996  0.058909  0.039273  0.176727  … -0.035293 -0.069685 -0.077819
997  0.008490  0.118860  0.050940  …  0.007011  0.007048 -0.005660
998  0.011332  0.013221  0.043441  …  0.024920  0.019415  0.027977
999  0.053598  0.059553  0.059553  … -0.003966  0.017891  0.013172

          X35       X36       X37       X38       X39       X40        X41
0    0.018827  0.002489 -0.020628  0.030929 -0.009930  0.047612  efectores
1    0.035345  0.028884  0.048376  0.043390  0.033323  0.035816  efectores
2    0.012287 -0.021127  0.005456  0.021648  0.010956  0.003906  efectores
3    0.004007  0.017947  0.018345  0.003254  0.007985  0.026064  efectores
4   -0.032694 -0.050100 -0.033249 -0.067597  0.095760  0.026429  efectores
..        …         …         …         …         …         …
995 -0.027123  0.019800  0.040458  0.009748  0.017662  0.040283  efectores
996  0.027427 -0.053674  0.076699 -0.011961  0.018483  0.015707  efectores
997 -0.001566  0.003026 -0.028376  0.006374  0.024661  0.007463  efectores
998  0.043491  0.019419  0.013942  0.027281  0.044753  0.025736  efectores
999  0.038533  0.004371  0.028595  0.016511  0.011861  0.021629  efectores

[1000 rows x 42 columns]
```

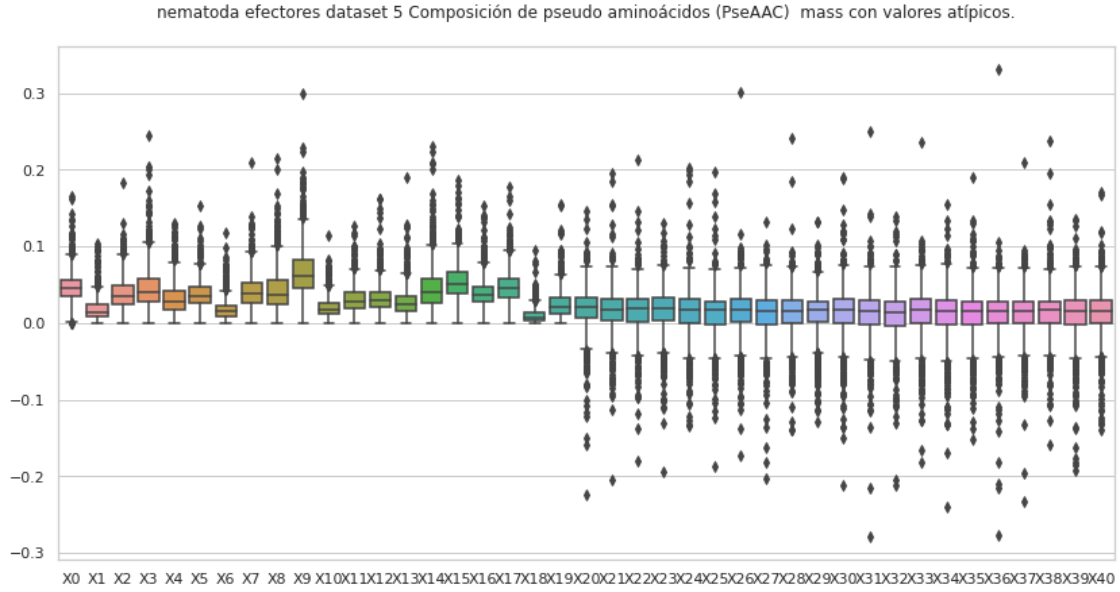
Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

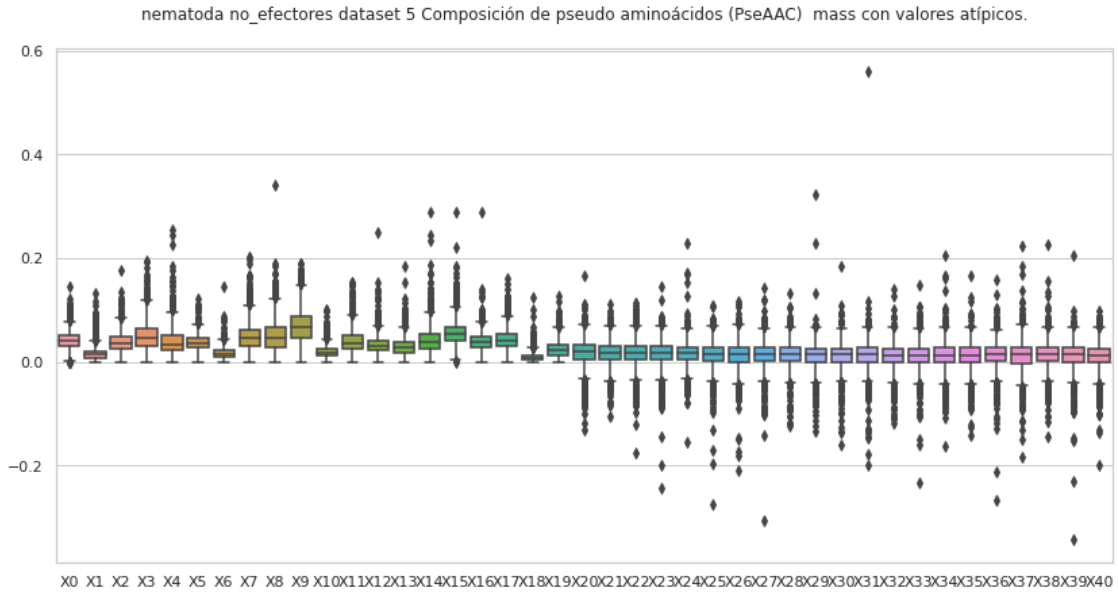
```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.047331     0.017438     0.037774     0.045936     0.032629
std       0.020025     0.015088     0.020580     0.028795     0.021000
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.034743     0.007658     0.023517     0.027163     0.017948
50%       0.045003     0.013910     0.035607     0.040625     0.028457
75%       0.056880     0.023574     0.049700     0.058594     0.042765
max       0.165505     0.104515     0.183241     0.245301     0.130788

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037479     0.017082     0.041242     0.042835     0.065843  …
std       0.016874     0.013057     0.022060     0.027454     0.032019  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.026706     0.008631     0.025559     0.024460     0.044818  …
```

```
50%       0.035537     0.014922     0.038312     0.037061     0.062119  …
75%       0.046841     0.022409     0.052401     0.055292     0.081841  …
max       0.153174     0.117818     0.209017     0.214290     0.299246  …

               X31          X32          X33          X34          X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.011637     0.011154     0.013938     0.012204     0.011182
std       0.031644     0.030504     0.030995     0.032145     0.030201
min      -0.278969    -0.211915    -0.181696    -0.240852    -0.152556
25%      -0.001727    -0.003367    -0.000097    -0.002446    -0.002718
50%       0.015298     0.014543     0.017663     0.015604     0.014791
75%       0.028840     0.028175     0.030587     0.029902     0.027339
max       0.249142     0.138615     0.236075     0.155085     0.190821

               X36          X37          X38          X39          X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012279     0.012267     0.014322     0.011360     0.013013
std       0.032686     0.029988     0.030498     0.033272     0.030831
min      -0.276893    -0.233347    -0.158337    -0.192015    -0.139844
25%      -0.000644    -0.000372    -0.000286    -0.001457     0.000086
50%       0.015613     0.015096     0.016438     0.015096     0.015274
75%       0.028490     0.028571     0.028688     0.028916     0.029471
max       0.330995     0.210329     0.238469     0.135403     0.170243

[8 rows x 41 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.038026  0.025351  0.025351  0.057039  0.012675  0.057039  0.006338
1    0.023186  0.011593  0.046371  0.023186  0.023186  0.057964  0.000000
2    0.040701  0.000000  0.021916  0.015654  0.112709  0.065747  0.021916
3    0.069145  0.000000  0.051858  0.077788  0.025929  0.034572  0.017286
4    0.040181  0.029223  0.047487  0.043834  0.062099  0.047487  0.029223
..        …         …         …         …         …         …         …
995  0.045305  0.060407  0.000000  0.015102  0.030203  0.045305  0.000000
996  0.052264  0.006149  0.009223  0.020752  0.026900  0.038429  0.011529
997  0.039779  0.010849  0.025314  0.036163  0.061477  0.021698  0.036163
998  0.035510  0.007795  0.055430  0.081413  0.017322  0.016456  0.022518
999  0.045534  0.033825  0.040330  0.048135  0.058543  0.026019  0.016912

           X7        X8        X9  …       X32       X33       X34  \
0    0.012675  0.076052  0.088727  …  0.060196  0.072220 -0.012693
```

```
1    0.046371  0.023186  0.046371  … -0.011339  0.033234 -0.000599
2    0.062616  0.025047  0.147148  …  0.004398  0.049589  0.002190
3    0.043215  0.043215  0.086431  … -0.008480  0.039966  0.018315
4    0.054793  0.058446  0.091321  …  0.028684 -0.009653  0.001141
..        …         …        …  …        …         …
995  0.075509  0.120814  0.045305  …  0.042228 -0.016642 -0.013708
996  0.027669  0.019215  0.033049  …  0.015514  0.040722  0.012889
997  0.075942  0.050628  0.094023  …  0.019029 -0.015382  0.040958
998  0.023385  0.053698  0.065823  …  0.029668  0.009344  0.010141
999  0.065048  0.052038  0.091067  … -0.000510  0.006353 -0.008993

          X35       X36       X37       X38       X39       X40           X41
0    0.016492  0.046296 -0.012834 -0.000121  0.000949 -0.011396  no_efectores
1   -0.061249  0.027347 -0.022097  0.065353 -0.001240  0.069309  no_efectores
2   -0.006026 -0.025005  0.027288  0.006053  0.035999 -0.021376  no_efectores
3   -0.000850 -0.017466 -0.063386  0.060918  0.023990  0.000224  no_efectores
4   -0.042705 -0.003968 -0.019615 -0.002974 -0.013359  0.015675  no_efectores
..        …         …         …         …         …         …           …
995 -0.052631 -0.003102  0.042043 -0.017693  0.076019  0.022365  no_efectores
996  0.014646  0.028842  0.031443  0.011448  0.017623  0.036066  no_efectores
997 -0.020306  0.004517 -0.039466 -0.032853 -0.046229 -0.013994  no_efectores
998  0.017117  0.000639  0.012343  0.011099  0.028109  0.022096  no_efectores
999 -0.000838  0.000080 -0.009406  0.044015 -0.005002  0.003610  no_efectores

[1000 rows x 42 columns]


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

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.040522     0.016526     0.037097     0.049304     0.039372
std       0.017093     0.015662     0.020287     0.029430     0.028063
min       0.000000     0.000000     0.000000     0.000000     0.000000
25%       0.030369     0.007194     0.023624     0.028968     0.021649
50%       0.039234     0.013184     0.035182     0.044798     0.033770
75%       0.049775     0.020890     0.048007     0.064901     0.050990
max       0.144431     0.131315     0.175956     0.194815     0.255066

                X5           X6           X7           X8           X9  …  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000  …
mean      0.037461     0.016976     0.048489     0.051040     0.067548  …
std       0.015713     0.012252     0.028769     0.032773     0.032801  …
min       0.000000     0.000000     0.000000     0.000000     0.000000  …
25%       0.027433     0.009115     0.029661     0.028351     0.045178  …
50%       0.035731     0.015145     0.044564     0.045241     0.065775  …
```

```
75%       0.045808      0.022874      0.061808      0.066380      0.086386  …
max       0.122200      0.144423      0.203451      0.339554      0.188748  …

                X31           X32           X33           X34           X35  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.012190     0.010772     0.011112     0.011314     0.010184
std       0.032902     0.025618     0.028408     0.028514     0.026970
min      -0.199279    -0.118596    -0.234445    -0.162978    -0.141966
25%      -0.000261    -0.001406    -0.001169    -0.001508    -0.001617
50%       0.014604     0.012132     0.012829     0.013055     0.012850
75%       0.027192     0.025764     0.026061     0.026317     0.026212
max       0.559696     0.138763     0.147701     0.205542     0.166290

                X36           X37           X38           X39           X40
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.011038     0.011130     0.012637     0.011007     0.009394
std       0.028854     0.030048     0.027633     0.030155     0.027814
min      -0.267316    -0.182691    -0.145869    -0.343173    -0.199054
25%       0.000132    -0.003092     0.000281    -0.000390    -0.002258
50%       0.013602     0.014075     0.014591     0.014425     0.012694
75%       0.026165     0.026599     0.027059     0.026876     0.025007
max       0.157818     0.222271     0.224615     0.204462     0.098304

[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 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.058835	0.009806	0.039223	0.049029	0.019612	0.049029	0.009806	
1	0.032546	0.000000	0.005660	0.016980	0.002830	0.009905	0.004245	
2	0.068885	0.016401	0.049203	0.049203	0.039363	0.049203	0.029522	
3	0.027721	0.043561	0.007920	0.007920	0.051481	0.039601	0.003960	
5	0.067416	0.003966	0.039656	0.031725	0.019828	0.031725	0.011897	
..	…	…	…	…	…	…	…	
993	0.033639	0.057666	0.048055	0.057666	0.048055	0.062472	0.014417	
995	0.037389	0.037389	0.029911	0.044867	0.059823	0.022433	0.014956	
997	0.021225	0.012735	0.059430	0.063675	0.012735	0.046695	0.004245	
998	0.028331	0.005666	0.041552	0.016998	0.007555	0.033997	0.011332	
999	0.041687	0.023821	0.065508	0.053598	0.029776	0.041687	0.029776	

	X7	X8	X9	…	X32	X33	X34	\
0	0.039223	0.049029	0.029417	…	-0.034988	0.030486	0.023527	
1	0.011320	0.014150	0.016980	…	0.026083	0.026544	0.037673	
2	0.042643	0.068885	0.068885	…	0.019186	-0.011132	0.003562	
3	0.063362	0.003960	0.095042	…	0.044825	-0.003311	0.011174	
5	0.055519	0.031725	0.055519	…	-0.004279	0.042369	0.013829	
..	…	…	…	…	…	…	…	
993	0.062472	0.043250	0.024028	…	0.013561	0.013406	-0.015610	
995	0.044867	0.059823	0.082256	…	0.010917	0.054794	-0.017944	
997	0.008490	0.118860	0.050940	…	0.007011	0.007048	-0.005660	
998	0.011332	0.013221	0.043441	…	0.024920	0.019415	0.027977	
999	0.053598	0.059553	0.059553	…	-0.003966	0.017891	0.013172	

	X35	X36	X37	X38	X39	X40	X41
0	0.018827	0.002489	-0.020628	0.030929	-0.009930	0.047612	efectores
1	0.035345	0.028884	0.048376	0.043390	0.033323	0.035816	efectores
2	0.012287	-0.021127	0.005456	0.021648	0.010956	0.003906	efectores
3	0.004007	0.017947	0.018345	0.003254	0.007985	0.026064	efectores
5	0.018401	0.036217	0.038470	0.050288	0.020562	0.015888	efectores
..	…	…	…	…	…	…	
993	0.022022	0.053473	-0.021164	0.026589	0.048437	-0.006192	efectores
995	-0.027123	0.019800	0.040458	0.009748	0.017662	0.040283	efectores
997	-0.001566	0.003026	-0.028376	0.006374	0.024661	0.007463	efectores
998	0.043491	0.019419	0.013942	0.027281	0.044753	0.025736	efectores
999	0.038533	0.004371	0.028595	0.016511	0.011861	0.021629	efectores

