ds4_nematoda_limpieza_de_datos

December 14, 2020

Limpieza de datos

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

1 Declaración de variables

[2]:
```
organismo ="nematoda"
dataset = 4
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 4, con valores
atípicos.
Valores del documento csv.

        X0     X1     X2     X3     X4      X5      X6     X7     X8     X9  \
0    1.795  2.821  4.359  1.026  0.513   4.872  31.026  0.769  2.051  4.872
1    7.463  6.343  2.985  2.985  3.731   5.597   4.478  6.343  3.731  9.701
2    8.166  5.701  2.465  5.547  2.465   6.163   3.544  3.082  4.006  6.471
3    4.819  7.229  5.622  3.815  4.016   4.418   3.414  5.020  2.610  6.426
4    6.797  5.882  5.490  6.797  1.176   9.412   4.052  4.183  2.353  6.928
..     …      …      …      …      …       …       …      …      …
495  7.399  5.489  4.535  5.012  1.432   7.637   2.864  4.057  3.580  6.444
496  4.255  8.511  4.255  6.383  2.128  11.702   3.191  4.787  2.660  3.191
497  7.692  7.240  4.525  3.620  0.905   4.072   4.072  6.787  4.072  8.597
498  6.667  7.879  2.424  4.242  4.242   4.242   1.818  9.091  4.242  4.848
```

```
499  4.895  5.944  3.846  6.643  0.350   8.042   1.049  6.643  2.098  8.042

     …     X11    X12    X13    X14    X15    X16    X17    X18    X19  \
0    …   5.897  2.821  2.821  3.077  4.359  5.385  0.000  0.513  3.590
1    …   4.104  2.612  2.612  3.358  5.597  6.343  2.239  2.985  5.597
2    …   7.088  4.622  3.852  5.393  6.009  4.777  0.616  3.082  6.009
3    …   2.811  1.406  7.229  6.426  9.639  6.024  1.004  2.008  7.229
4    …   8.497  2.222  3.399  1.699  6.144  4.706  0.654  4.052  5.752
..   …     …    …    …    …    …    …    …    …    …
495  …   5.967  2.387  6.205  3.580  7.876  5.728  0.477  3.580  5.489
496  …  12.234  5.851  3.723  2.660  6.915  2.660  0.532  3.191  6.383
497  …   7.240  3.167  3.167  4.072  5.882  6.335  1.357  2.715  4.072
498  …   3.030  3.636  7.273  4.242  6.667  6.667  1.212  1.818  6.061
499  …   9.091  2.098  5.944  4.895  4.895  3.846  1.049  3.497  4.895

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

[500 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     7.078322    6.069014    4.228156    5.017088    2.426000    6.198264
std      3.016393    2.660361    1.933351    2.167032    2.073409    3.018674
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      5.362250    4.512500    3.004250    3.530500    1.225250    4.348000
50%      6.770500    5.752000    4.125000    5.187000    1.942000    5.837000
75%      8.592250    7.426250    5.235000    6.349000    3.079750    7.882250
max     30.400000   19.101000   11.688000   15.385000   14.844000   23.438000

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     3.830904    5.787726    2.484834    5.705706    8.891136    5.916308
```

```
std      2.583322    3.911075    1.504228    2.423381    3.022545    3.265887
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      2.375250    3.738000    1.503500    4.196000    7.143000    3.753750
50%      3.508000    5.235500    2.292000    5.545500    8.696000    5.415500
75%      4.727250    6.963750    3.260250    7.064250   10.593750    7.547000
max     31.026000   55.670000    9.091000   15.385000   20.833000   37.870000

               X12         X13         X14         X15         X16         X17  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     2.893974    4.218564    4.949380    7.829622    5.648902    1.202980
std      1.630186    2.070328    3.152553    2.977459    2.460813    1.047696
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      1.880750    2.827750    3.231750    5.920750    4.297250    0.542500
50%      2.569000    4.108000    4.523000    7.413500    5.367000    1.057000
75%      3.575000    5.449000    6.062250    9.474500    6.635750    1.710250
max     12.500000   12.583000   31.065000   19.672000   21.918000    8.333000

               X18         X19
count  500.000000  500.000000
mean     3.301686    6.321532
std      2.392134    2.355625
min      0.000000    0.000000
25%      2.035750    4.762000
50%      2.965500    6.186000
75%      4.130000    7.692000
max     27.869000   20.370000
```

no_efectores

Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, con valores atípicos.

Valores del documento csv.

```
         X0      X1      X2     X3     X4     X5     X6     X7     X8      X9  \
0    11.538   2.564   5.128  2.564  0.000  6.410  5.128  6.410  5.128   5.128
1     2.963   4.444   2.963  5.185  1.481  8.889  5.926  3.704  0.741   2.963
2     6.383   1.064   4.255  5.319  6.383  3.191  9.574  5.319  1.064  11.702
3     3.297   5.495   7.692  3.297  5.495  3.297  5.495  4.396  2.198   7.692
4     8.374   2.709   2.709  3.695  2.709  5.911  1.970  7.635  7.882   4.926
..      …       …       …      …      …      …      …      …      …       …
495   8.036  12.500   5.357  2.679  1.786  3.571  1.786  6.250  0.893   7.143
496   0.000   2.151  13.978  6.452  0.000  9.677  4.301  1.075  2.151   9.677
497   5.161   3.871   5.806  4.516  1.935  9.677  4.516  3.226  4.516   6.452
498   2.681   5.362   8.043  7.239  0.536  6.434  3.485  4.021  0.804   5.898
499   5.435   6.739   5.870  4.783  1.087  7.391  2.391  4.565  1.957  10.435

     …     X11    X12    X13    X14     X15    X16    X17    X18    X19  \
```

```
0    …  25.641  3.846  1.282  3.846   0.000  3.846  0.000  0.000  8.974
1    …   8.889  2.963  7.407  5.185  10.370  6.667  0.000  3.704  6.667
2    …  12.766  1.064  6.383  0.000   8.511  2.128  1.064  3.191  2.128
3    …  14.286  2.198  4.396  3.297   7.692  6.593  1.099  6.593  2.198
4    …   3.448  4.433  5.172  3.448   7.389  5.172  0.246  1.478  7.143
..   …     …    …    …    …     …    …    …    …    …
495  …   4.464  2.679  4.464  2.679   9.821  7.143  0.000  5.357  4.464
496  …  12.903  2.151  4.301  2.151  13.978  2.151  0.000  2.151  1.075
497  …   5.806  3.226  5.806  2.581   2.581  3.871  0.000  5.806  5.806
498  …   7.507  3.217  2.681  5.362  12.332  6.166  2.145  1.609  5.362
499  …   5.435  2.174  6.304  2.391   6.304  6.739  2.174  2.826  6.739

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

[500 rows x 21 columns]
```