[821 rows x 42 columns]

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	821.000000	821.000000	821.000000	821.000000	821.000000	821.000000	
mean	0.045146	0.015917	0.035482	0.042232	0.029100	0.035707	
std	0.016030	0.011889	0.017383	0.023506	0.016683	0.013573	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.034669	0.007668	0.022989	0.026157	0.017433	0.026744	
50%	0.044224	0.013489	0.033921	0.038448	0.026387	0.034801	
75%	0.053927	0.021545	0.047375	0.054325	0.038044	0.044258	
max	0.101848	0.061534	0.098031	0.121817	0.092915	0.085841	

	X6	X7	X8	X9	…	X31	\
count	821.000000	821.000000	821.000000	821.000000	…	821.000000	
mean	0.015234	0.038349	0.038589	0.060546	…	0.014228	
std	0.009415	0.018703	0.022287	0.025400	…	0.022311	
min	0.000000	0.000000	0.000000	0.000000	…	-0.075933	
25%	0.008125	0.024763	0.023647	0.043330	…	0.001643	
50%	0.014182	0.036464	0.034024	0.058977	…	0.016453	
75%	0.020434	0.048791	0.051316	0.076821	…	0.028516	
max	0.047105	0.103429	0.124393	0.156678	…	0.092075	

	X32	X33	X34	X35	X36	X37	\
count	821.000000	821.000000	821.000000	821.000000	821.000000	821.000000	
mean	0.014678	0.017608	0.016026	0.013526	0.015654	0.014307	
std	0.021945	0.021738	0.022603	0.021456	0.022061	0.022162	
min	-0.071024	-0.078501	-0.076730	-0.075373	-0.066198	-0.075945	
25%	0.002826	0.004526	0.002330	0.001287	0.003298	0.003839	
50%	0.016593	0.018881	0.017403	0.016140	0.017176	0.015896	
75%	0.028716	0.030737	0.029858	0.026561	0.028957	0.027981	

```
max      0.080486    0.094170    0.096305    0.099518    0.104824    0.087592

              X38         X39         X40
count  821.000000  821.000000  821.000000
mean     0.015847    0.014937    0.015570
std      0.022198    0.022480    0.022387
min     -0.055596   -0.066203   -0.059861
25%      0.002734    0.002043    0.003689
50%      0.017569    0.016326    0.016211
75%      0.028110    0.028739    0.029373
max      0.102182    0.081923    0.095009

[8 rows x 41 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.038026  0.025351  0.025351  0.057039  0.012675  0.057039  0.006338
4    0.040181  0.029223  0.047487  0.043834  0.062099  0.047487  0.029223
5    0.042215  0.005277  0.073875  0.094983  0.052768  0.031661  0.015830
6    0.046472  0.023236  0.034854  0.054218  0.058090  0.019363  0.015491
7    0.041393  0.012694  0.041393  0.086098  0.025940  0.028699  0.015453
..        …         …         …         …         …         …         …
995  0.045305  0.060407  0.000000  0.015102  0.030203  0.045305  0.000000
996  0.052264  0.006149  0.009223  0.020752  0.026900  0.038429  0.011529
997  0.039779  0.010849  0.025314  0.036163  0.061477  0.021698  0.036163
998  0.035510  0.007795  0.055430  0.081413  0.017322  0.016456  0.022518
999  0.045534  0.033825  0.040330  0.048135  0.058543  0.026019  0.016912

           X7        X8        X9  …       X32       X33       X34  \
0    0.012675  0.076052  0.088727  …  0.060196  0.072220 -0.012693
4    0.054793  0.058446  0.091321  …  0.028684 -0.009653  0.001141
5    0.058045  0.063322  0.073875  … -0.024683  0.008706  0.012246
6    0.061963  0.042600  0.096817  …  0.019972  0.005770 -0.015001
7    0.048568  0.072852  0.081131  …  0.006862  0.014759  0.009555
..        …         …         …  …         …         …         …
995  0.075509  0.120814  0.045305  …  0.042228 -0.016642 -0.013708
996  0.027669  0.019215  0.033049  …  0.015514  0.040722  0.012889
997  0.075942  0.050628  0.094023  …  0.019029 -0.015382  0.040958
998  0.023385  0.053698  0.065823  …  0.029668  0.009344  0.010141
999  0.065048  0.052038  0.091067  … -0.000510  0.006353 -0.008993

           X35       X36       X37       X38       X39       X40           X41
0    0.016492  0.046296 -0.012834 -0.000121  0.000949 -0.011396  no_efectores
```

```
4   -0.042705 -0.003968 -0.019615 -0.002974 -0.013359  0.015675  no_efectores
5   -0.019064 -0.006006 -0.044173  0.029521  0.057180 -0.002936  no_efectores
6   -0.015817 -0.000468  0.027565  0.015778  0.020754 -0.015942  no_efectores
7    0.003214  0.012755  0.016226  0.003626  0.012831  0.006526  no_efectores
..        …         …         …         …         …         …
995 -0.052631 -0.003102  0.042043 -0.017693  0.076019  0.022365  no_efectores
996  0.014646  0.028842  0.031443  0.011448  0.017623  0.036066  no_efectores
997 -0.020306  0.004517 -0.039466 -0.032853 -0.046229 -0.013994  no_efectores
998  0.017117  0.000639  0.012343  0.011099  0.028109  0.022096  no_efectores
999 -0.000838  0.000080 -0.009406  0.044015 -0.005002  0.003610  no_efectores

[809 rows x 42 columns]
```

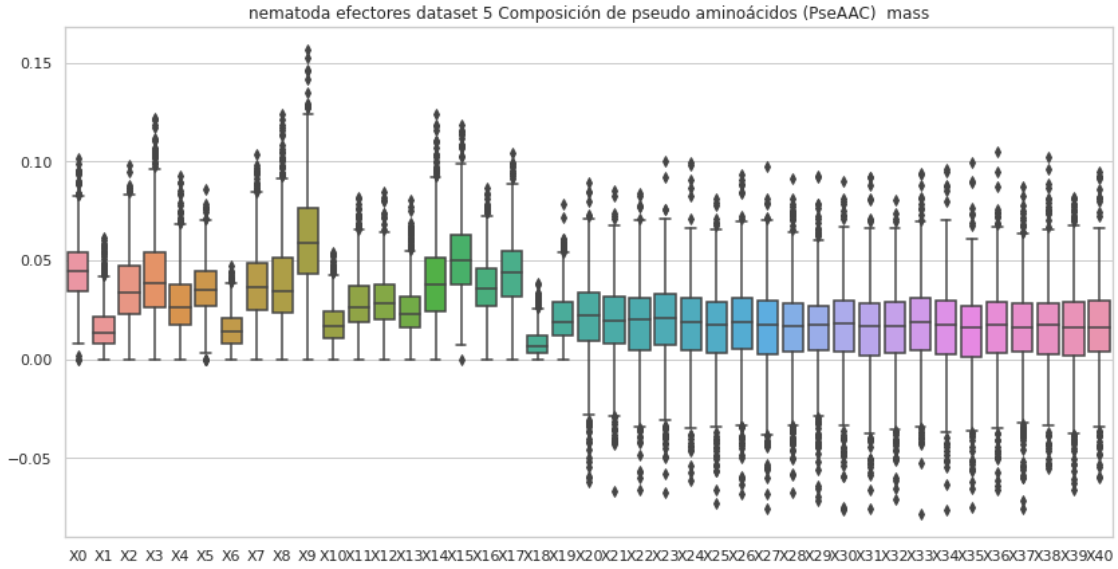
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

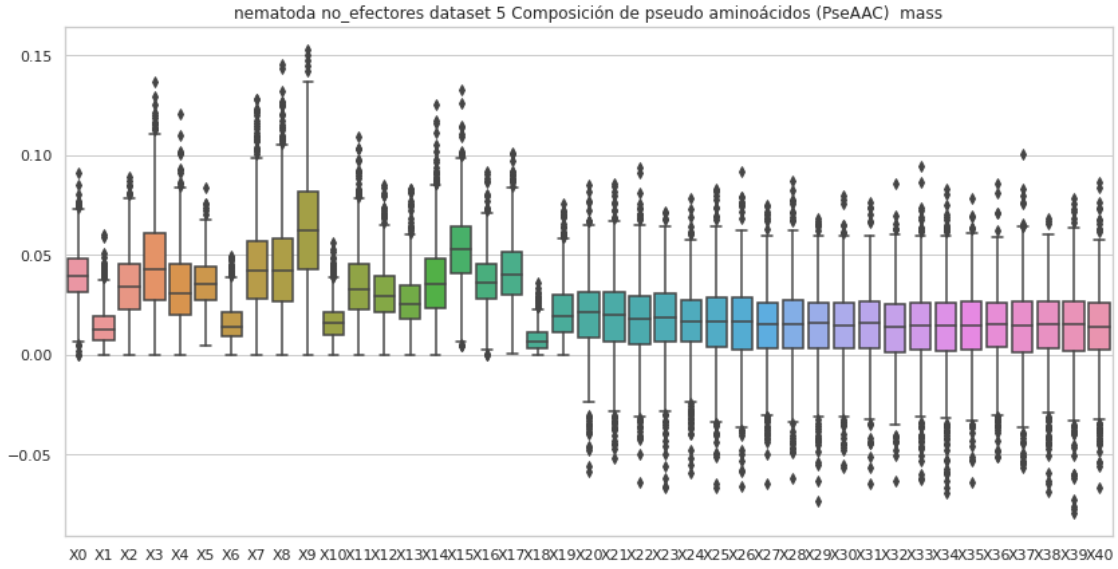
	X0	X1	X2	X3	X4	X5	\
count	809.000000	809.000000	809.000000	809.000000	809.000000	809.000000	
mean	0.039929	0.014183	0.034851	0.045526	0.034109	0.036389	
std	0.013557	0.009798	0.016892	0.024840	0.019555	0.012757	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.004537	
25%	0.031293	0.007266	0.023049	0.027247	0.020019	0.027502	
50%	0.039390	0.012555	0.034200	0.042615	0.030993	0.035244	
75%	0.048097	0.019457	0.045512	0.060741	0.045525	0.044316	
max	0.091033	0.060407	0.089095	0.136961	0.120498	0.083964	

	X6	X7	X8	X9	…	X31	\
count	809.000000	809.000000	809.000000	809.000000	…	809.000000	
mean	0.015648	0.044270	0.045295	0.062943	…	0.014578	
std	0.009550	0.023664	0.025443	0.028092	…	0.019715	
min	0.000000	0.000000	0.000000	0.000000	…	-0.064216	
25%	0.009095	0.028335	0.026733	0.043072	…	0.003321	
50%	0.014327	0.042186	0.042317	0.062201	…	0.015874	
75%	0.021292	0.057127	0.058396	0.081554	…	0.027031	
max	0.049494	0.128031	0.145470	0.152910	…	0.076705	

	X32	X33	X34	X35	X36	X37	\
count	809.000000	809.000000	809.000000	809.000000	809.000000	809.000000	
mean	0.012771	0.013935	0.013515	0.013688	0.014293	0.013608	
std	0.019338	0.020364	0.021007	0.019436	0.019098	0.021065	
min	-0.063040	-0.062188	-0.069345	-0.063520	-0.051000	-0.056573	
25%	0.001071	0.002564	0.002284	0.002572	0.003690	0.001404	
50%	0.013749	0.014701	0.014822	0.014469	0.015392	0.014997	
75%	0.025650	0.026136	0.025934	0.026795	0.026466	0.026606	
max	0.085526	0.094611	0.083327	0.078801	0.086072	0.100640	

```
              X38         X39         X40
count  809.000000  809.000000  809.000000
mean     0.013568    0.013984    0.013708
std      0.020041    0.021107    0.020331
min     -0.068171   -0.079215   -0.066732
25%      0.003626    0.002309    0.002368
50%      0.015299    0.015653    0.014340
75%      0.026489    0.027004    0.026130
max      0.068712    0.078397    0.086747

[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 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.032244	0.005374	0.021496	0.026870	0.010748	0.026870	0.005374	
1	0.032970	0.000000	0.005734	0.017202	0.002867	0.010034	0.004300	
2	0.070347	0.016749	0.050248	0.050248	0.040198	0.050248	0.030149	
3	0.008625	0.013553	0.002464	0.002464	0.016018	0.012321	0.001232	
4	0.080984	0.026995	0.053989	0.053989	0.053989	0.134973	0.053989	
..	…	…	…	…	…	…	…	
995	0.044228	0.044228	0.035383	0.053074	0.070765	0.026537	0.017691	
996	0.290962	0.290962	0.000000	0.193974	0.096987	0.096987	0.581923	
997	0.004125	0.002475	0.011551	0.012376	0.002475	0.009076	0.000825	
998	0.057114	0.011423	0.083767	0.034268	0.015230	0.068537	0.022846	
999	0.048704	0.027831	0.076534	0.062619	0.034788	0.048704	0.034788	

	X7	X8	X9	…	X53	X54	X55	\
0	0.021496	0.026870	0.016122	…	0.017732	0.011209	0.064377	
1	0.011468	0.014335	0.017202	…	0.010983	0.011291	0.005428	
2	0.043548	0.070347	0.070347	…	0.046909	0.000991	-0.001745	
3	0.019714	0.001232	0.029571	…	0.024358	0.017191	0.006320	
4	0.053989	0.053989	0.080984	…	-0.249029	0.081204	-0.074862	
..	…	…	…	…	…	…	…	
995	0.053074	0.070765	0.097302	…	-0.031718	-0.064409	-0.030646	
996	0.290962	0.193974	0.872885	…	0.159044	0.806299	0.245170	
997	0.001650	0.023102	0.009901	…	0.027422	0.006377	0.021885	
998	0.022846	0.026653	0.087574	…	0.024294	-0.046495	-0.025202	
999	0.062619	0.069577	0.069577	…	0.047302	0.022103	0.070670	

	X56	X57	X58	X59	X60	X61	X62
0	0.034353	0.023234	-0.011788	-0.023507	0.026855	0.043853	efectores
1	0.034719	0.031538	0.010736	0.018224	0.018578	0.011523	efectores
2	-0.047165	0.017910	-0.001448	-0.023740	0.023861	-0.004263	efectores
3	0.023081	0.014427	0.024769	0.020683	0.015157	0.009697	efectores
4	0.018670	0.027723	0.076102	-0.114519	0.063559	-0.062317	efectores
..	…	…	…	…	…	…	…
995	0.054532	0.072216	0.069325	0.002636	0.112367	0.064465	efectores
996	-0.406243	-0.191395	-0.713182	0.215061	-0.825434	-0.228370	efectores
997	0.011919	0.030867	0.005097	0.026144	0.012425	0.028981	efectores

```
998  0.025247  0.022797  0.025701  0.027565 -0.028040  0.037045  efectores
999  0.076098  0.057368 -0.028971  0.089598 -0.001025 -0.030124  efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.053801	0.020271	0.040185	0.046203	0.035084	
std	0.038130	0.023581	0.025367	0.028143	0.028217	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.030080	0.006566	0.021544	0.028014	0.017337	
50%	0.045838	0.014185	0.037589	0.042358	0.030096	
75%	0.066903	0.025458	0.054536	0.059709	0.045254	
max	0.453844	0.290962	0.206945	0.260310	0.378203	

	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.043881	0.019214	0.044565	0.043057	0.071707	…	
std	0.033207	0.024820	0.030824	0.028270	0.053489	…	
min	0.000000	0.000000	0.000000	0.000000	0.000000	…	
25%	0.021713	0.007290	0.024075	0.024677	0.041467	…	
50%	0.037151	0.014270	0.038565	0.037711	0.062619	…	
75%	0.057398	0.025021	0.056757	0.054860	0.090242	…	
max	0.302562	0.581923	0.290962	0.223123	0.872885	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.002090	0.004943	0.004696	0.008818	0.002009	
std	0.053935	0.044734	0.061159	0.042820	0.057418	
min	-0.437569	-0.488176	-0.311540	-0.306958	-0.616130	
25%	-0.015978	-0.007852	-0.014555	-0.007549	-0.013865	
50%	0.003465	0.008325	0.005746	0.009082	0.007026	
75%	0.020064	0.025096	0.022476	0.025546	0.023420	
max	0.227198	0.233426	0.806299	0.377979	0.485081	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008647	-0.000750	0.005715	-0.000387	0.007476
std	0.046603	0.058710	0.041486	0.071287	0.045446
min	-0.520150	-0.720278	-0.310069	-1.107197	-0.448002
25%	-0.005780	-0.016338	-0.008681	-0.018287	-0.007582
50%	0.011421	0.004499	0.009412	0.005633	0.009717
75%	0.027545	0.022070	0.023969	0.022446	0.026839

```
max       0.398552     0.429838     0.267561     0.587958     0.213285

[8 rows x 62 columns]
```

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.018444	0.012296	0.012296	0.027666	0.006148	0.027666	0.003074	
1	0.022148	0.011074	0.044296	0.022148	0.022148	0.055370	0.000000	
2	0.010202	0.000000	0.005493	0.003924	0.028252	0.016480	0.005493	
3	0.107801	0.000000	0.080851	0.121276	0.040425	0.053900	0.026950	
4	0.052668	0.038304	0.062243	0.057456	0.081395	0.062243	0.038304	
..	…	…	…	…	…	…	…	
995	0.063568	0.084758	0.000000	0.021189	0.042379	0.063568	0.000000	
996	0.058175	0.006844	0.010266	0.023099	0.029943	0.042776	0.012833	
997	0.017011	0.004639	0.010825	0.015465	0.026290	0.009279	0.015465	
998	0.016138	0.003543	0.025191	0.037000	0.007872	0.007479	0.010234	
999	0.033480	0.024871	0.029653	0.035393	0.043045	0.019131	0.012435	

	X7	X8	X9	…	X53	X54	X55	\
0	0.006148	0.036887	0.043035	…	0.035387	0.004123	0.017663	
1	0.044296	0.022148	0.044296	…	-0.009151	-0.084706	-0.060372	
2	0.015695	0.006278	0.036884	…	0.003583	0.016180	0.005631	
3	0.067376	0.067376	0.134751	…	0.079712	-0.007298	-0.014638	
4	0.071819	0.076607	0.119699	…	0.057662	0.074447	0.054964	
..	…	…	…	…	…	…	…	
995	0.105947	0.169515	0.063568	…	0.034117	-0.097267	-0.061580	
996	0.030798	0.021388	0.036787	…	-0.001825	0.011681	0.006528	
997	0.032476	0.021650	0.040208	…	0.011004	-0.000822	-0.000939	
998	0.010628	0.024404	0.029915	…	0.035688	0.024086	0.031608	
999	0.047828	0.038263	0.066959	…	-0.004803	-0.007940	0.012529	