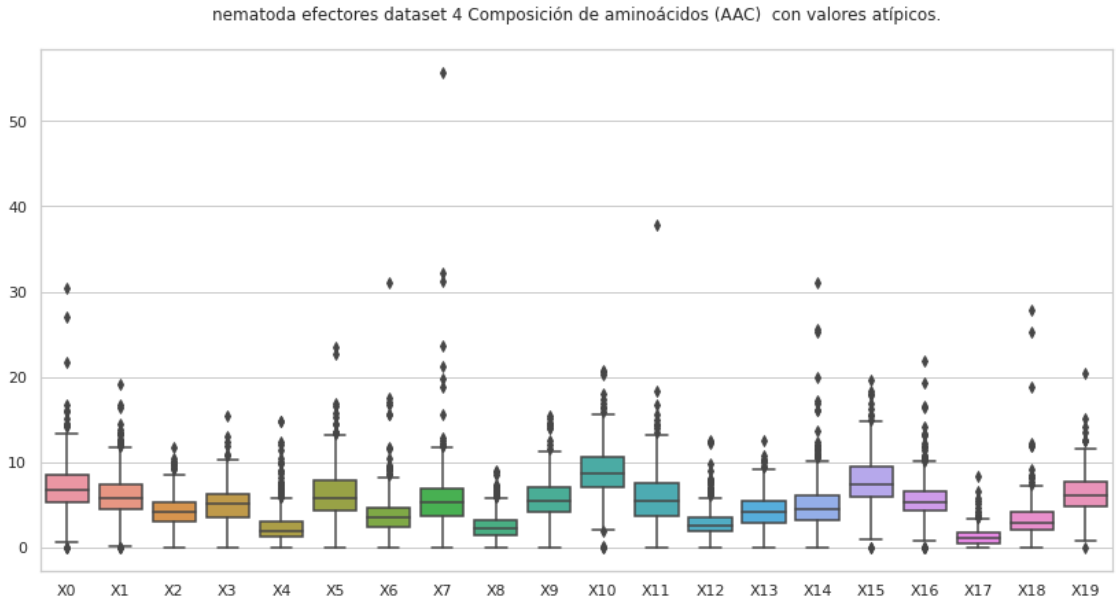
Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

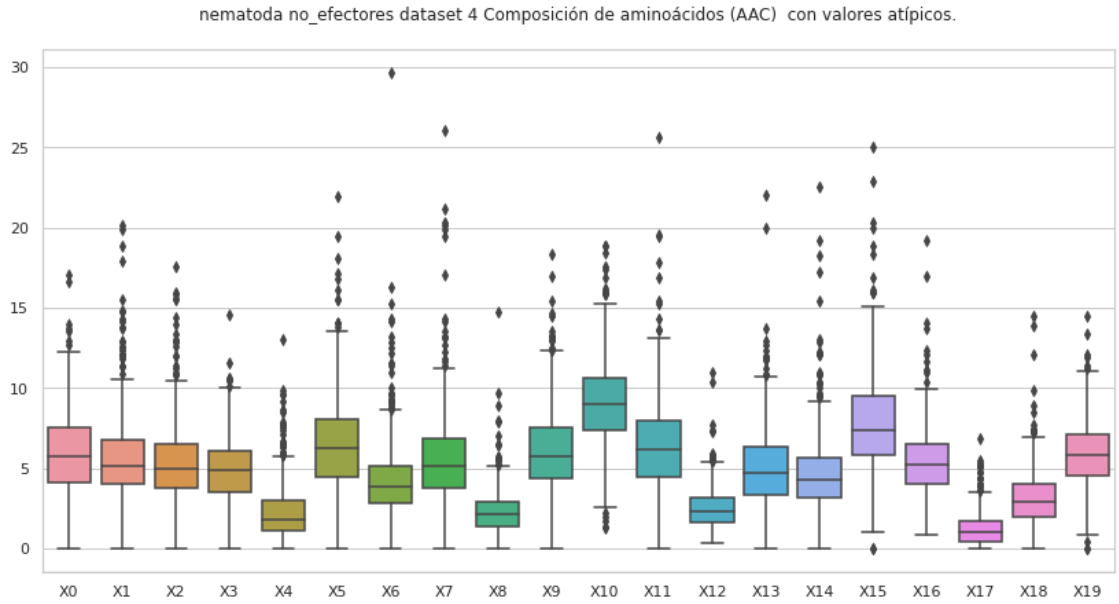
```
                X0          X1          X2          X3          X4          X5  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      5.940320    5.592456    5.381820    4.885208    2.269586    6.450014
std       2.559892    2.832871    2.511947    2.035037    1.821853    3.013807
min       0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%       4.118250    4.023250    3.808500    3.494000    1.106500    4.449250
50%       5.714000    5.172000    4.950000    4.925000    1.830000    6.283000
75%       7.537000    6.749250    6.539000    6.101500    2.957250    8.082000
max      17.045000   20.144000   17.544000   14.545000   13.000000   21.975000

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      4.251664    5.491208    2.259932    6.154378    9.138794    6.446390
std       2.491003    2.967416    1.440426    2.682986    2.809700    3.053874
min       0.000000    0.000000    0.000000    0.000000    1.266000    0.000000
25%       2.791500    3.772500    1.358250    4.396000    7.396000    4.490750
```

50%	3.896000	5.125500	2.145000	5.754000	9.007500	6.136500
75%	5.128000	6.825750	2.883250	7.577000	10.635500	7.985000
max	29.647000	26.087000	14.768000	18.310000	18.841000	25.641000

	X12	X13	X14	X15	X16	X17	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	2.529078	5.018232	4.676862	7.884540	5.347748	1.231944	
std	1.285755	2.515803	2.533571	3.126905	2.121322	1.057799	
min	0.379000	0.000000	0.000000	0.000000	0.844000	0.000000	
25%	1.606000	3.332500	3.205250	5.797000	4.008250	0.466000	
50%	2.352500	4.715500	4.284500	7.357500	5.204000	1.061500	
75%	3.150500	6.320000	5.635750	9.524000	6.494000	1.697500	
max	10.959000	22.011000	22.581000	25.000000	19.205000	6.897000	

	X18	X19
count	500.000000	500.000000
mean	3.152068	5.897734
std	1.811745	2.144770
min	0.000000	0.000000
25%	1.973500	4.545000
50%	2.948000	5.807000
75%	4.000000	7.143000
max	14.493000	14.493000





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 4, sin valores
atípicos.
Valores del documento csv.

        X0     X1     X2     X3     X4      X5     X6     X7     X8     X9  \
1    7.463  6.343  2.985  2.985  3.731   5.597  4.478  6.343  3.731  9.701
2    8.166  5.701  2.465  5.547  2.465   6.163  3.544  3.082  4.006  6.471
3    4.819  7.229  5.622  3.815  4.016   4.418  3.414  5.020  2.610  6.426
4    6.797  5.882  5.490  6.797  1.176   9.412  4.052  4.183  2.353  6.928
5    8.116  3.188  6.957  6.667  1.159   5.507  2.609  5.797  0.870  6.087
..     …      …      …      …      …       …      …      …      …      …
495  7.399  5.489  4.535  5.012  1.432   7.637  2.864  4.057  3.580  6.444
496  4.255  8.511  4.255  6.383  2.128  11.702  3.191  4.787  2.660  3.191
497  7.692  7.240  4.525  3.620  0.905   4.072  4.072  6.787  4.072  8.597
498  6.667  7.879  2.424  4.242  4.242   4.242  1.818  9.091  4.242  4.848
499  4.895  5.944  3.846  6.643  0.350   8.042  1.049  6.643  2.098  8.042

     …     X11    X12    X13    X14    X15    X16    X17    X18     X19  \
1    …   4.104  2.612  2.612  3.358  5.597  6.343  2.239  2.985   5.597
2    …   7.088  4.622  3.852  5.393  6.009  4.777  0.616  3.082   6.009
3    …   2.811  1.406  7.229  6.426  9.639  6.024  1.004  2.008   7.229
4    …   8.497  2.222  3.399  1.699  6.144  4.706  0.654  4.052   5.752
5    …   6.087  2.899  4.348  2.319  8.116  8.406  0.580  2.609  10.145
..   …     …      …      …      …      …      …      …      …       …
```

```
495  …   5.967  2.387  6.205  3.580  7.876  5.728  0.477  3.580   5.489
496  …  12.234  5.851  3.723  2.660  6.915  2.660  0.532  3.191   6.383
497  …   7.240  3.167  3.167  4.072  5.882  6.335  1.357  2.715   4.072
498  …   3.030  3.636  7.273  4.242  6.667  6.667  1.212  1.818   6.061
499  …   9.091  2.098  5.944  4.895  4.895  3.846  1.049  3.497   4.895

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

[413 rows x 21 columns]
```

Composición de aminoácidos (AAC) efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  413.000000  413.000000  413.000000  413.000000  413.000000  413.000000
mean     6.953797    6.015114    4.318799    5.199061    2.260240    6.381973
std      2.330069    2.303826    1.782314    1.835352    1.428819    2.451587
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.971000
25%      5.515000    4.630000    3.147000    3.980000    1.299000    4.769000
50%      6.842000    5.794000    4.167000    5.357000    2.030000    6.077000
75%      8.290000    7.330000    5.215000    6.383000    2.947000    7.968000
max     15.942000   13.402000   10.000000   10.881000    7.729000   14.516000

                X6          X7          X8          X9         X10         X11  \
count  413.000000  413.000000  413.000000  413.000000  413.000000  413.000000
mean     3.634402    5.518981    2.514559    5.880952    9.138620    5.896232
std      1.732949    2.252924    1.311178    2.031460    2.613531    2.645768
min      0.000000    0.000000    0.000000    0.833000    0.000000    0.000000
25%      2.439000    3.855000    1.639000    4.469000    7.418000    4.018000
50%      3.507000    5.263000    2.353000    5.747000    9.015000    5.484000
75%      4.624000    6.787000    3.279000    7.205000   10.714000    7.377000
max     10.490000   12.857000    6.780000   12.500000   16.883000   14.925000

               X12         X13         X14         X15         X16         X17  \
count  413.000000  413.000000  413.000000  413.000000  413.000000  413.000000
```

mean	2.804441	4.463690	4.694649	7.881738	5.559935	1.203048
std	1.289798	1.869171	2.177730	2.786345	1.914165	0.893205
min	0.000000	0.000000	0.000000	1.010000	0.000000	0.000000
25%	1.942000	3.147000	3.258000	5.970000	4.451000	0.611000
50%	2.574000	4.301000	4.464000	7.451000	5.381000	1.072000
75%	3.521000	5.590000	5.814000	9.821000	6.534000	1.724000
max	7.778000	10.356000	12.108000	16.270000	12.286000	4.202000

	X18	X19
count	413.000000	413.000000
mean	3.200203	6.479639
std	1.558190	2.024788
min	0.000000	1.156000
25%	2.116000	5.089000
50%	2.997000	6.364000
75%	4.124000	7.692000
max	9.127000	12.621000

no_efectores
Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	\
1	2.963	4.444	2.963	5.185	1.481	8.889	5.926	3.704	0.741	2.963	
2	6.383	1.064	4.255	5.319	6.383	3.191	9.574	5.319	1.064	11.702	
3	3.297	5.495	7.692	3.297	5.495	3.297	5.495	4.396	2.198	7.692	
5	12.245	8.503	4.195	6.122	0.113	14.059	9.184	1.814	2.381	3.855	
6	3.871	10.323	5.161	9.032	0.000	10.323	4.516	3.871	1.935	2.581	
..	…	…	…	…	…	…	…	…	…	…	
494	8.025	8.642	3.704	0.617	3.086	4.938	6.173	5.556	1.235	6.790	
495	8.036	12.500	5.357	2.679	1.786	3.571	1.786	6.250	0.893	7.143	
497	5.161	3.871	5.806	4.516	1.935	9.677	4.516	3.226	4.516	6.452	
498	2.681	5.362	8.043	7.239	0.536	6.434	3.485	4.021	0.804	5.898	
499	5.435	6.739	5.870	4.783	1.087	7.391	2.391	4.565	1.957	10.435	

	…	X11	X12	X13	X14	X15	X16	X17	X18	X19	\
1	…	8.889	2.963	7.407	5.185	10.370	6.667	0.000	3.704	6.667	
2	…	12.766	1.064	6.383	0.000	8.511	2.128	1.064	3.191	2.128	
3	…	14.286	2.198	4.396	3.297	7.692	6.593	1.099	6.593	2.198	
5	…	7.937	1.587	0.680	0.567	4.875	3.061	0.113	1.134	4.422	
6	…	3.871	1.290	3.871	6.452	14.194	5.806	0.645	3.871	2.581	
..	…	…	…	…	…	…	…	…	…	…	
494	…	3.704	3.086	3.704	2.469	13.580	6.790	1.235	1.852	4.938	
495	…	4.464	2.679	4.464	2.679	9.821	7.143	0.000	5.357	4.464	
497	…	5.806	3.226	5.806	2.581	2.581	3.871	0.000	5.806	5.806	

```
498  …   7.507  3.217  2.681  5.362  12.332  6.166  2.145  1.609  5.362
499  …   5.435  2.174  6.304  2.391   6.304  6.739  2.174  2.826  6.739

              X20
1    no_efectores
2    no_efectores
3    no_efectores
5    no_efectores
6    no_efectores
..            …
494  no_efectores
495  no_efectores
497  no_efectores
498  no_efectores
499  no_efectores

[404 rows x 21 columns]
```

Composición de aminoácidos (AAC) no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

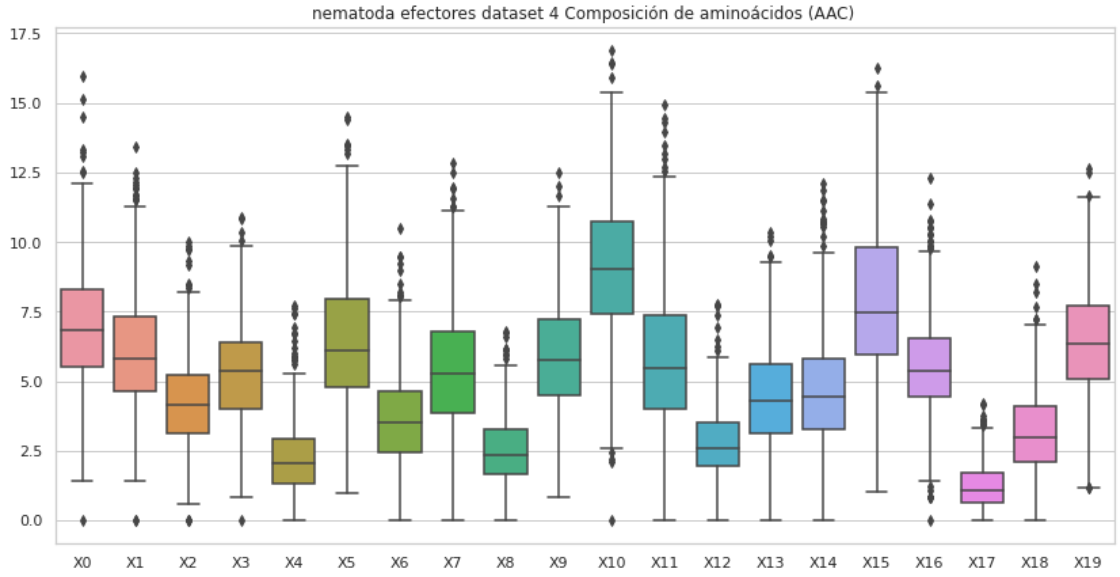
	X0	X1	X2	X3	X4	X5	\
count	404.000000	404.000000	404.000000	404.00000	404.000000	404.000000	
mean	5.936780	5.523998	5.241126	5.13777	2.177827	6.594876	
std	2.281499	2.085941	1.987599	1.78810	1.484526	2.439775	
min	1.087000	0.866000	0.806000	0.00000	0.000000	0.000000	
25%	4.343250	4.227250	3.961250	4.01800	1.151750	4.950000	
50%	5.740500	5.286000	4.940500	5.18050	1.853500	6.533500	
75%	7.362750	6.591500	6.420000	6.14525	2.807250	8.104750	
max	13.514000	13.768000	12.877000	10.60600	7.692000	14.082000	