	X56	X57	X58	X59	X60	X61	X62
0	0.013480	0.017481	0.029679	0.020314	-0.034894	-0.021317	no_efectores
1	0.030224	0.059740	0.170626	0.134469	0.120070	0.079633	no_efectores
2	0.014444	0.007849	0.021692	0.007357	0.026463	0.010118	no_efectores
3	-0.141017	-0.084639	0.019196	0.009602	0.112015	0.163459	no_efectores
4	-0.016596	0.088878	-0.008691	-0.005520	-0.010067	-0.011901	no_efectores
..	…	…	…	…	…	…	…
995	-0.099872	-0.102246	0.084126	0.092403	0.035099	0.042429	no_efectores
996	0.015075	0.001926	0.009163	-0.005483	0.010271	-0.005023	no_efectores
997	0.002703	0.004535	-0.021452	-0.004555	-0.012058	-0.018929	no_efectores
998	0.002370	0.009697	0.005412	0.023229	0.011673	0.042460	no_efectores

```
999 -0.011302  0.002302 -0.002021 -0.007189  0.020701  0.029824  no_efectores

[1000 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

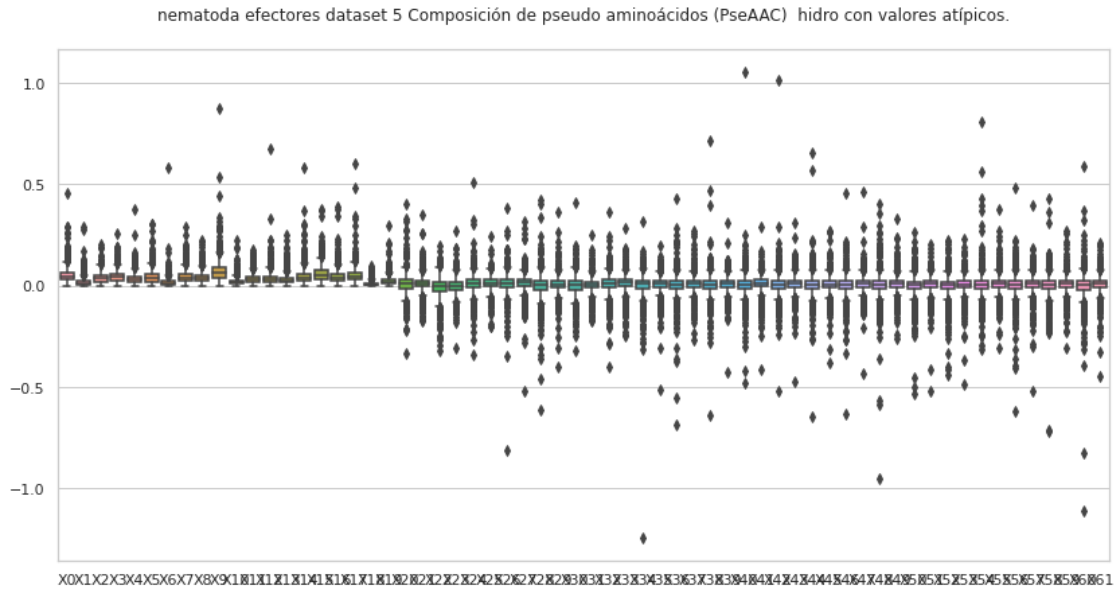
	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.040304	0.017159	0.035910	0.045126	0.037785	
std	0.028625	0.029791	0.044504	0.059639	0.065269	
min	-0.076109	-0.025370	-0.177588	-0.101479	-0.101479	
25%	0.021912	0.005678	0.017305	0.024292	0.017733	
50%	0.035555	0.011478	0.031803	0.041532	0.029959	
75%	0.051770	0.020528	0.047787	0.057249	0.046675	
max	0.344321	0.688642	1.154034	1.721606	1.721606	

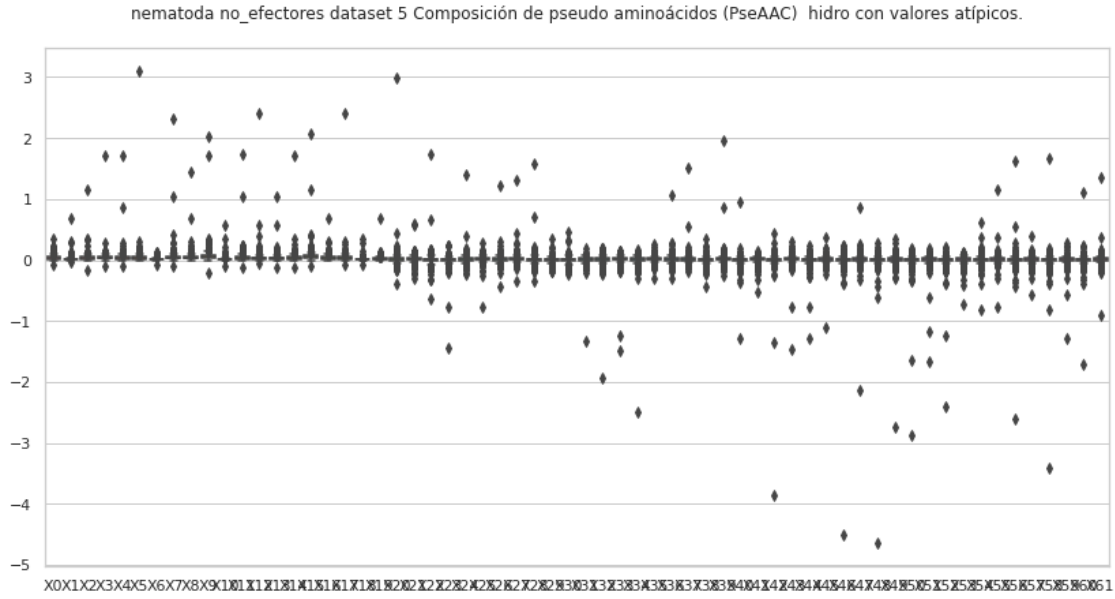
	X5	X6	X7	X8	X9	…	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	…	
mean	0.041832	0.015947	0.047029	0.046384	0.065522	…	
std	0.101277	0.014157	0.084629	0.057207	0.090867	…	
min	0.000000	-0.076109	-0.101479	0.000000	-0.202958	…	
25%	0.019074	0.006754	0.023851	0.024539	0.036075	…	
50%	0.031096	0.012716	0.038434	0.039423	0.054741	…	
75%	0.050255	0.021730	0.055439	0.057795	0.080025	…	
max	3.098891	0.118080	2.308068	1.442542	2.019559	…	

	X52	X53	X54	X55	X56	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.000955	0.006884	0.001435	0.009689	0.000906	
std	0.094667	0.042057	0.051136	0.054852	0.107010	
min	-2.400858	-0.729949	-0.806995	-0.767039	-2.602005	
25%	-0.009944	-0.003433	-0.011891	-0.003966	-0.010773	
50%	0.007080	0.011166	0.004873	0.009881	0.005341	
75%	0.020075	0.025034	0.017895	0.024031	0.018606	
max	0.208142	0.133179	0.621133	1.153628	1.626163	

	X57	X58	X59	X60	X61
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.008734	-0.001198	0.006124	0.000472	0.009708
std	0.039820	0.129464	0.056584	0.076530	0.061282
min	-0.566692	-3.408297	-1.293629	-1.718187	-0.909697
25%	-0.003775	-0.011209	-0.003699	-0.012040	-0.003448
50%	0.011163	0.006292	0.009935	0.005444	0.010922
75%	0.024377	0.020002	0.023229	0.017860	0.024864
max	0.394228	1.677702	0.289847	1.103251	1.363343

[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 5, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.032244	0.005374	0.021496	0.026870	0.010748	0.026870	0.005374	
1	0.032970	0.000000	0.005734	0.017202	0.002867	0.010034	0.004300	
2	0.070347	0.016749	0.050248	0.050248	0.040198	0.050248	0.030149	
3	0.008625	0.013553	0.002464	0.002464	0.016018	0.012321	0.001232	
5	0.111670	0.006569	0.065688	0.052550	0.032844	0.052550	0.019706	
..	…	…	…	…	…	…	…	
993	0.052148	0.089397	0.074498	0.089397	0.074498	0.096847	0.022349	
994	0.042295	0.018798	0.023497	0.042295	0.046995	0.014098	0.070492	
997	0.004125	0.002475	0.011551	0.012376	0.002475	0.009076	0.000825	
998	0.057114	0.011423	0.083767	0.034268	0.015230	0.068537	0.022846	
999	0.048704	0.027831	0.076534	0.062619	0.034788	0.048704	0.034788	

	X7	X8	X9	…	X53	X54	X55	\
0	0.021496	0.026870	0.016122	…	0.017732	0.011209	0.064377	
1	0.011468	0.014335	0.017202	…	0.010983	0.011291	0.005428	
2	0.043548	0.070347	0.070347	…	0.046909	0.000991	-0.001745	
3	0.019714	0.001232	0.029571	…	0.024358	0.017191	0.006320	
5	0.091963	0.052550	0.091963	…	0.038252	0.005102	0.010000	
..	…	…	…	…	…	…	…	
993	0.096847	0.067048	0.037249	…	-0.016959	0.044780	0.066049	
994	0.037596	0.014098	0.061093	…	0.001892	0.001416	0.018985	
997	0.001650	0.023102	0.009901	…	0.027422	0.006377	0.021885	
998	0.022846	0.026653	0.087574	…	0.024294	-0.046495	-0.025202	
999	0.062619	0.069577	0.069577	…	0.047302	0.022103	0.070670	

	X56	X57	X58	X59	X60	X61	X62
0	0.034353	0.023234	-0.011788	-0.023507	0.026855	0.043853	efectores
1	0.034719	0.031538	0.010736	0.018224	0.018578	0.011523	efectores
2	-0.047165	0.017910	-0.001448	-0.023740	0.023861	-0.004263	efectores
3	0.023081	0.014427	0.024769	0.020683	0.015157	0.009697	efectores
5	0.000825	-0.010420	-0.018873	-0.018266	0.021526	-0.003964	efectores
..	…	…	…	…	…	…	…
993	0.041701	0.100934	0.077504	0.022065	0.007629	0.011847	efectores
994	-0.017800	-0.011809	0.002958	-0.002961	-0.018663	-0.044791	efectores
997	0.011919	0.030867	0.005097	0.026144	0.012425	0.028981	efectores
998	0.025247	0.022797	0.025701	0.027565	-0.028040	0.037045	efectores
999	0.076098	0.057368	-0.028971	0.089598	-0.001025	-0.030124	efectores

[841 rows x 63 columns]

Composición de pseudo aminoácidos (PseAAC) efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	841.000000	841.000000	841.000000	841.000000	841.000000	841.000000	
mean	0.046828	0.016685	0.036361	0.041385	0.029421	0.037489	
std	0.026048	0.015101	0.020254	0.021206	0.018052	0.022644	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.028220	0.006326	0.020244	0.026329	0.016120	0.020338	
50%	0.043354	0.013267	0.035429	0.039690	0.027332	0.034168	
75%	0.060725	0.022581	0.049849	0.056453	0.039034	0.049373	
max	0.157015	0.090807	0.108412	0.124909	0.107264	0.129593	

	X6	X7	X8	X9	…	X52	\
count	841.000000	841.000000	841.000000	841.000000	…	841.000000	
mean	0.015985	0.038090	0.037691	0.061210	…	0.003368	
std	0.012092	0.021121	0.020356	0.031237	…	0.031738	
min	0.000000	0.000000	0.000000	0.001568	…	-0.142294	
25%	0.007013	0.021894	0.023351	0.037303	…	-0.011231	
50%	0.013132	0.034743	0.035348	0.057711	…	0.005169	
75%	0.022290	0.051066	0.048933	0.082237	…	0.019638	
max	0.070492	0.111452	0.122399	0.175191	…	0.136101	

	X53	X54	X55	X56	X57	X58	\
count	841.000000	841.000000	841.000000	841.000000	841.000000	841.000000	
mean	0.009082	0.003564	0.010423	0.006235	0.012110	0.004035	
std	0.026254	0.031238	0.026808	0.032088	0.026472	0.031246	
min	-0.098398	-0.164504	-0.105285	-0.122744	-0.098657	-0.129215	
25%	-0.003930	-0.011807	-0.002936	-0.009625	-0.001791	-0.011945	
50%	0.010202	0.005765	0.010646	0.007949	0.012126	0.005509	
75%	0.024607	0.019841	0.024872	0.022288	0.026039	0.021540	
max	0.098434	0.120795	0.127817	0.167283	0.113187	0.139924	

	X59	X60	X61
count	841.000000	841.000000	841.000000
mean	0.009352	0.004869	0.010644
std	0.026410	0.033775	0.027465
min	-0.108302	-0.167031	-0.119795
25%	-0.003308	-0.011226	-0.004408
50%	0.010906	0.007343	0.010414
75%	0.024055	0.022212	0.026451
max	0.125893	0.141292	0.140126

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.018444  0.012296  0.012296  0.027666  0.006148  0.027666  0.003074
1    0.022148  0.011074  0.044296  0.022148  0.022148  0.055370  0.000000
2    0.010202  0.000000  0.005493  0.003924  0.028252  0.016480  0.005493
4    0.052668  0.038304  0.062243  0.057456  0.081395  0.062243  0.038304
5    0.026280  0.003285  0.045991  0.059131  0.032851  0.019710  0.009855
..        …         …         …         …         …         …         …
994  0.035492  0.003944  0.007887  0.015774  0.015774  0.003944  0.007887
996  0.058175  0.006844  0.010266  0.023099  0.029943  0.042776  0.012833
997  0.017011  0.004639  0.010825  0.015465  0.026290  0.009279  0.015465
998  0.016138  0.003543  0.025191  0.037000  0.007872  0.007479  0.010234
999  0.033480  0.024871  0.029653  0.035393  0.043045  0.019131  0.012435

           X7        X8        X9  …       X53       X54       X55  \
0    0.006148  0.036887  0.043035  …  0.035387  0.004123  0.017663
1    0.044296  0.022148  0.044296  … -0.009151 -0.084706 -0.060372
2    0.015695  0.006278  0.036884  …  0.003583  0.016180  0.005631
4    0.071819  0.076607  0.119699  …  0.057662  0.074447  0.054964
5    0.036136  0.039421  0.045991  …  0.006492  0.005157  0.055422
..        …         …         …  …         …         …         …
994  0.015774  0.063098  0.035492  …  0.034349 -0.048518 -0.031717
996  0.030798  0.021388  0.036787  … -0.001825  0.011681  0.006528
997  0.032476  0.021650  0.040208  …  0.011004 -0.000822 -0.000939
998  0.010628  0.024404  0.029915  …  0.035688  0.024086  0.031608
999  0.047828  0.038263  0.066959  … -0.004803 -0.007940  0.012529

          X56       X57       X58       X59       X60       X61           X62
0    0.013480  0.017481  0.029679  0.020314 -0.034894 -0.021317  no_efectores
1    0.030224  0.059740  0.170626  0.134469  0.120070  0.079633  no_efectores
2    0.014444  0.007849  0.021692  0.007357  0.026463  0.010118  no_efectores
4   -0.016596  0.088878 -0.008691 -0.005520 -0.010067 -0.011901  no_efectores
5    0.011508  0.047584 -0.004583 -0.005444  0.026671  0.047838  no_efectores
..        …         …         …         …         …         …           …
994  0.062035  0.053858 -0.005601 -0.011366  0.026768  0.028369  no_efectores
996  0.015075  0.001926  0.009163 -0.005483  0.010271 -0.005023  no_efectores
997  0.002703  0.004535 -0.021452 -0.004555 -0.012058 -0.018929  no_efectores
998  0.002370  0.009697  0.005412  0.023229  0.011673  0.042460  no_efectores
999 -0.011302  0.002302 -0.002021 -0.007189  0.020701  0.029824  no_efectores

[925 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

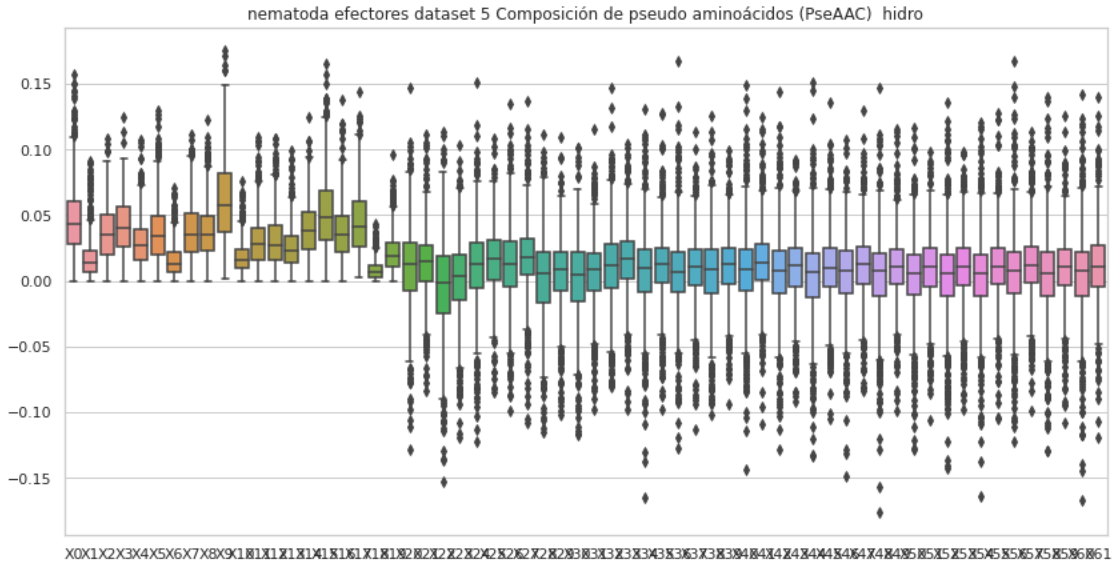
	X0	X1	X2	X3	X4	X5	\
count	925.000000	925.000000	925.000000	925.000000	925.000000	925.000000	
mean	0.037450	0.014561	0.032740	0.041440	0.032294	0.035573	
std	0.021965	0.014384	0.018967	0.023172	0.020345	0.024416	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.021243	0.005422	0.017043	0.023770	0.017480	0.018547	
50%	0.034256	0.010887	0.030927	0.040115	0.029058	0.029921	
75%	0.049575	0.019165	0.045751	0.054746	0.043555	0.047566	
max	0.125647	0.106110	0.100725	0.150878	0.122872	0.213670	

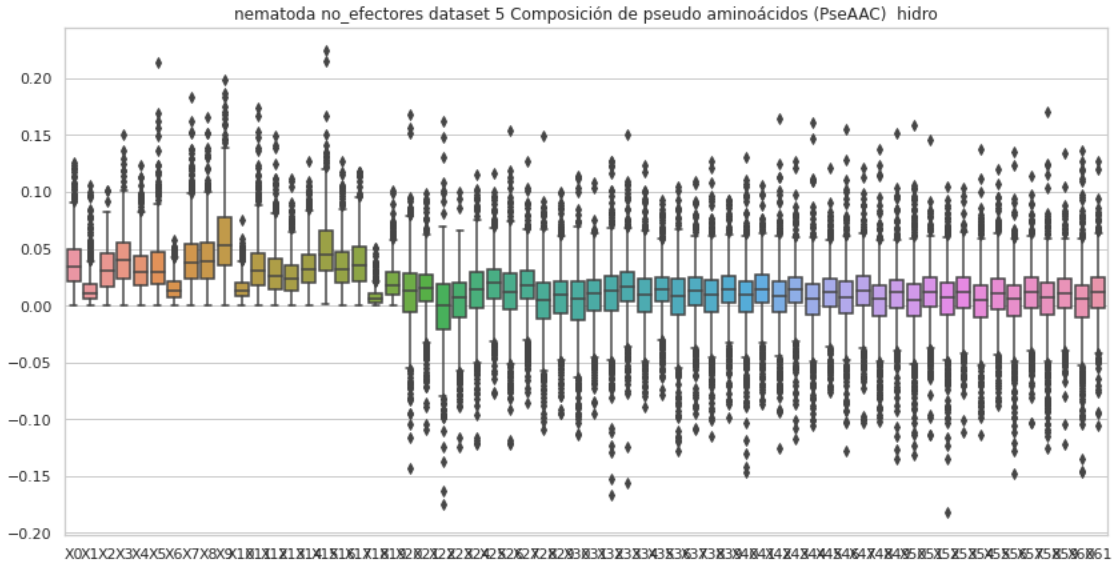
	X6	X7	X8	X9	…	X52	\
count	925.000000	925.000000	925.000000	925.000000	…	925.000000	
mean	0.014767	0.041116	0.041429	0.057964	…	0.004392	
std	0.010672	0.025410	0.023878	0.032011	…	0.028152	
min	0.000000	0.000000	0.000000	0.000000	…	-0.181861	
25%	0.006785	0.023475	0.024164	0.035216	…	-0.007589	
50%	0.012452	0.037464	0.038360	0.052931	…	0.007340	
75%	0.020875	0.053610	0.055413	0.077036	…	0.019766	
max	0.057889	0.183678	0.165598	0.198717	…	0.104335	

	X53	X54	X55	X56	X57	X58	\
count	925.000000	925.000000	925.000000	925.000000	925.000000	925.000000	
mean	0.010251	0.002650	0.009489	0.003073	0.010366	0.004217	
std	0.023849	0.027086	0.025062	0.027682	0.024141	0.029468	
min	-0.111861	-0.113424	-0.087682	-0.148392	-0.103045	-0.125622	
25%	-0.001825	-0.010427	-0.003268	-0.008917	-0.001738	-0.008070	
50%	0.011510	0.005245	0.010159	0.005686	0.011322	0.006913	
75%	0.024885	0.017491	0.023640	0.018162	0.024215	0.019604	
max	0.103376	0.137609	0.120022	0.135126	0.113857	0.170626	

	X59	X60	X61
count	925.000000	925.000000	925.000000
mean	0.009602	0.003221	0.010574
std	0.023424	0.029376	0.023863
min	-0.122006	-0.146222	-0.104971
25%	-0.002185	-0.010520	-0.002083
50%	0.010427	0.005824	0.011279
75%	0.023067	0.017433	0.024496
max	0.134469	0.136446	0.126358

[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 5,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.198306 -0.096564  0.003400 -0.028014  0.039079 -0.020450  0.109634
1    0.215057  0.099904  0.180423  0.135070  0.171050  0.230267  0.175587
2   -0.032040 -0.000362 -0.045080  0.002972 -0.051593 -0.047613  0.012504
3    0.023837 -0.034744  0.120064  0.089214  0.171561  0.033669  0.003346
4   -0.066733  0.017327  0.141441 -0.122540 -0.001408 -0.286970  0.004042
..        …         …         …         …         …         …         …
995 -0.058570 -0.044187 -0.004299 -0.009991  0.039203  0.014471  0.159008
996 -0.078046 -0.133529 -0.045082  0.026367 -0.094609 -0.084586  0.141933
997  0.108639  0.093044  0.103261  0.063113  0.104137  0.023857  0.061714
998  0.029922 -0.068364  0.033858  0.108529  0.096435  0.049748 -0.007200
999  0.017326 -0.095379 -0.136142  0.078188  0.013554  0.055849 -0.101645

           X7        X8        X9       X10       X11       X12        X13
0   -0.071913 -0.153224  0.216654  0.026026  0.221540 -0.133203  efectores
```

```
1    0.032970  0.139461  0.044398  0.175364  0.255569  0.098884  efectores
2    0.119015  0.003834  0.012017  0.063383  0.057021  0.019342  efectores
3    0.236380  0.080222  0.043116  0.054423  0.028470  0.144641  efectores
4   -0.092650 -0.127373  0.134217  0.245591 -0.054127 -0.117980  efectores
..        …         …         …         …         …         …
995 -0.006929 -0.044089 -0.041960 -0.008103  0.138331  0.019279  efectores
996  0.014520 -0.157158  0.043103  0.202248  0.013317 -0.067702  efectores
997  0.070705  0.029685 -0.055110  0.014563  0.119203  0.003765  efectores
998  0.036398  0.070076  0.034989  0.031137  0.041295  0.015967  efectores
999 -0.022477 -0.141220  0.084689  0.011441  0.159226 -0.065709  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.014910	0.008770	0.006702	0.008283	0.005633	
std	0.077829	0.072842	0.069108	0.071784	0.076672	
min	-0.730653	-0.344726	-0.271595	-0.561751	-0.371485	
25%	-0.024945	-0.033161	-0.031352	-0.029133	-0.033886	
50%	0.016749	0.007791	0.008179	0.010388	0.004396	
75%	0.054595	0.049217	0.044923	0.050431	0.045511	
max	0.609494	0.537329	0.326743	0.274181	0.686111	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	-0.000506	0.005856	-0.001235	0.000341	0.001216	
std	0.072155	0.075740	0.072899	0.073808	0.076085	
min	-0.618578	-0.469635	-0.598750	-0.346286	-0.381541	
25%	-0.038934	-0.035579	-0.043917	-0.037862	-0.037026	
50%	0.001417	0.006072	0.002389	0.001853	0.002975	
75%	0.037168	0.044491	0.041393	0.038597	0.038135	
max	0.352317	0.849472	0.303462	0.676845	0.711217	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.004567	-0.000315	-0.000987
std	0.076611	0.075857	0.075130
min	-0.352250	-0.478147	-0.664122
25%	-0.034485	-0.041081	-0.039002
50%	0.005967	-0.000503	0.001008
75%	0.045549	0.037871	0.038164
max	0.369355	0.449763	0.351251

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.061032	0.071436	0.062879	-0.043620	-0.052055	-0.128933	-0.226321	
1	0.106856	-0.097198	-0.035023	0.058737	0.113749	-0.215712	0.017200	
2	-0.005547	-0.121427	0.034395	0.018228	0.020976	0.014566	-0.044512	
3	-0.174648	-0.090001	0.027514	-0.114697	0.072602	-0.114611	0.155499	
4	0.123675	0.031222	-0.061732	0.000232	-0.083922	0.020906	0.152406	
..	…	…	…	…	…	…	…	
995	-0.068811	0.135098	-0.000125	0.019719	-0.111221	-0.174060	-0.011653	
996	-0.060728	0.035252	0.060683	0.130416	0.023056	-0.035884	0.040423	
997	0.086872	0.039709	0.061814	-0.010889	0.004952	-0.016297	-0.106189	
998	-0.000132	0.044601	0.033549	0.035455	0.022138	0.030854	0.105316	
999	-0.005405	-0.012749	0.031845	-0.051491	0.001390	0.007281	0.056732	

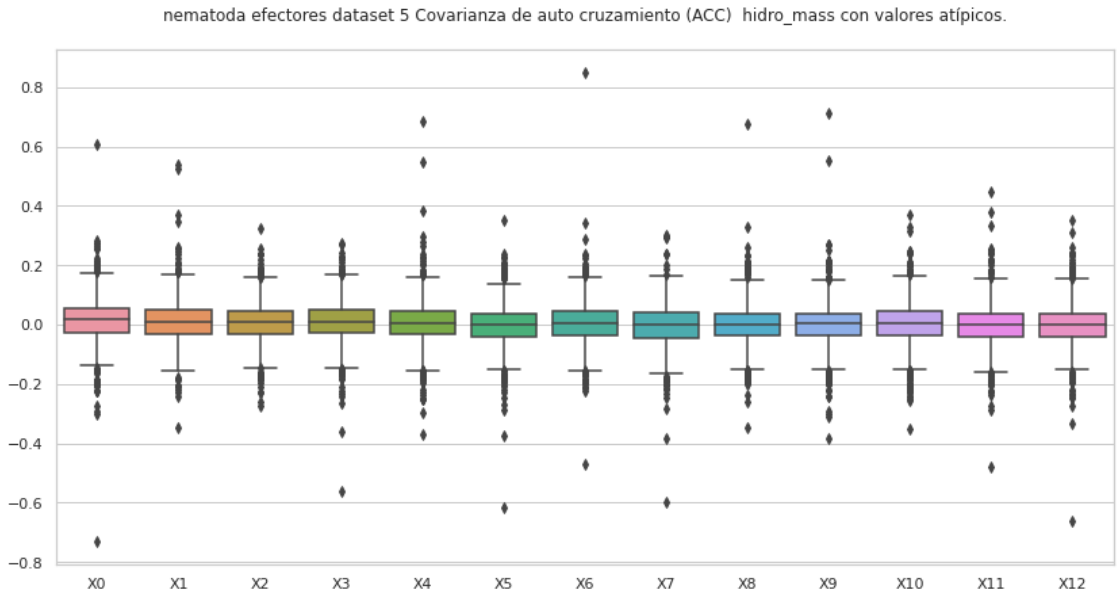
	X7	X8	X9	X10	X11	X12	X13
0	-0.008808	-0.007547	-0.023080	0.177946	0.023195	0.122853	no_efectores
1	-0.052645	0.021088	-0.116181	-0.256115	-0.004686	-0.070947	no_efectores
2	-0.022856	0.025484	-0.031919	-0.050281	-0.085112	-0.007090	no_efectores
3	0.003875	-0.032213	-0.127983	0.036870	0.041078	-0.029101	no_efectores
4	0.047955	-0.000315	-0.002075	0.037371	-0.004019	0.056219	no_efectores
..	…	…	…	…	…	…	…
995	-0.124622	0.129634	-0.083762	0.170637	-0.068227	0.077656	no_efectores
996	0.005172	0.128852	-0.031606	-0.020342	0.080540	-0.027630	no_efectores
997	-0.072978	0.060772	0.057945	-0.025211	0.041341	0.029541	no_efectores
998	0.013676	0.044018	0.027886	0.039204	0.028717	0.073240	no_efectores
999	0.003359	0.032317	0.001227	-0.056464	0.026003	-0.016598	no_efectores

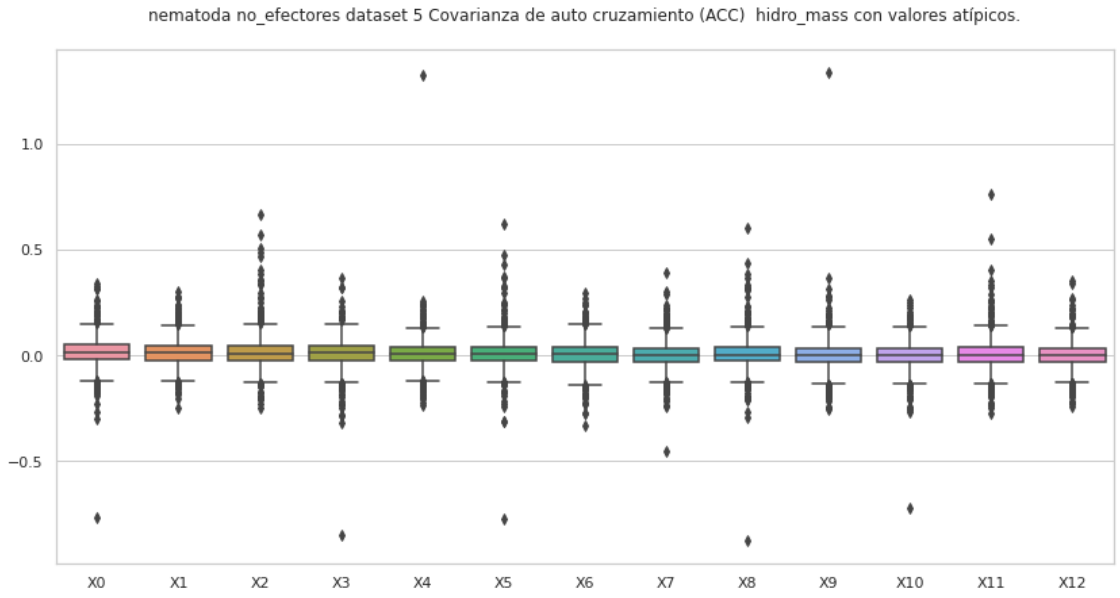
[1000 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.015009	0.013571	0.013755	0.009617	0.007762	
std	0.070985	0.063506	0.078104	0.071528	0.072244	
min	-0.764690	-0.250623	-0.246332	-0.847524	-0.234991	
25%	-0.018567	-0.022058	-0.025272	-0.023866	-0.024718	
50%	0.014904	0.012235	0.010388	0.011090	0.007350	
75%	0.049143	0.045627	0.044430	0.045824	0.038735	