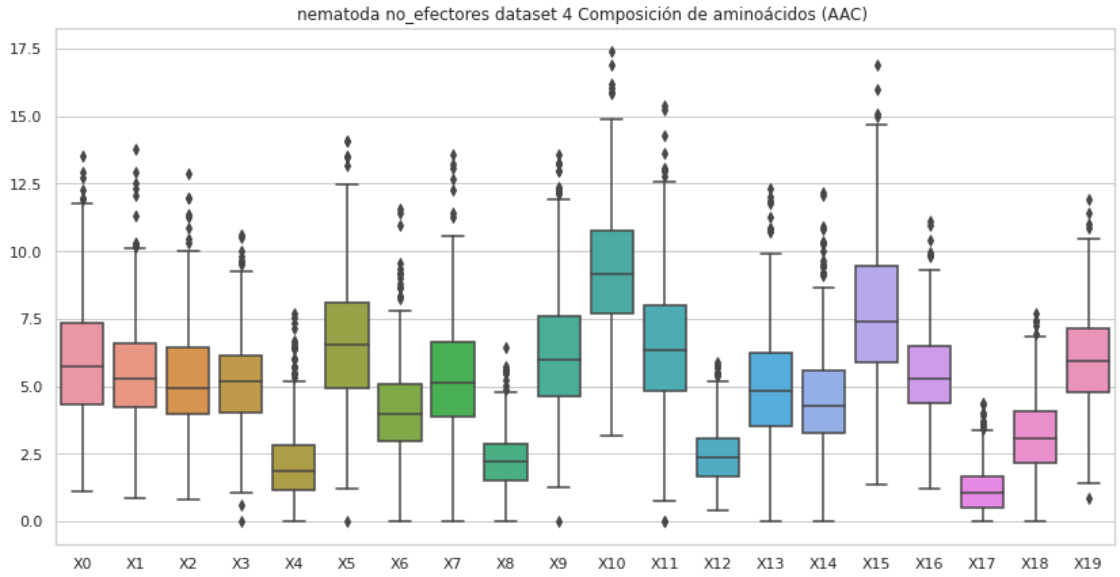
	X6	X7	X8	X9	X10	X11	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	4.137431	5.340366	2.253455	6.247050	9.325062	6.526621	
std	1.727217	2.138496	1.080492	2.410361	2.456405	2.605195	
min	0.000000	0.000000	0.000000	0.000000	3.175000	0.000000	
25%	2.972000	3.871000	1.527750	4.641250	7.696500	4.835000	
50%	3.962500	5.137500	2.193000	5.978500	9.161000	6.324000	
75%	5.057750	6.658000	2.842750	7.577000	10.750000	8.000000	
max	11.550000	13.580000	6.422000	13.576000	17.391000	15.385000	

	X12	X13	X14	X15	X16	X17	\
count	404.000000	404.000000	404.000000	404.000000	404.000000	404.000000	
mean	2.461245	4.989079	4.526361	7.784064	5.404394	1.165092	
std	1.092263	2.106766	1.944062	2.650299	1.707300	0.849674	
min	0.379000	0.000000	0.000000	1.370000	1.190000	0.000000	

25%	1.638500	3.498000	3.270000	5.891000	4.372000	0.506000
50%	2.352500	4.827500	4.284500	7.365500	5.294000	1.058000
75%	3.086000	6.214750	5.556000	9.441250	6.497750	1.674250
max	5.882000	12.329000	12.146000	16.892000	11.111000	4.396000

	X18	X19
count	404.000000	404.000000
mean	3.195230	6.032191
std	1.485085	1.937981
min	0.000000	0.847000
25%	2.142000	4.793750
50%	3.084500	5.943500
75%	4.054000	7.146000
max	7.671000	11.905000





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 4, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.009484	0.002710	0.005419	0.025743	0.014904	0.004065	0.010839	
1	0.056885	0.028443	0.022754	0.042664	0.019910	0.048352	0.028443	
2	0.064304	0.019412	0.043678	0.048531	0.030332	0.024266	0.031545	
3	0.015740	0.013116	0.012461	0.014428	0.023610	0.016396	0.008526	
4	0.039163	0.006778	0.039163	0.054225	0.019581	0.024100	0.013556	
..	…	…	…	…	…	…	…	
495	0.049165	0.009516	0.033305	0.050751	0.041235	0.026961	0.023789	
496	0.008540	0.004270	0.012810	0.023484	0.007472	0.009607	0.005337	
497	0.054985	0.006469	0.025875	0.029110	0.022641	0.048517	0.029110	
498	0.036731	0.023374	0.023374	0.023374	0.040070	0.050088	0.023374	
499	0.031972	0.002284	0.043390	0.052525	0.038823	0.043390	0.013702	

	X7	X8	X9	…	X74	X75	X76	\
0	0.025743	0.031162	0.092131	…	-0.006582	-0.002322	0.005185	
1	0.073951	0.031287	0.085328	…	0.044503	-0.012358	0.021168	
2	0.050958	0.055811	0.086143	…	-0.028719	0.010205	0.008395	
3	0.020986	0.009181	0.028856	…	0.000534	-0.001245	0.012452	
4	0.039916	0.048953	0.056485	…	0.003529	0.034409	0.015253	
..	…	…	…	…	…	…	…	
495	0.042821	0.039649	0.068196	…	0.014936	0.017298	0.026946	
496	0.006405	0.024552	0.009607	…	0.016584	0.035413	0.001535	
497	0.061454	0.051751	0.074392	…	-0.085717	-0.024341	0.009091	
498	0.026714	0.016696	0.053427	…	-0.008072	-0.016631	0.009312	
499	0.052525	0.059376	0.079930	…	0.011291	0.041894	0.021689	

	X77	X78	X79	X80	X81	X82	X83
0	0.023543	0.006198	0.012157	0.051436	0.016675	0.009882	efectores
1	0.023485	-0.008586	0.074855	-0.006360	0.007689	0.029175	efectores
2	0.040032	0.021009	0.023155	-0.004703	-0.007410	0.006830	efectores
3	0.012568	0.010776	0.016449	0.007001	0.002673	0.003125	efectores
4	-0.011430	0.022799	-0.005674	0.001476	0.014946	0.013294	efectores
..	…	…	…	…	…	…	…
495	-0.017405	0.000032	0.010747	0.004275	-0.004908	-0.005098	efectores
496	-0.003076	0.021799	-0.003532	0.023042	0.024783	-0.002184	efectores

```
497 -0.029664 -0.003875  0.033179 -0.038601 -0.044441  0.011890  efectores
498 -0.034865 -0.013903  0.049038 -0.011598 -0.032337  0.004556  efectores
499  0.021851  0.043277 -0.010367 -0.014000  0.020478  0.013313  efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.035504	0.012225	0.027991	0.032532	0.023351	0.028945	
std	0.093981	0.027334	0.042238	0.037744	0.082684	0.050548	
min	-1.227891	-0.445955	-0.445955	-0.445955	-1.473469	-0.891911	
25%	0.023544	0.005222	0.014916	0.018453	0.010880	0.016328	
50%	0.033277	0.010480	0.024977	0.029150	0.020937	0.026369	
75%	0.045462	0.017738	0.037468	0.043996	0.033515	0.038950	
max	1.308764	0.140631	0.480321	0.218568	0.747865	0.270180	

	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.012020	0.032118	0.029309	0.045150	…	0.007806	
std	0.066980	0.054811	0.107630	0.130044	…	0.048853	
min	-1.114889	-0.668933	-2.006799	-1.841836	…	-0.818050	
25%	0.005767	0.017123	0.016015	0.027685	…	0.000128	
50%	0.010936	0.028246	0.026898	0.044660	…	0.011283	
75%	0.018915	0.040610	0.041589	0.063494	…	0.023200	
max	0.560899	0.747865	1.121798	0.934831	…	0.141836	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.003592	0.003023	0.014591	0.004314	0.005218	0.012487	
std	0.044814	0.064073	0.070522	0.182358	0.140772	0.082335	
min	-0.381148	-0.981136	-0.863115	-1.680141	-1.642813	-1.058860	
25%	-0.008133	-0.005411	0.000792	-0.010756	-0.005302	0.000891	
50%	0.003961	0.006411	0.011554	0.003453	0.006799	0.012093	
75%	0.016774	0.020859	0.025204	0.015925	0.018189	0.023831	
max	0.256591	0.126809	0.866926	3.539541	2.517423	1.259305	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	-0.002991	0.002837	0.009663
std	0.137816	0.112263	0.061421
min	-2.637764	-2.033348	-0.677070
25%	-0.011798	-0.007007	-0.000221
50%	0.003259	0.006052	0.009770