```
max       0.339533     0.301987     0.664530     0.362259     1.321828

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009632     0.004548     0.003472     0.007056     0.000966
std       0.076810     0.065033     0.065359     0.075024     0.077110
min      -0.773181    -0.332875    -0.450146    -0.874222    -0.252828
25%      -0.025090    -0.030471    -0.030230    -0.026825    -0.033087
50%       0.005765     0.005900     0.003395     0.004189     0.001333
75%       0.040892     0.040293     0.033517     0.038305     0.035201
max       0.621165     0.292385     0.394128     0.599925     1.336145

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean     -0.000374     0.007194    -0.000335
std       0.066223     0.075094     0.063284
min      -0.723087    -0.275350    -0.243021
25%      -0.032910    -0.029616    -0.032952
50%       0.000396     0.004112    -0.001533
75%       0.035570     0.037803     0.032226
max       0.265687     0.763901     0.355340
```





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 5,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.198306 -0.096564  0.003400 -0.028014  0.039079 -0.020450  0.109634
2   -0.032040 -0.000362 -0.045080  0.002972 -0.051593 -0.047613  0.012504
6    0.058300  0.090867 -0.025260  0.066682 -0.002370 -0.013914  0.083186
7   -0.188491  0.090376 -0.140627 -0.147295  0.015910  0.040622 -0.071555
8    0.012802  0.033627 -0.001965  0.052073  0.024316 -0.010292  0.029020
..        …         …         …         …         …         …         …
995 -0.058570 -0.044187 -0.004299 -0.009991  0.039203  0.014471  0.159008
996 -0.078046 -0.133529 -0.045082  0.026367 -0.094609 -0.084586  0.141933
997  0.108639  0.093044  0.103261  0.063113  0.104137  0.023857  0.061714
998  0.029922 -0.068364  0.033858  0.108529  0.096435  0.049748 -0.007200
999  0.017326 -0.095379 -0.136142  0.078188  0.013554  0.055849 -0.101645

           X7        X8        X9       X10       X11       X12        X13
0   -0.071913 -0.153224  0.216654  0.026026  0.221540 -0.133203  efectores
2    0.119015  0.003834  0.012017  0.063383  0.057021  0.019342  efectores
6    0.031707  0.040955 -0.032714  0.023982  0.048818  0.018615  efectores
7    0.048117  0.075445  0.006778  0.121414 -0.093226  0.143531  efectores
8   -0.006345 -0.008183 -0.015819 -0.014203  0.010486  0.052660  efectores
..        …         …         …         …         …         …         …
995 -0.006929 -0.044089 -0.041960 -0.008103  0.138331  0.019279  efectores
```

```
996  0.014520 -0.157158  0.043103  0.202248  0.013317 -0.067702  efectores
997  0.070705  0.029685 -0.055110  0.014563  0.119203  0.003765  efectores
998  0.036398  0.070076  0.034989  0.031137  0.041295  0.015967  efectores
999 -0.022477 -0.141220  0.084689  0.011441  0.159226 -0.065709  efectores

[918 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

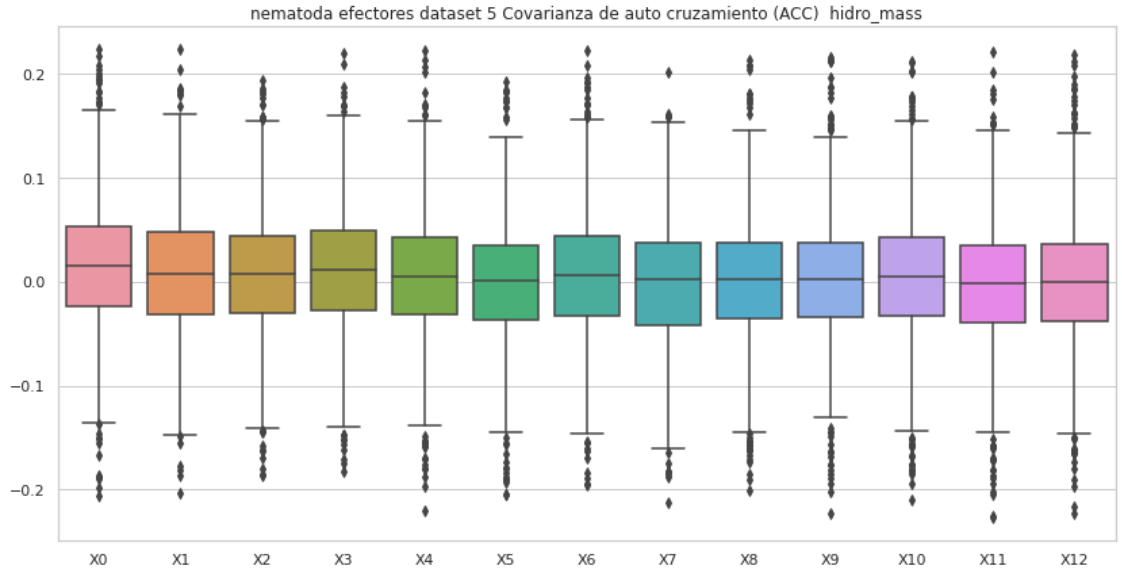
```
                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.015307    0.007436    0.006010    0.009055    0.004728   -0.000409
std      0.063950    0.059636    0.060669    0.061696    0.062293    0.059491
min     -0.206230   -0.203284   -0.186346   -0.182046   -0.220533   -0.205238
25%     -0.023268   -0.030981   -0.030521   -0.027840   -0.031396   -0.037281
50%      0.016013    0.007791    0.008146    0.011014    0.004546    0.001161
75%      0.052701    0.047634    0.043565    0.048836    0.043210    0.035253
max      0.223913    0.223931    0.194317    0.219842    0.222314    0.193302

                X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.006132   -0.001001    0.000570    0.001892    0.003695   -0.002696
std      0.063555    0.061697    0.063057    0.061286    0.065774    0.063506
min     -0.195144   -0.212417   -0.201025   -0.223076   -0.210373   -0.227005
25%     -0.032472   -0.042531   -0.036013   -0.033880   -0.033031   -0.038899
50%      0.006344    0.001928    0.001824    0.003025    0.005642   -0.001824
75%      0.044003    0.037099    0.037412    0.037118    0.042538    0.035274
max      0.222849    0.201909    0.213486    0.216654    0.211735    0.221540

               X12
count  918.000000
mean    -0.001048
std      0.063877
min     -0.222792
25%     -0.037850
50%      0.000442
75%      0.035811
max      0.218312
```

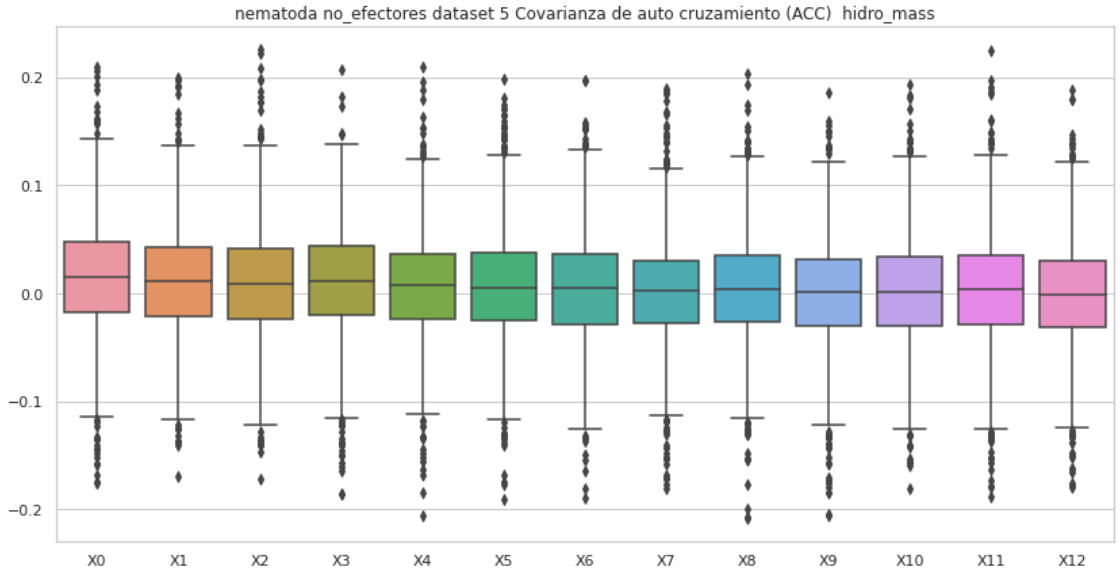
no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
2   -0.005547 -0.121427  0.034395  0.018228  0.020976  0.014566 -0.044512
3   -0.174648 -0.090001  0.027514 -0.114697  0.072602 -0.114611  0.155499
4    0.123675  0.031222 -0.061732  0.000232 -0.083922  0.020906  0.152406
5    0.063299 -0.112670 -0.010102 -0.003871 -0.003073 -0.027940 -0.098261
6    0.071370 -0.016682  0.055829  0.006114 -0.069399 -0.093705 -0.006958
..        …         …         …         …         …         …
995 -0.068811  0.135098 -0.000125  0.019719 -0.111221 -0.174060 -0.011653
996 -0.060728  0.035252  0.060683  0.130416  0.023056 -0.035884  0.040423
997  0.086872  0.039709  0.061814 -0.010889  0.004952 -0.016297 -0.106189
998 -0.000132  0.044601  0.033549  0.035455  0.022138  0.030854  0.105316
999 -0.005405 -0.012749  0.031845 -0.051491  0.001390  0.007281  0.056732

           X7        X8        X9       X10       X11       X12           X13
2   -0.022856  0.025484 -0.031919 -0.050281 -0.085112 -0.007090  no_efectores
3    0.003875 -0.032213 -0.127983  0.036870  0.041078 -0.029101  no_efectores
4    0.047955 -0.000315 -0.002075  0.037371 -0.004019  0.056219  no_efectores
5   -0.019385  0.042201  0.081564 -0.021323 -0.091489 -0.072680  no_efectores
6   -0.025126  0.002378  0.122339 -0.022878  0.015668  0.040485  no_efectores
..        …         …         …         …         …         …           …
995 -0.124622  0.129634 -0.083762  0.170637 -0.068227  0.077656  no_efectores
996  0.005172  0.128852 -0.031606 -0.020342  0.080540 -0.027630  no_efectores
997 -0.072978  0.060772  0.057945 -0.025211  0.041341  0.029541  no_efectores
998  0.013676  0.044018  0.027886  0.039204  0.028717  0.073240  no_efectores
999  0.003359  0.032317  0.001227 -0.056464  0.026003 -0.016598  no_efectores

[918 rows x 14 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.014757    0.011325    0.008550    0.011182    0.005872    0.005798
std      0.057564    0.053722    0.055070    0.053648    0.051312    0.054381
min     -0.175914   -0.169354   -0.172028   -0.186163   -0.205463   -0.190172
25%     -0.017913   -0.021282   -0.024163   -0.019797   -0.023567   -0.025042
50%      0.014807    0.011495    0.009386    0.011090    0.007177    0.004705
75%      0.047225    0.042804    0.041341    0.043794    0.035987    0.037170
max      0.209380    0.200221    0.225905    0.207706    0.209778    0.199002

                X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.004186    0.001867    0.004665   -0.000823    0.001910    0.002960
std      0.053948    0.052922    0.051956    0.054360    0.052140    0.054513
```

min	-0.189507	-0.181022	-0.208166	-0.205915	-0.180734	-0.188148
25%	-0.028834	-0.027190	-0.026014	-0.029911	-0.030017	-0.029006
50%	0.005766	0.003210	0.003586	0.001333	0.001253	0.003450
75%	0.036347	0.030270	0.035222	0.031662	0.033944	0.035651
max	0.196615	0.189135	0.203086	0.185817	0.193163	0.225067