```
75%      0.015826    0.017998    0.024073
max      1.206501    1.249137    0.833411

[8 rows x 83 columns]
```

```
no_efectores
Composición de pseudo aminoácidos (PseAAC)  hidro_mass no_efectores nematoda
dataset 4, con valores atípicos.
Valores del documento csv.
```

	X0	X1	X2	X3	X4	X5	X6	\
0	0.026753	0.000000	0.005945	0.014863	0.002973	0.014863	0.011890	
1	0.021146	0.010573	0.037006	0.063439	0.052866	0.026433	0.005287	
2	0.030624	0.030624	0.025520	0.015312	0.030624	0.025520	0.005104	
3	0.014637	0.024395	0.014637	0.014637	0.019516	0.019516	0.009758	
4	0.023607	0.007637	0.010415	0.016663	0.014581	0.021524	0.022218	
..	…	…	…	…	…	…	…	
495	0.019081	0.004240	0.006360	0.008480	0.010601	0.014841	0.002120	
496	0.000000	0.000000	0.015613	0.023420	0.010409	0.002602	0.005204	
497	0.042227	0.015835	0.036949	0.079176	0.047505	0.026392	0.036949	
498	0.011317	0.002263	0.030557	0.027161	0.011317	0.016976	0.003395	
499	0.031136	0.006227	0.027399	0.042345	0.036117	0.026154	0.011209	

	X7	X8	X9	…	X74	X75	X76	\
0	0.011890	0.059451	0.005945	…	0.000058	0.022358	0.017932	
1	0.021146	0.063439	0.063439	…	-0.006111	0.062541	-0.010966	
2	0.056144	0.061248	0.040832	…	0.007171	0.011965	-0.005600	
3	0.034154	0.063428	0.014637	…	0.037093	0.029841	0.009403	
4	0.013886	0.009720	0.038187	…	0.014104	-0.002535	0.008638	
..	…	…	…	…	…	…	…	
495	0.016961	0.010601	0.021201	…	0.037309	0.018925	0.020604	
496	0.023420	0.031226	0.023420	…	0.020197	0.041822	-0.004732	
497	0.052784	0.047505	0.121403	…	0.028858	0.040223	0.029465	
498	0.024898	0.031688	0.038479	…	0.008383	0.009355	0.018789	
499	0.059781	0.031136	0.047326	…	-0.013857	0.011521	0.002764	

	X77	X78	X79	X80	X81	X82	X83
0	0.010004	0.020081	0.002376	0.015672	0.025134	0.005404	no_efectores
1	-0.051913	0.004850	-0.030758	-0.062627	0.027220	-0.026197	no_efectores
2	0.011878	-0.009692	0.004800	-0.018735	-0.025044	0.016202	no_efectores
3	0.030223	0.001448	0.018868	-0.029601	-0.018949	0.019438	no_efectores
4	0.004110	-0.004637	0.020361	0.007299	-0.000690	0.013438	no_efectores
..	…	…	…	…	…	…	…
495	0.030327	0.028074	-0.014946	-0.005381	0.000305	0.014526	no_efectores
496	-0.004004	0.031412	-0.002649	0.032669	0.029324	0.003304	no_efectores
497	-0.055642	-0.017564	-0.023648	0.037139	0.007915	-0.063889	no_efectores

```
498  0.006627  0.014938  0.013105  0.004567  0.016807  0.008657  no_efectores
499 -0.009985 -0.009743 -0.014758 -0.002890  0.011736 -0.000821  no_efectores

[500 rows x 84 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, con valores atípicos.
Estadísticas.

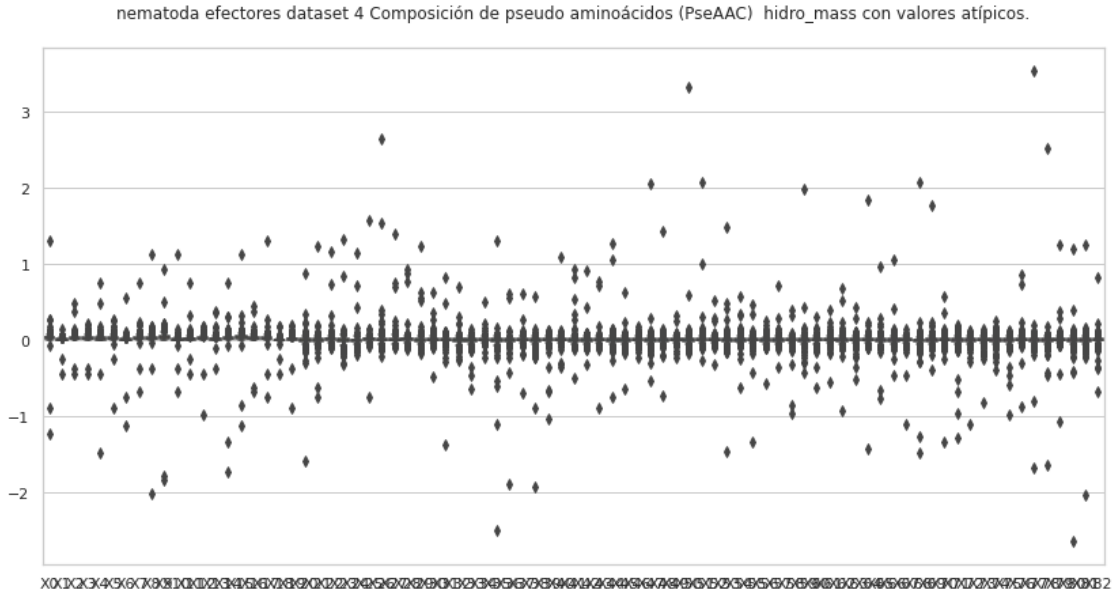
	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.031143	0.012369	0.027790	0.035983	0.029076	0.029433	
std	0.027379	0.026683	0.020114	0.028047	0.045351	0.026293	
min	-0.364511	-0.486015	0.000000	-0.121504	-0.729022	-0.243007	
25%	0.019181	0.004333	0.015306	0.018910	0.013737	0.016717	
50%	0.028576	0.009018	0.024785	0.031373	0.024002	0.025608	
75%	0.039784	0.017384	0.036633	0.046982	0.037999	0.035979	
max	0.167424	0.145248	0.248351	0.239020	0.261446	0.203347	

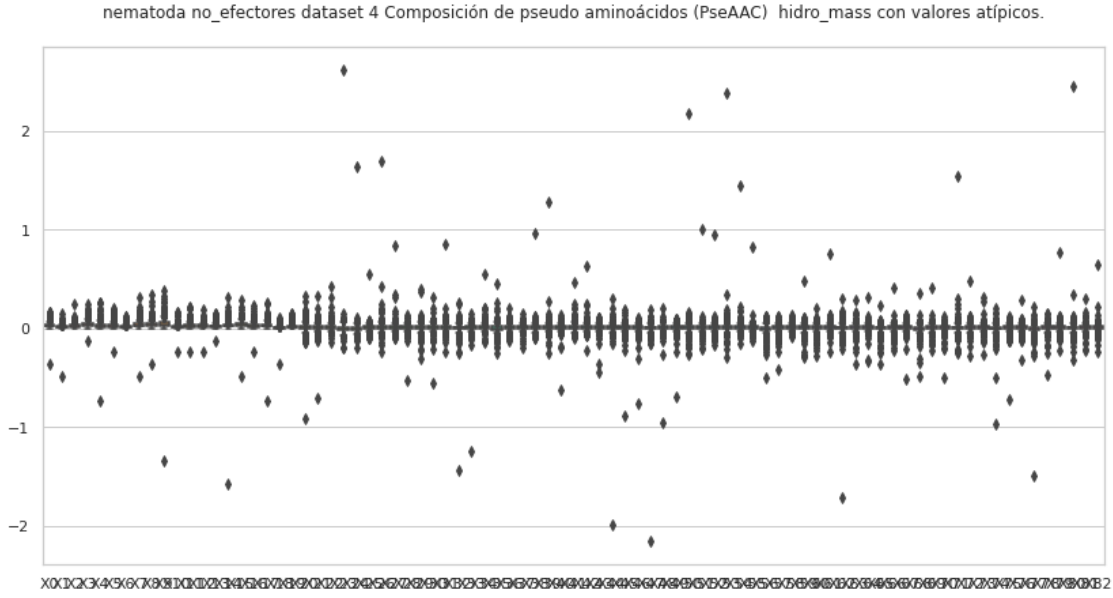
	X6	X7	X8	X9	…	X73	\
count	500.000000	500.000000	500.000000	500.000000	…	500.000000	
mean	0.013875	0.035012	0.036104	0.051537	…	0.007923	
std	0.014802	0.037896	0.035976	0.075870	…	0.033273	
min	0.000000	-0.486015	-0.364511	-1.336540	…	-0.272624	
25%	0.005643	0.017487	0.018943	0.029933	…	-0.001653	
50%	0.010404	0.029123	0.031496	0.043665	…	0.009069	
75%	0.017018	0.045032	0.045612	0.064135	…	0.019242	
max	0.126551	0.313970	0.340795	0.378661	…	0.316673	

	X74	X75	X76	X77	X78	X79	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	-0.002873	0.004916	0.006963	-0.004016	0.003821	0.010657	
std	0.059237	0.040925	0.031401	0.076450	0.037826	0.042309	
min	-0.966898	-0.717322	-0.317851	-1.494626	-0.472014	-0.224041	
25%	-0.009337	-0.002458	-0.000857	-0.011710	-0.004705	0.000536	
50%	0.002598	0.007480	0.008406	0.002825	0.006309	0.009874	
75%	0.012923	0.018339	0.019380	0.013229	0.018177	0.020433	
max	0.203358	0.118019	0.291419	0.218513	0.150869	0.770672	