	X12
count	918.000000
mean	-0.001097
std	0.052214
min	-0.179347
25%	-0.031497
50%	-0.001555
75%	0.030242
max	0.187762





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))
```

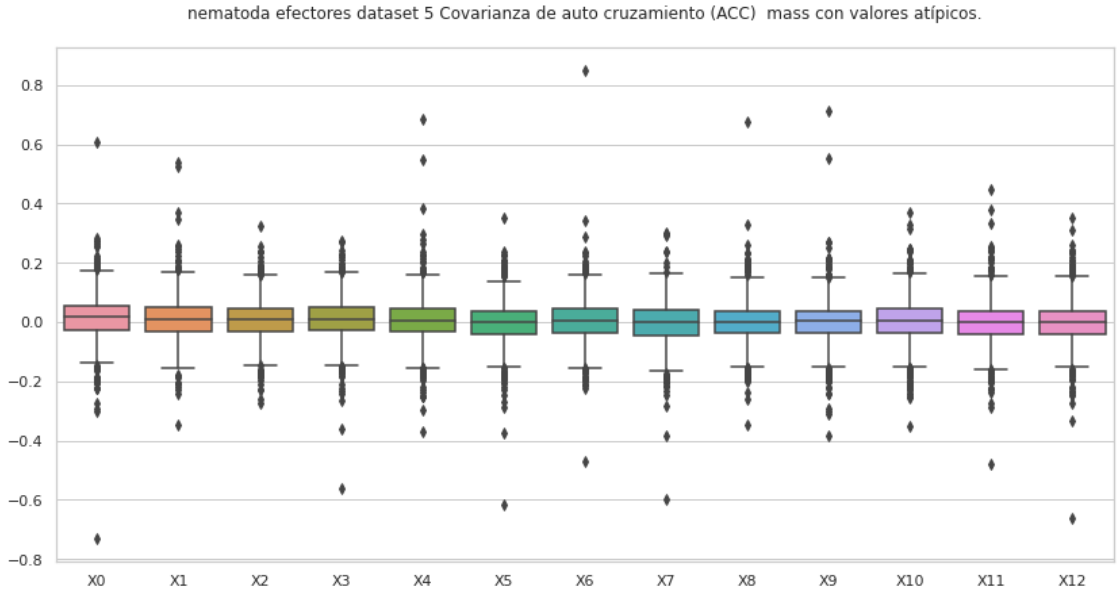
```
efectores
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.198306 -0.096564  0.003400 -0.028014  0.039079 -0.020450  0.109634
1    0.215057  0.099904  0.180423  0.135070  0.171050  0.230267  0.175587
2   -0.032040 -0.000362 -0.045080  0.002972 -0.051593 -0.047613  0.012504
3    0.023837 -0.034744  0.120064  0.089214  0.171561  0.033669  0.003346
4   -0.066733  0.017327  0.141441 -0.122540 -0.001408 -0.286970  0.004042
..        …         …         …         …         …         …         …
995 -0.058570 -0.044187 -0.004299 -0.009991  0.039203  0.014471  0.159008
996 -0.078046 -0.133529 -0.045082  0.026367 -0.094609 -0.084586  0.141933
997  0.108639  0.093044  0.103261  0.063113  0.104137  0.023857  0.061714
998  0.029922 -0.068364  0.033858  0.108529  0.096435  0.049748 -0.007200
999  0.017326 -0.095379 -0.136142  0.078188  0.013554  0.055849 -0.101645

           X7        X8        X9       X10       X11       X12        X13
0   -0.071913 -0.153224  0.216654  0.026026  0.221540 -0.133203  efectores
1    0.032970  0.139461  0.044398  0.175364  0.255569  0.098884  efectores
2    0.119015  0.003834  0.012017  0.063383  0.057021  0.019342  efectores
3    0.236380  0.080222  0.043116  0.054423  0.028470  0.144641  efectores
4   -0.092650 -0.127373  0.134217  0.245591 -0.054127 -0.117980  efectores
..        …         …         …         …         …         …         …
995 -0.006929 -0.044089 -0.041960 -0.008103  0.138331  0.019279  efectores
996  0.014520 -0.157158  0.043103  0.202248  0.013317 -0.067702  efectores
997  0.070705  0.029685 -0.055110  0.014563  0.119203  0.003765  efectores
998  0.036398  0.070076  0.034989  0.031137  0.041295  0.015967  efectores
999 -0.022477 -0.141220  0.084689  0.011441  0.159226 -0.065709  efectores

[1000 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 5, con
valores atípicos.
Estadísticas.

                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.014910     0.008770     0.006702     0.008283     0.005633
```

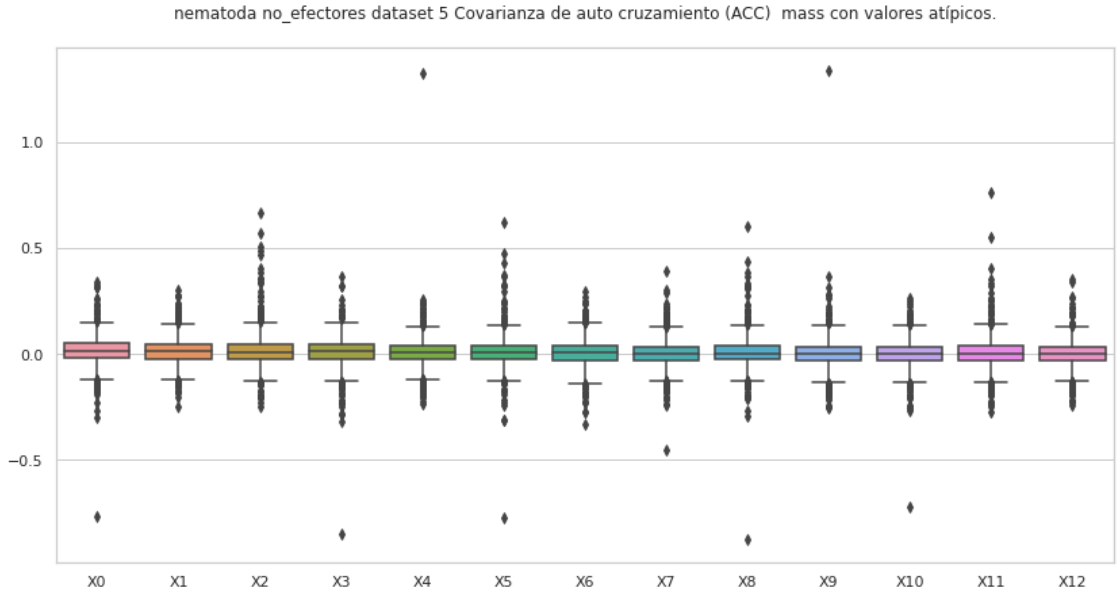
```
std       0.077829     0.072842     0.069108     0.071784     0.076672
min      -0.730653    -0.344726    -0.271595    -0.561751    -0.371485
25%      -0.024945    -0.033161    -0.031352    -0.029133    -0.033886
50%       0.016749     0.007791     0.008179     0.010388     0.004396
75%       0.054595     0.049217     0.044923     0.050431     0.045511
max       0.609494     0.537329     0.326743     0.274181     0.686111

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.000506     0.005856    -0.001235     0.000341     0.001216
std       0.072155     0.075740     0.072899     0.073808     0.076085
min      -0.618578    -0.469635    -0.598750    -0.346286    -0.381541
25%      -0.038934    -0.035579    -0.043917    -0.037862    -0.037026
50%       0.001417     0.006072     0.002389     0.001853     0.002975
75%       0.037168     0.044491     0.041393     0.038597     0.038135
max       0.352317     0.849472     0.303462     0.676845     0.711217

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.004567    -0.000315    -0.000987
std       0.076611     0.075857     0.075130
min      -0.352250    -0.478147    -0.664122
25%      -0.034485    -0.041081    -0.039002
50%       0.005967    -0.000503     0.001008
75%       0.045549     0.037871     0.038164
max       0.369355     0.449763     0.351251



no_efectores
Covarianza de auto cruzamiento (ACC)  mass no_efectores nematoda dataset 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.061032  0.071436  0.062879 -0.043620 -0.052055 -0.128933 -0.226321
1    0.106856 -0.097198 -0.035023  0.058737  0.113749 -0.215712  0.017200
2   -0.005547 -0.121427  0.034395  0.018228  0.020976  0.014566 -0.044512
3   -0.174648 -0.090001  0.027514 -0.114697  0.072602 -0.114611  0.155499
4    0.123675  0.031222 -0.061732  0.000232 -0.083922  0.020906  0.152406
..        …         …         …         …         …         …         …
995 -0.068811  0.135098 -0.000125  0.019719 -0.111221 -0.174060 -0.011653
996 -0.060728  0.035252  0.060683  0.130416  0.023056 -0.035884  0.040423
997  0.086872  0.039709  0.061814 -0.010889  0.004952 -0.016297 -0.106189
998 -0.000132  0.044601  0.033549  0.035455  0.022138  0.030854  0.105316
999 -0.005405 -0.012749  0.031845 -0.051491  0.001390  0.007281  0.056732

           X7        X8        X9       X10       X11       X12           X13
```

```
0   -0.008808 -0.007547 -0.023080  0.177946  0.023195  0.122853  no_efectores
1   -0.052645  0.021088 -0.116181 -0.256115 -0.004686 -0.070947  no_efectores
2   -0.022856  0.025484 -0.031919 -0.050281 -0.085112 -0.007090  no_efectores
3    0.003875 -0.032213 -0.127983  0.036870  0.041078 -0.029101  no_efectores
4    0.047955 -0.000315 -0.002075  0.037371 -0.004019  0.056219  no_efectores
..        …       …       …       …       …       …           …
995 -0.124622  0.129634 -0.083762  0.170637 -0.068227  0.077656  no_efectores
996  0.005172  0.128852 -0.031606 -0.020342  0.080540 -0.027630  no_efectores
997 -0.072978  0.060772  0.057945 -0.025211  0.041341  0.029541  no_efectores
998  0.013676  0.044018  0.027886  0.039204  0.028717  0.073240  no_efectores
999  0.003359  0.032317  0.001227 -0.056464  0.026003 -0.016598  no_efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.015009     0.013571     0.013755     0.009617     0.007762
std       0.070985     0.063506     0.078104     0.071528     0.072244
min      -0.764690    -0.250623    -0.246332    -0.847524    -0.234991
25%      -0.018567    -0.022058    -0.025272    -0.023866    -0.024718
50%       0.014904     0.012235     0.010388     0.011090     0.007350
75%       0.049143     0.045627     0.044430     0.045824     0.038735
max       0.339533     0.301987     0.664530     0.362259     1.321828

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.009632     0.004548     0.003472     0.007056     0.000966
std       0.076810     0.065033     0.065359     0.075024     0.077110
min      -0.773181    -0.332875    -0.450146    -0.874222    -0.252828
25%      -0.025090    -0.030471    -0.030230    -0.026825    -0.033087
50%       0.005765     0.005900     0.003395     0.004189     0.001333
75%       0.040892     0.040293     0.033517     0.038305     0.035201
max       0.621165     0.292385     0.394128     0.599925     1.336145

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean     -0.000374     0.007194    -0.000335
std       0.066223     0.075094     0.063284
min      -0.723087    -0.275350    -0.243021
25%      -0.032910    -0.029616    -0.032952
50%       0.000396     0.004112    -0.001533
75%       0.035570     0.037803     0.032226
max       0.265687     0.763901     0.355340
```


nematoda efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) mass con valores atípicos.
1.0
0.5
0.0
−0.5
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 5, sin
valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.198306	-0.096564	0.003400	-0.028014	0.039079	-0.020450	0.109634	
2	-0.032040	-0.000362	-0.045080	0.002972	-0.051593	-0.047613	0.012504	
6	0.058300	0.090867	-0.025260	0.066682	-0.002370	-0.013914	0.083186	
7	-0.188491	0.090376	-0.140627	-0.147295	0.015910	0.040622	-0.071555	
8	0.012802	0.033627	-0.001965	0.052073	0.024316	-0.010292	0.029020	
..	…	…	…	…	…	…	…	
995	-0.058570	-0.044187	-0.004299	-0.009991	0.039203	0.014471	0.159008	
996	-0.078046	-0.133529	-0.045082	0.026367	-0.094609	-0.084586	0.141933	
997	0.108639	0.093044	0.103261	0.063113	0.104137	0.023857	0.061714	
998	0.029922	-0.068364	0.033858	0.108529	0.096435	0.049748	-0.007200	
999	0.017326	-0.095379	-0.136142	0.078188	0.013554	0.055849	-0.101645	

	X7	X8	X9	X10	X11	X12	X13
0	-0.071913	-0.153224	0.216654	0.026026	0.221540	-0.133203	efectores
2	0.119015	0.003834	0.012017	0.063383	0.057021	0.019342	efectores
6	0.031707	0.040955	-0.032714	0.023982	0.048818	0.018615	efectores
7	0.048117	0.075445	0.006778	0.121414	-0.093226	0.143531	efectores
8	-0.006345	-0.008183	-0.015819	-0.014203	0.010486	0.052660	efectores
..	…	…	…	…	…	…	…
995	-0.006929	-0.044089	-0.041960	-0.008103	0.138331	0.019279	efectores
996	0.014520	-0.157158	0.043103	0.202248	0.013317	-0.067702	efectores
997	0.070705	0.029685	-0.055110	0.014563	0.119203	0.003765	efectores
998	0.036398	0.070076	0.034989	0.031137	0.041295	0.015967	efectores
999	-0.022477	-0.141220	0.084689	0.011441	0.159226	-0.065709	efectores

```
[918 rows x 14 columns]
```

```
Covarianza de auto cruzamiento (ACC)  mass efectores nematoda dataset 5, sin
valores atípicos.
Estadísticas.
```

	X0	X1	X2	X3	X4	X5	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.015307	0.007436	0.006010	0.009055	0.004728	-0.000409	
std	0.063950	0.059636	0.060669	0.061696	0.062293	0.059491	
min	-0.206230	-0.203284	-0.186346	-0.182046	-0.220533	-0.205238	
25%	-0.023268	-0.030981	-0.030521	-0.027840	-0.031396	-0.037281	
50%	0.016013	0.007791	0.008146	0.011014	0.004546	0.001161	
75%	0.052701	0.047634	0.043565	0.048836	0.043210	0.035253	

```
max      0.223913    0.223931    0.194317    0.219842    0.222314    0.193302

               X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.006132   -0.001001    0.000570    0.001892    0.003695   -0.002696
std      0.063555    0.061697    0.063057    0.061286    0.065774    0.063506
min     -0.195144   -0.212417   -0.201025   -0.223076   -0.210373   -0.227005
25%     -0.032472   -0.042531   -0.036013   -0.033880   -0.033031   -0.038899
50%      0.006344    0.001928    0.001824    0.003025    0.005642   -0.001824
75%      0.044003    0.037099    0.037412    0.037118    0.042538    0.035274
max      0.222849    0.201909    0.213486    0.216654    0.211735    0.221540

              X12
count  918.000000
mean    -0.001048
std      0.063877
min     -0.222792
25%     -0.037850
50%      0.000442
75%      0.035811
max      0.218312
```

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
2   -0.005547 -0.121427  0.034395  0.018228  0.020976  0.014566 -0.044512
3   -0.174648 -0.090001  0.027514 -0.114697  0.072602 -0.114611  0.155499
4    0.123675  0.031222 -0.061732  0.000232 -0.083922  0.020906  0.152406
5    0.063299 -0.112670 -0.010102 -0.003871 -0.003073 -0.027940 -0.098261
6    0.071370 -0.016682  0.055829  0.006114 -0.069399 -0.093705 -0.006958
..        …         …         …         …         …         …
995 -0.068811  0.135098 -0.000125  0.019719 -0.111221 -0.174060 -0.011653
996 -0.060728  0.035252  0.060683  0.130416  0.023056 -0.035884  0.040423
997  0.086872  0.039709  0.061814 -0.010889  0.004952 -0.016297 -0.106189
998 -0.000132  0.044601  0.033549  0.035455  0.022138  0.030854  0.105316
999 -0.005405 -0.012749  0.031845 -0.051491  0.001390  0.007281  0.056732

           X7        X8        X9       X10       X11       X12           X13
2   -0.022856  0.025484 -0.031919 -0.050281 -0.085112 -0.007090  no_efectores
3    0.003875 -0.032213 -0.127983  0.036870  0.041078 -0.029101  no_efectores
4    0.047955 -0.000315 -0.002075  0.037371 -0.004019  0.056219  no_efectores
5   -0.019385  0.042201  0.081564 -0.021323 -0.091489 -0.072680  no_efectores
6   -0.025126  0.002378  0.122339 -0.022878  0.015668  0.040485  no_efectores
..        …         …         …         …         …         …           …
```

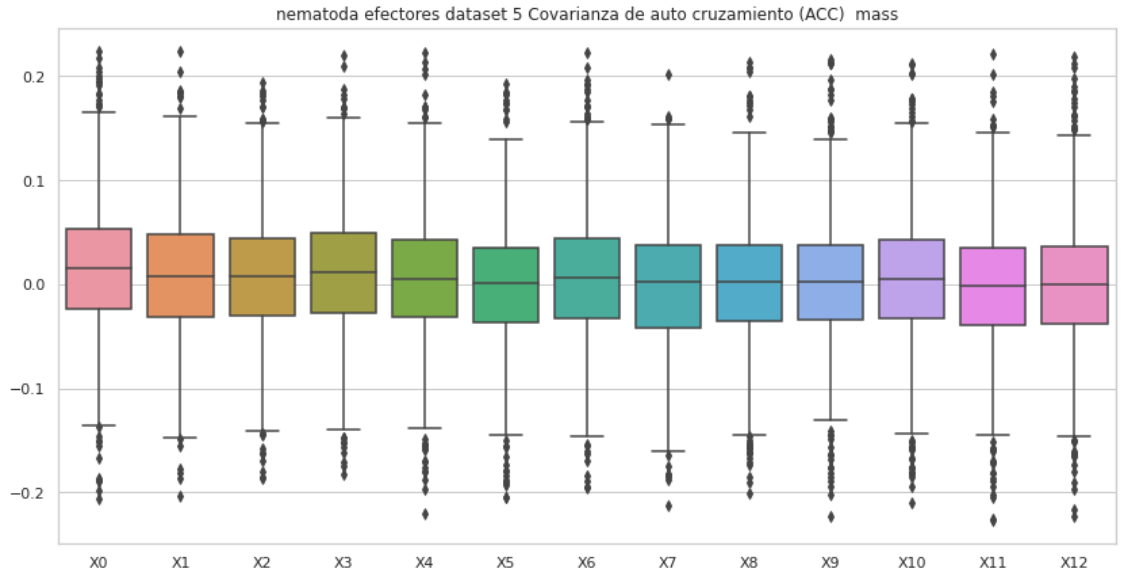
```
995 -0.124622  0.129634 -0.083762  0.170637 -0.068227  0.077656  no_efectores
996  0.005172  0.128852 -0.031606 -0.020342  0.080540 -0.027630  no_efectores
997 -0.072978  0.060772  0.057945 -0.025211  0.041341  0.029541  no_efectores
998  0.013676  0.044018  0.027886  0.039204  0.028717  0.073240  no_efectores
999  0.003359  0.032317  0.001227 -0.056464  0.026003 -0.016598  no_efectores

[918 rows x 14 columns]
```

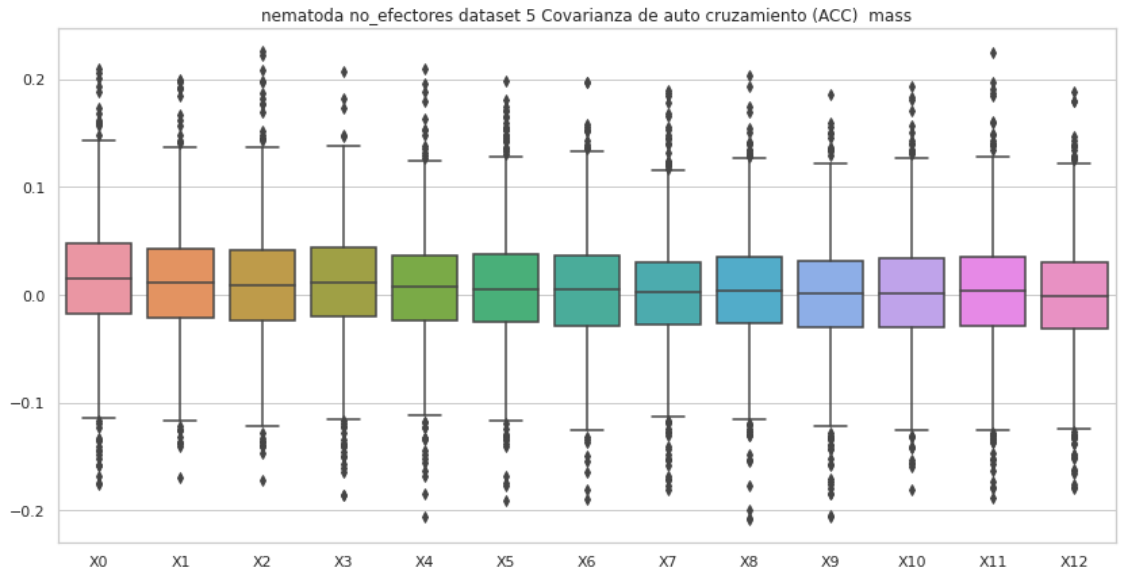
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.014757	0.011325	0.008550	0.011182	0.005872	0.005798	
std	0.057564	0.053722	0.055070	0.053648	0.051312	0.054381	
min	-0.175914	-0.169354	-0.172028	-0.186163	-0.205463	-0.190172	
25%	-0.017913	-0.021282	-0.024163	-0.019797	-0.023567	-0.025042	
50%	0.014807	0.011495	0.009386	0.011090	0.007177	0.004705	
75%	0.047225	0.042804	0.041341	0.043794	0.035987	0.037170	
max	0.209380	0.200221	0.225905	0.207706	0.209778	0.199002	

	X6	X7	X8	X9	X10	X11	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.004186	0.001867	0.004665	-0.000823	0.001910	0.002960	
std	0.053948	0.052922	0.051956	0.054360	0.052140	0.054513	
min	-0.189507	-0.181022	-0.208166	-0.205915	-0.180734	-0.188148	
25%	-0.028834	-0.027190	-0.026014	-0.029911	-0.030017	-0.029006	
50%	0.005766	0.003210	0.003586	0.001333	0.001253	0.003450	
75%	0.036347	0.030270	0.035222	0.031662	0.033944	0.035651	
max	0.196615	0.189135	0.203086	0.185817	0.193163	0.225067	

	X12
count	918.000000
mean	-0.001097
std	0.052214
min	-0.179347
25%	-0.031497
50%	-0.001555
75%	0.030242
max	0.187762





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 5, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.128452 -0.000162  0.102696  0.167912 -0.084527 -0.240126  0.024804
1    0.135194  0.100501  0.100002  0.026687  0.122726  0.107765  0.203718
2   -0.033040 -0.072263  0.047394 -0.011275  0.042384 -0.111268  0.001604
3    0.069386  0.060163  0.094871  0.057057  0.042433 -0.043737  0.002637
4    0.054367  0.190111 -0.023659  0.063262  0.108061  0.009113 -0.069412
..        …         …         …         …         …         …         …
995  0.179546  0.044590  0.065685 -0.035581 -0.189708 -0.106743 -0.017193
996  0.120294  0.093305 -0.009688 -0.180956 -0.088899 -0.038287 -0.048829
997  0.071745  0.065549  0.117622  0.077274  0.106744 -0.009052  0.008855
998 -0.064379  0.003856 -0.099411  0.017018  0.040957 -0.073808  0.031955
999  0.005325  0.036251 -0.043201 -0.031132 -0.087924  0.027403 -0.048407

           X7        X8        X9       X10       X11       X12        X13
0   -0.099054 -0.252367 -0.158609  0.113583 -0.109426 -0.034376  efectores
1   -0.027204  0.123327  0.121245  0.087161  0.137966  0.063475  efectores
2    0.067682 -0.026528  0.007899  0.131103 -0.024099 -0.063176  efectores
```

```
3    0.070348  0.032627 -0.032609  0.000171  0.043627 -0.122681  efectores
4    0.038323 -0.234348  0.026135 -0.121069  0.000743  0.121557  efectores
..        …       …        …        …         …         …        …
995  0.005925 -0.049907 -0.060016  0.027915 -0.101107 -0.299604  efectores
996 -0.180543 -0.095023 -0.148668  0.100779  0.107914  0.092290  efectores
997  0.021389  0.121020  0.091602  0.074864  0.046938  0.035093  efectores
998  0.019434  0.079115 -0.034108 -0.056138  0.021228 -0.091615  efectores
999  0.003140 -0.068103 -0.164210 -0.091879 -0.106407 -0.152882  efectores