	X80	X81	X82
count	500.000000	500.000000	500.000000
mean	0.007467	0.008393	0.010111
std	0.115601	0.032782	0.039900
min	-0.320083	-0.224047	-0.234177
25%	-0.006897	-0.002505	0.000844
50%	0.004484	0.009349	0.010345
75%	0.016463	0.021380	0.021276

```
max       2.445776     0.294724     0.641759

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

	X0	X1	X2	X3	X4	X5	X6	\
1	0.056885	0.028443	0.022754	0.042664	0.019910	0.048352	0.028443	
2	0.064304	0.019412	0.043678	0.048531	0.030332	0.024266	0.031545	
3	0.015740	0.013116	0.012461	0.014428	0.023610	0.016396	0.008526	
4	0.039163	0.006778	0.039163	0.054225	0.019581	0.024100	0.013556	
5	0.036753	0.005250	0.030190	0.024939	0.019689	0.026252	0.003938	
..	…	…	…	…	…	…	…	
495	0.049165	0.009516	0.033305	0.050751	0.041235	0.026961	0.023789	
496	0.008540	0.004270	0.012810	0.023484	0.007472	0.009607	0.005337	
497	0.054985	0.006469	0.025875	0.029110	0.022641	0.048517	0.029110	
498	0.036731	0.023374	0.023374	0.023374	0.040070	0.050088	0.023374	
499	0.031972	0.002284	0.043390	0.052525	0.038823	0.043390	0.013702	

	X7	X8	X9	…	X74	X75	X76	\
1	0.073951	0.031287	0.085328	…	0.044503	-0.012358	0.021168	
2	0.050958	0.055811	0.086143	…	-0.028719	0.010205	0.008395	
3	0.020986	0.009181	0.028856	…	0.000534	-0.001245	0.012452	
4	0.039916	0.048953	0.056485	…	0.003529	0.034409	0.015253	
5	0.027565	0.027565	0.034128	…	0.000528	0.011106	0.015441	
..	…	…	…	…	…	…	…	
495	0.042821	0.039649	0.068196	…	0.014936	0.017298	0.026946	
496	0.006405	0.024552	0.009607	…	0.016584	0.035413	0.001535	
497	0.061454	0.051751	0.074392	…	-0.085717	-0.024341	0.009091	
498	0.026714	0.016696	0.053427	…	-0.008072	-0.016631	0.009312	
499	0.052525	0.059376	0.079930	…	0.011291	0.041894	0.021689	

	X77	X78	X79	X80	X81	X82	X83
1	0.023485	-0.008586	0.074855	-0.006360	0.007689	0.029175	efectores
2	0.040032	0.021009	0.023155	-0.004703	-0.007410	0.006830	efectores
3	0.012568	0.010776	0.016449	0.007001	0.002673	0.003125	efectores
4	-0.011430	0.022799	-0.005674	0.001476	0.014946	0.013294	efectores
5	0.005429	-0.004663	0.014633	-0.000900	-0.013163	0.021987	efectores
..	…	…	…	…	…	…	
495	-0.017405	0.000032	0.010747	0.004275	-0.004908	-0.005098	efectores
496	-0.003076	0.021799	-0.003532	0.023042	0.024783	-0.002184	efectores
497	-0.029664	-0.003875	0.033179	-0.038601	-0.044441	0.011890	efectores
498	-0.034865	-0.013903	0.049038	-0.011598	-0.032337	0.004556	efectores
499	0.021851	0.043277	-0.010367	-0.014000	0.020478	0.013313	efectores

[470 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass efectores nematoda
dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	470.000000	470.000000	470.000000	470.000000	470.000000	470.000000	
mean	0.035031	0.012877	0.026401	0.031649	0.022524	0.027723	
std	0.017519	0.012084	0.015709	0.018777	0.015854	0.015611	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.023484	0.005259	0.014897	0.018413	0.010575	0.015875	
50%	0.032854	0.010215	0.024535	0.028286	0.019799	0.025569	
75%	0.044030	0.016775	0.035424	0.042670	0.031433	0.036984	
max	0.181827	0.078097	0.109096	0.133340	0.102009	0.093508	

	X6	X7	X8	X9	…	X73	\
count	470.000000	470.000000	470.000000	470.000000	…	470.000000	
mean	0.013211	0.029646	0.029386	0.046573	…	0.011464	
std	0.011094	0.017720	0.018098	0.026386	…	0.019438	
min	0.000000	0.000000	0.000000	0.000000	…	-0.050655	
25%	0.005753	0.016956	0.015524	0.027446	…	0.001280	
50%	0.010644	0.027480	0.025936	0.042792	…	0.011643	
75%	0.017154	0.038725	0.039638	0.060736	…	0.022315	
max	0.076829	0.100541	0.113868	0.191027	…	0.091655	

	X74	X75	X76	X77	X78	X79	\
count	470.000000	470.000000	470.000000	470.000000	470.000000	470.000000	
mean	0.005917	0.008922	0.012797	0.003133	0.006922	0.011762	
std	0.026771	0.023148	0.020872	0.027715	0.024544	0.020738	
min	-0.100945	-0.069071	-0.047719	-0.128009	-0.197717	-0.069122	
25%	-0.007447	-0.004540	0.001209	-0.009304	-0.003856	0.001573	
50%	0.004241	0.006780	0.011281	0.003597	0.007263	0.012068	
75%	0.016802	0.020939	0.024417	0.014790	0.018169	0.023350	
max	0.135006	0.126809	0.096610	0.117325	0.090385	0.105496	

	X80	X81	X82
count	470.000000	470.000000	470.000000
mean	0.002439	0.006280	0.011750
std	0.026733	0.023717	0.021720
min	-0.153934	-0.123056	-0.129744
25%	-0.010224	-0.005450	0.001026
50%	0.003707	0.006682	0.010044
75%	0.015719	0.017845	0.024001
max	0.117034	0.091968	0.102648

[8 rows x 83 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.026753	0.000000	0.005945	0.014863	0.002973	0.014863	0.011890	
1	0.021146	0.010573	0.037006	0.063439	0.052866	0.026433	0.005287	
2	0.030624	0.030624	0.025520	0.015312	0.030624	0.025520	0.005104	
3	0.014637	0.024395	0.014637	0.014637	0.019516	0.019516	0.009758	
4	0.023607	0.007637	0.010415	0.016663	0.014581	0.021524	0.022218	
..	…	…	…	…	…	…	…	
495	0.019081	0.004240	0.006360	0.008480	0.010601	0.014841	0.002120	
496	0.000000	0.000000	0.015613	0.023420	0.010409	0.002602	0.005204	
497	0.042227	0.015835	0.036949	0.079176	0.047505	0.026392	0.036949	
498	0.011317	0.002263	0.030557	0.027161	0.011317	0.016976	0.003395	
499	0.031136	0.006227	0.027399	0.042345	0.036117	0.026154	0.011209	

	X7	X8	X9	…	X74	X75	X76	\
0	0.011890	0.059451	0.005945	…	0.000058	0.022358	0.017932	
1	0.021146	0.063439	0.063439	…	-0.006111	0.062541	-0.010966	
2	0.056144	0.061248	0.040832	…	0.007171	0.011965	-0.005600	
3	0.034154	0.063428	0.014637	…	0.037093	0.029841	0.009403	
4	0.013886	0.009720	0.038187	…	0.014104	-0.002535	0.008638	
..	…	…	…	…	…	…	…	
495	0.016961	0.010601	0.021201	…	0.037309	0.018925	0.020604	
496	0.023420	0.031226	0.023420	…	0.020197	0.041822	-0.004732	
497	0.052784	0.047505	0.121403	…	0.028858	0.040223	0.029465	
498	0.024898	0.031688	0.038479	…	0.008383	0.009355	0.018789	
499	0.059781	0.031136	0.047326	…	-0.013857	0.011521	0.002764	

	X77	X78	X79	X80	X81	X82	X83
0	0.010004	0.020081	0.002376	0.015672	0.025134	0.005404	no_efectores
1	-0.051913	0.004850	-0.030758	-0.062627	0.027220	-0.026197	no_efectores
2	0.011878	-0.009692	0.004800	-0.018735	-0.025044	0.016202	no_efectores