[1000 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

```
                X0           X1           X2           X3           X4  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean      0.013716    -0.022624     0.019216     0.025902    -0.005849
std       0.086284     0.091905     0.082619     0.087664     0.088191
min      -0.304425    -0.334397    -0.277291    -0.301561    -0.322481
25%      -0.039367    -0.077922    -0.033633    -0.022884    -0.059873
50%       0.011454    -0.024229     0.018338     0.026066    -0.006464
75%       0.061546     0.031581     0.066044     0.072968     0.041786
max       0.443122     0.395286     0.395733     0.368211     0.539266

                X5           X6           X7           X8           X9  \
count  1000.000000  1000.000000  1000.000000  1000.000000  1000.000000
mean     -0.008317     0.019914     0.005441     0.000351     0.007478
std       0.084846     0.088969     0.084963     0.086239     0.082610
min      -0.358802    -0.409835    -0.513972    -0.342471    -0.368382
25%      -0.054941    -0.028722    -0.038493    -0.047278    -0.038393
50%      -0.008225     0.019926     0.007274    -0.000204     0.005924
75%       0.037889     0.065535     0.050582     0.046406     0.052698
max       0.452901     0.392374     0.358559     0.524962     0.592679

               X10          X11          X12
count  1000.000000  1000.000000  1000.000000
mean      0.011951     0.003334     0.001307
std       0.084677     0.084023     0.088686
min      -0.408205    -0.331700    -0.427011
25%      -0.028124    -0.042536    -0.047183
50%       0.011607     0.002142    -0.000391
75%       0.056869     0.050347     0.052024
max       0.455874     0.403142     0.363955
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.161990	-0.007426	0.151202	0.092768	0.045317	0.074516	0.133144	
1	0.085818	-0.146357	0.017810	-0.023908	0.082598	0.059489	-0.170718	
2	0.055878	0.057311	0.040485	-0.008638	-0.012641	-0.005477	0.001211	
3	-0.103230	-0.054478	0.069063	-0.166415	-0.045693	-0.149392	0.092383	
4	-0.069450	-0.019427	-0.089064	-0.066985	-0.004365	-0.101446	-0.030350	
..	…	…	…	…	…	…	…	
995	0.052054	-0.043221	-0.024911	-0.039809	0.092963	-0.020511	-0.114742	
996	0.056339	-0.027295	0.010792	0.058589	0.049617	0.026123	0.011344	
997	0.071599	0.187168	0.183859	0.063273	0.045617	-0.040309	0.077601	
998	-0.027486	-0.128328	-0.031883	0.034670	-0.031900	-0.042074	0.089526	
999	0.046114	0.051110	0.042044	-0.007135	0.086706	0.067162	0.053208	

	X7	X8	X9	X10	X11	X12	X13
0	0.196866	0.200918	0.164710	-0.098171	0.063882	0.119362	no_efectores
1	-0.244734	0.061753	0.103424	-0.063126	-0.236811	-0.076397	no_efectores
2	0.012846	-0.015772	-0.004101	0.007424	-0.049223	-0.044114	no_efectores
3	-0.123357	0.034708	0.130875	0.045462	0.212005	-0.201032	no_efectores
4	-0.105785	0.106633	0.035750	-0.071841	0.044751	0.116815	no_efectores
..	…	…	…	…	…	…	…
995	-0.076427	0.009143	0.273273	0.038402	-0.053683	0.046609	no_efectores
996	0.011312	0.041981	0.020503	0.052589	0.087481	-0.044838	no_efectores
997	-0.053126	0.087933	0.044219	0.077988	0.096999	0.117131	no_efectores
998	-0.036950	-0.073947	0.080444	0.065114	0.017460	0.005581	no_efectores
999	0.011418	0.038474	0.086059	0.047189	0.057657	0.035279	no_efectores

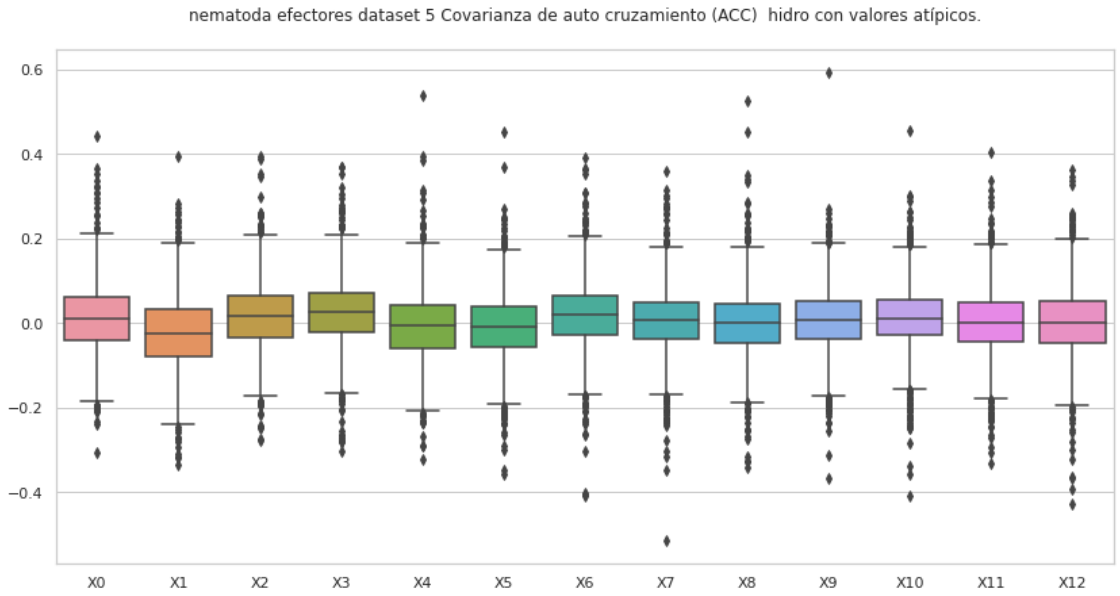
[1000 rows x 14 columns]

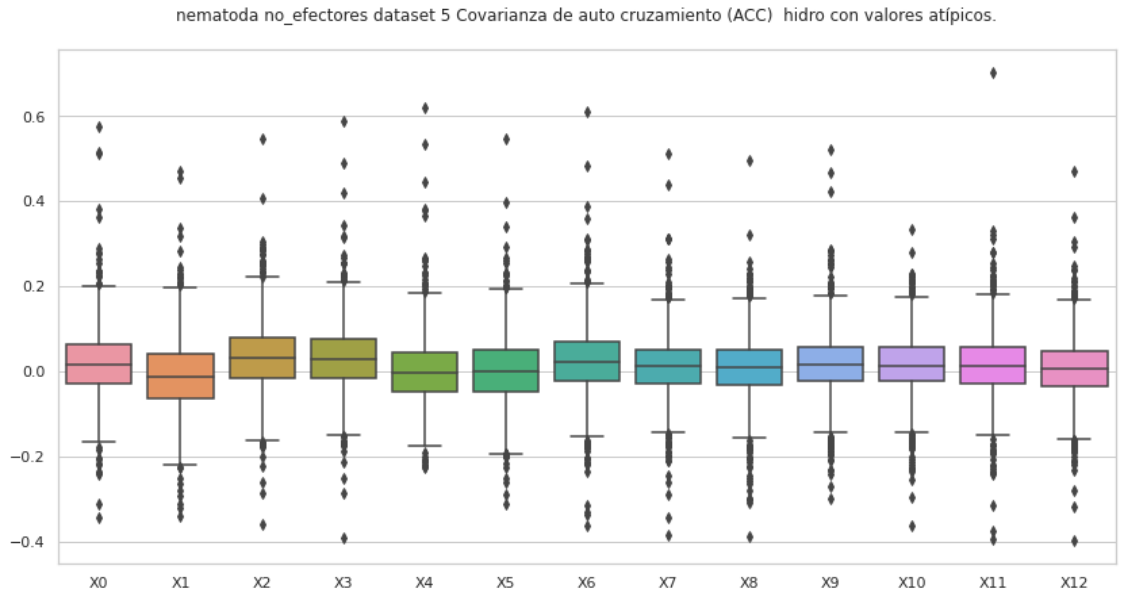
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 5, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.020954	-0.008515	0.035293	0.032728	0.001245	
std	0.084779	0.088851	0.082781	0.081610	0.085394	
min	-0.341432	-0.341046	-0.359098	-0.389394	-0.226311	
25%	-0.027735	-0.062452	-0.016713	-0.014298	-0.048389	
50%	0.016640	-0.012968	0.031560	0.029590	-0.004422	
75%	0.063649	0.041739	0.079264	0.076484	0.044652	
max	0.574231	0.469443	0.546224	0.588738	0.619122	

	X5	X6	X7	X8	X9	\
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	
mean	0.001171	0.025551	0.012442	0.008959	0.017665	
std	0.083077	0.086483	0.077996	0.080179	0.078101	
min	-0.312421	-0.360768	-0.384432	-0.386232	-0.299825	
25%	-0.047352	-0.021771	-0.027374	-0.031544	-0.023392	
50%	0.000374	0.021759	0.011775	0.008521	0.015646	
75%	0.049860	0.070413	0.051616	0.051552	0.057820	
max	0.545517	0.610655	0.510624	0.496627	0.520394	

	X10	X11	X12
count	1000.000000	1000.000000	1000.000000
mean	0.015495	0.013672	0.006681
std	0.075341	0.081264	0.076196
min	-0.362667	-0.393206	-0.396673
25%	-0.023092	-0.028004	-0.035440
50%	0.014333	0.012063	0.005891
75%	0.057868	0.056184	0.047520
max	0.334525	0.701620	0.470655





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 5, sin valores
atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0   -0.128452 -0.000162  0.102696  0.167912 -0.084527 -0.240126  0.024804
1    0.135194  0.100501  0.100002  0.026687  0.122726  0.107765  0.203718
2   -0.033040 -0.072263  0.047394 -0.011275  0.042384 -0.111268  0.001604
3    0.069386  0.060163  0.094871  0.057057  0.042433 -0.043737  0.002637
4    0.054367  0.190111 -0.023659  0.063262  0.108061  0.009113 -0.069412
..        …         …         …         …         …         …         …
994  0.036998 -0.050945  0.066860 -0.003897  0.066903  0.015017 -0.068114
996  0.120294  0.093305 -0.009688 -0.180956 -0.088899 -0.038287 -0.048829
997  0.071745  0.065549  0.117622  0.077274  0.106744 -0.009052  0.008855
998 -0.064379  0.003856 -0.099411  0.017018  0.040957 -0.073808  0.031955
999  0.005325  0.036251 -0.043201 -0.031132 -0.087924  0.027403 -0.048407

           X7        X8        X9       X10       X11       X12        X13
0   -0.099054 -0.252367 -0.158609  0.113583 -0.109426 -0.034376  efectores
1   -0.027204  0.123327  0.121245  0.087161  0.137966  0.063475  efectores
2    0.067682 -0.026528  0.007899  0.131103 -0.024099 -0.063176  efectores
3    0.070348  0.032627 -0.032609  0.000171  0.043627 -0.122681  efectores
4    0.038323 -0.234348  0.026135 -0.121069  0.000743  0.121557  efectores
..        …         …         …         …         …         …         …
994 -0.016209  0.122240  0.027416  0.033067 -0.001349  0.091234  efectores
```

```
996 -0.180543 -0.095023 -0.148668  0.100779  0.107914  0.092290  efectores
997  0.021389  0.121020  0.091602  0.074864  0.046938  0.035093  efectores
998  0.019434  0.079115 -0.034108 -0.056138  0.021228 -0.091615  efectores
999  0.003140 -0.068103 -0.164210 -0.091879 -0.106407 -0.152882  efectores

[918 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.011736	-0.022384	0.017324	0.026059	-0.005938	-0.007579	
std	0.076085	0.083198	0.074045	0.076144	0.076261	0.074701	
min	-0.208958	-0.279976	-0.212260	-0.231822	-0.235857	-0.256734	
25%	-0.038675	-0.074327	-0.031067	-0.021588	-0.055181	-0.052522	
50%	0.011126	-0.024081	0.016688	0.026443	-0.006732	-0.007959	
75%	0.059324	0.028998	0.062355	0.071528	0.039581	0.036528	
max	0.272329	0.245496	0.257785	0.265589	0.252273	0.224611	

	X6	X7	X8	X9	X10	X11	\
count	918.000000	918.000000	918.000000	918.000000	918.000000	918.000000	
mean	0.020166	0.005980	0.000610	0.006417	0.012166	0.001997	
std	0.077021	0.073025	0.072676	0.073507	0.073768	0.074056	
min	-0.233427	-0.235608	-0.254586	-0.236821	-0.241500	-0.245993	
25%	-0.027364	-0.035676	-0.042640	-0.037244	-0.026583	-0.041743	
50%	0.019926	0.008338	0.000097	0.005124	0.011241	0.001585	
75%	0.063860	0.048357	0.042522	0.048387	0.055451	0.046329	
max	0.285682	0.256325	0.253627	0.246196	0.262823	0.246045	

	X12
count	918.000000
mean	0.001807
std	0.075950
min	-0.250969
25%	-0.045254
50%	-0.000849
75%	0.047386
max	0.255617

no_efectores
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 5, sin valores atípicos.
Valores del documento csv.

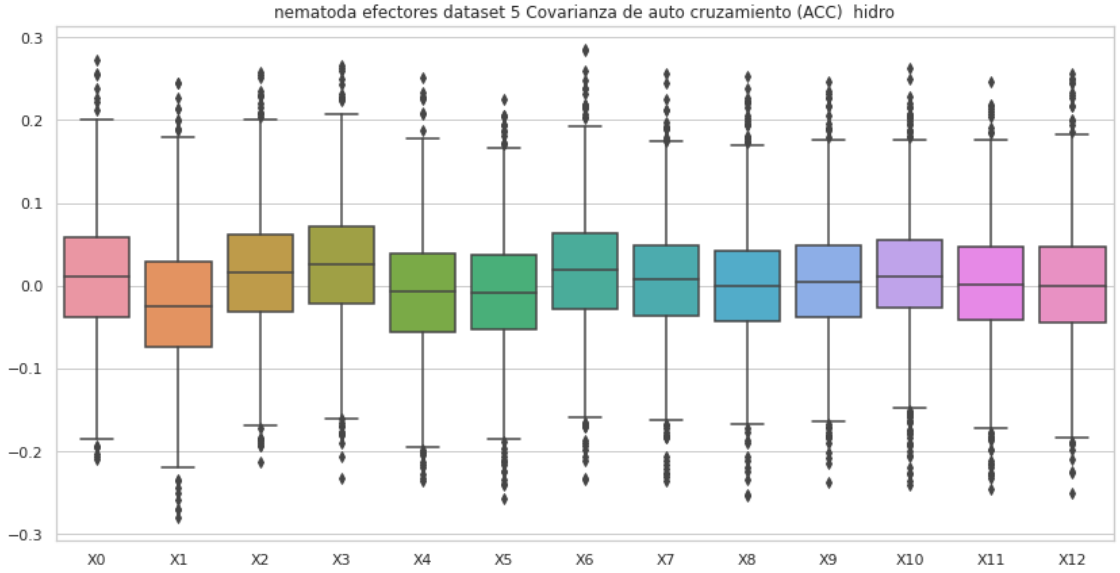
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.161990 -0.007426  0.151202  0.092768  0.045317  0.074516  0.133144
2    0.055878  0.057311  0.040485 -0.008638 -0.012641 -0.005477  0.001211
3   -0.103230 -0.054478  0.069063 -0.166415 -0.045693 -0.149392  0.092383
4   -0.069450 -0.019427 -0.089064 -0.066985 -0.004365 -0.101446 -0.030350
5   -0.157782 -0.075530  0.055190 -0.211713 -0.047742  0.045582 -0.002715
..        …         …         …         …         …         …         …
994 -0.025595 -0.141954 -0.022330  0.174807 -0.065224 -0.038963  0.120157
996  0.056339 -0.027295  0.010792  0.058589  0.049617  0.026123  0.011344
997  0.071599  0.187168  0.183859  0.063273  0.045617 -0.040309  0.077601
998 -0.027486 -0.128328 -0.031883  0.034670 -0.031900 -0.042074  0.089526
999  0.046114  0.051110  0.042044 -0.007135  0.086706  0.067162  0.053208

           X7        X8        X9       X10       X11       X12           X13
0    0.196866  0.200918  0.164710 -0.098171  0.063882  0.119362  no_efectores
2    0.012846 -0.015772 -0.004101  0.007424 -0.049223 -0.044114  no_efectores
3   -0.123357  0.034708  0.130875  0.045462  0.212005 -0.201032  no_efectores
4   -0.105785  0.106633  0.035750 -0.071841  0.044751  0.116815  no_efectores
5    0.184904 -0.018907 -0.033557  0.104354 -0.103988 -0.072245  no_efectores
..        …         …         …         …         …         …           …
994  0.000015 -0.122084 -0.209639  0.068282  0.179329  0.012810  no_efectores
996  0.011312  0.041981  0.020503  0.052589  0.087481 -0.044838  no_efectores
997 -0.053126  0.087933  0.044219  0.077988  0.096999  0.117131  no_efectores
998 -0.036950 -0.073947  0.080444  0.065114  0.017460  0.005581  no_efectores
999  0.011418  0.038474  0.086059  0.047189  0.057657  0.035279  no_efectores

[918 rows x 14 columns]
```

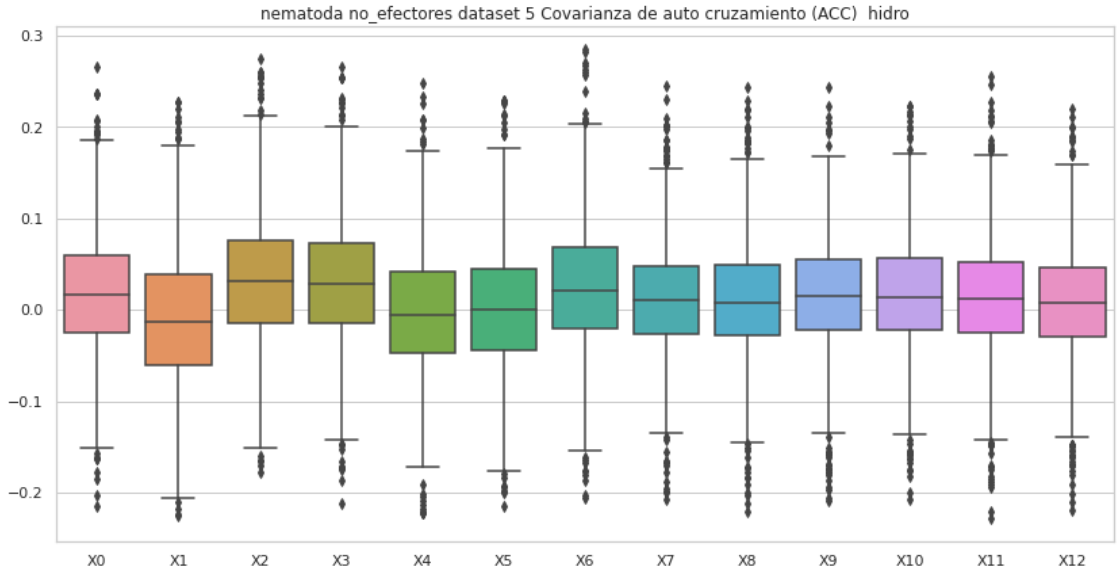
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 5, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.019452   -0.010332    0.033287    0.029356   -0.004300   -0.000301
std      0.070775    0.077301    0.071888    0.070685    0.071982    0.071352
min     -0.214289   -0.224518   -0.178209   -0.211713   -0.222834   -0.214651
25%     -0.024972   -0.060265   -0.015126   -0.014255   -0.048079   -0.043945
50%      0.016245   -0.013066    0.030953    0.027700   -0.006312    0.000227
75%      0.059880    0.038277    0.076121    0.072008    0.041100    0.045246
max      0.265174    0.227557    0.274972    0.266085    0.246976    0.228446

                X6          X7          X8          X9         X10         X11  \
count  918.000000  918.000000  918.000000  918.000000  918.000000  918.000000
mean     0.024003    0.010668    0.009805    0.015382    0.016617    0.013013
std      0.072998    0.065045    0.068533    0.065602    0.065475    0.068505
```

min	-0.206069	-0.207756	-0.220900	-0.209639	-0.206974	-0.228217
25%	-0.021112	-0.026004	-0.028374	-0.022212	-0.021561	-0.025530
50%	0.020905	0.010989	0.008213	0.014950	0.013714	0.012292
75%	0.068521	0.047827	0.049571	0.054800	0.056489	0.052253
max	0.284582	0.244128	0.242491	0.243910	0.222150	0.254710

	X12
count	918.000000
mean	0.006937
std	0.065183
min	-0.218724
25%	-0.030010
50%	0.007822
75%	0.045921
max	0.219826




nematoda no_efectores dataset 5 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12