3	0.030223	0.001448	0.018868	-0.029601	-0.018949	0.019438	no_efectores
4	0.004110	-0.004637	0.020361	0.007299	-0.000690	0.013438	no_efectores
..	…	…	…	…	…	…	…
495	0.030327	0.028074	-0.014946	-0.005381	0.000305	0.014526	no_efectores
496	-0.004004	0.031412	-0.002649	0.032669	0.029324	0.003304	no_efectores
497	-0.055642	-0.017564	-0.023648	0.037139	0.007915	-0.063889	no_efectores
498	0.006627	0.014938	0.013105	0.004567	0.016807	0.008657	no_efectores
499	-0.009985	-0.009743	-0.014758	-0.002890	0.011736	-0.000821	no_efectores

[451 rows x 84 columns]

Composición de pseudo aminoácidos (PseAAC) hidro_mass no_efectores nematoda
dataset 4, sin valores atípicos.
Estadísticas.

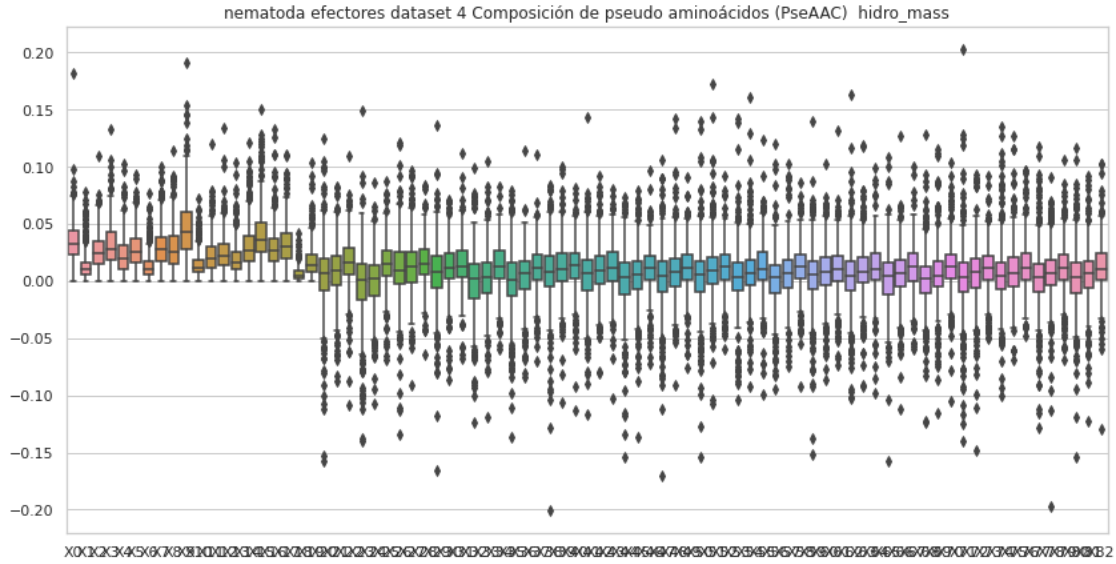
	X0	X1	X2	X3	X4	X5	\
count	451.000000	451.000000	451.000000	451.000000	451.000000	451.000000	
mean	0.028571	0.011793	0.025018	0.031952	0.025559	0.026431	
std	0.013669	0.010779	0.014149	0.018260	0.016917	0.014070	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.018360	0.004302	0.014447	0.018218	0.012847	0.016269	
50%	0.027707	0.008400	0.023443	0.029940	0.022496	0.024712	
75%	0.037370	0.015943	0.034328	0.044099	0.034399	0.034217	
max	0.078458	0.072761	0.083176	0.105508	0.100789	0.087065	

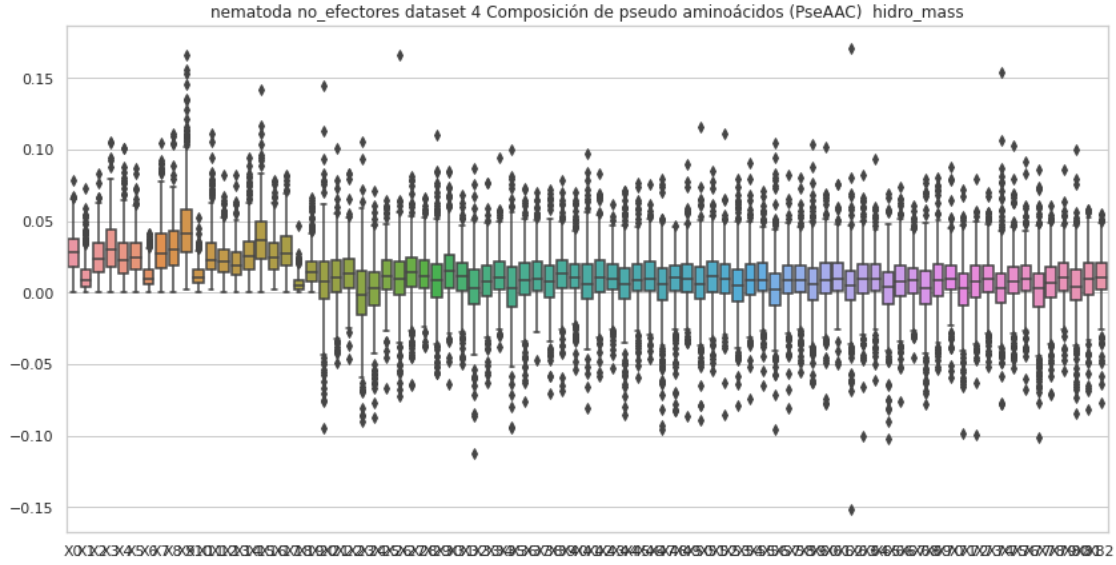
	X6	X7	X8	X9	…	X73	\
count	451.000000	451.000000	451.000000	451.000000	…	451.000000	
mean	0.011277	0.031370	0.032178	0.046069	…	0.009044	
std	0.008110	0.018937	0.018727	0.025771	…	0.017350	
min	0.000000	0.000000	0.000000	0.002348	…	-0.052136	
25%	0.005464	0.016798	0.018451	0.028418	…	-0.000202	
50%	0.009807	0.027327	0.029722	0.040970	…	0.009576	
75%	0.015664	0.041619	0.042695	0.057996	…	0.019046	
max	0.041498	0.104264	0.111215	0.165539	…	0.069678	

	X74	X75	X76	X77	X78	X79	\
count	451.000000	451.000000	451.000000	451.000000	451.000000	451.000000	
mean	0.003405	0.008068	0.009632	0.001223	0.006326	0.010296	
std	0.022812	0.019081	0.017360	0.022906	0.019173	0.016362	
min	-0.078448	-0.052588	-0.067395	-0.100878	-0.071546	-0.054696	
25%	-0.007302	-0.001120	0.000703	-0.010297	-0.003768	0.001531	
50%	0.003429	0.007810	0.009138	0.003155	0.006624	0.010216	
75%	0.012946	0.017863	0.019251	0.012984	0.017639	0.020262	
max	0.153711	0.103045	0.091704	0.086280	0.060821	0.085981	

	X80	X81	X82
count	451.000000	451.000000	451.000000
mean	0.004159	0.008477	0.010283
std	0.021445	0.019733	0.017346
min	-0.084466	-0.081921	-0.076652
25%	-0.005750	-0.001951	0.001850
50%	0.004339	0.009138	0.010398
75%	0.015871	0.020315	0.020645
max	0.099464	0.057627	0.055539

[8 rows x 83 columns]


nematoda efectores dataset 4 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20


nematoda no_efectores dataset 4 Composición de pseudo aminoácidos (PseAAC) hidro_mass
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15

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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.014014  0.004004  0.008008  0.038038  0.022022  0.006006  0.016016
1    0.049563  0.024782  0.019825  0.037172  0.017347  0.042129  0.024782
2    0.063488  0.019166  0.043124  0.047915  0.029947  0.023958  0.031145
3    0.030813  0.025677  0.024393  0.028245  0.046219  0.032096  0.016690
4    0.059807  0.010351  0.059807  0.082810  0.029904  0.036805  0.020703
..        …         …         …         …         …         …         …
495  0.055731  0.010787  0.037753  0.057528  0.046742  0.030562  0.026966
```

```
496  0.039800  0.019900  0.059699  0.109449  0.034825  0.044775  0.024875
497  0.056048  0.006594  0.026375  0.029672  0.023078  0.049454  0.029672
498  0.044382  0.028243  0.028243  0.028243  0.048416  0.060520  0.028243
499  0.041170  0.002941  0.055873  0.067636  0.049992  0.055873  0.017644

          X7        X8        X9  …       X32       X33       X34  \
0    0.038038  0.046046  0.136136  …  0.008002  0.016129 -0.000133
1    0.064432  0.027260  0.074344  …  0.029623  0.005525 -0.048522
2    0.050311  0.055103  0.085050  …  0.015025 -0.014201 -0.006890
3    0.041083  0.017974  0.056490  …  0.017134  0.017833  0.005448
4    0.060958  0.074759  0.086261  …  0.002217  0.025039 -0.003161
..        …         …         …  …         …         …         …
495  0.048540  0.044944  0.077304  …  0.003257  0.011664  0.001980
496  0.029850  0.114424  0.044775  …  0.007986 -0.004807 -0.010993
497  0.062641  0.052751  0.075829  …  0.039356 -0.018896  0.015645
498  0.032277  0.020173  0.064555  … -0.004592 -0.017043  0.010810
499  0.067636  0.076458  0.102925  …  0.068603 -0.001901  0.027613

          X35       X36       X37       X38       X39       X40        X41
0    0.007889  0.007962  0.011842  0.007661  0.017963  0.014603  efectores
1    0.022512  0.039377  0.004523  0.018443  0.065220  0.025420  efectores
2    0.019783 -0.001932  0.007936  0.008288  0.022861  0.006743  efectores
3    0.017829  0.020139  0.020475  0.024376  0.032200  0.006118  efectores
4    0.007612 -0.008852  0.013990  0.023293 -0.008666  0.020302  efectores
..        …         …         …         …         …         …         …
495  0.015257  0.000503  0.005190  0.030545  0.012183 -0.005779  efectores
496  0.027231 -0.005796  0.001902  0.007155 -0.016459 -0.010179  efectores
497  0.037919  0.047508  0.005813  0.009266  0.033820  0.012120  efectores
498 -0.002832  0.008804 -0.000417  0.011251  0.059252  0.005504  efectores
499 -0.009943  0.038463 -0.019438  0.027928 -0.013349  0.017144  efectores

[500 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.046299    0.016396    0.035367    0.044748    0.030902    0.037110
std      0.022448    0.014809    0.019294    0.028317    0.020223    0.024362
min     -0.204382    0.000000    0.000000    0.000000    0.000000   -0.136255
25%      0.034782    0.007433    0.021984    0.026107    0.016809    0.025123
50%      0.044960    0.014102    0.034035    0.039665    0.026766    0.034697
75%      0.056058    0.020446    0.045918    0.057552    0.041434    0.045436
max      0.139827    0.141917    0.150348    0.195979    0.125403    0.375977
```

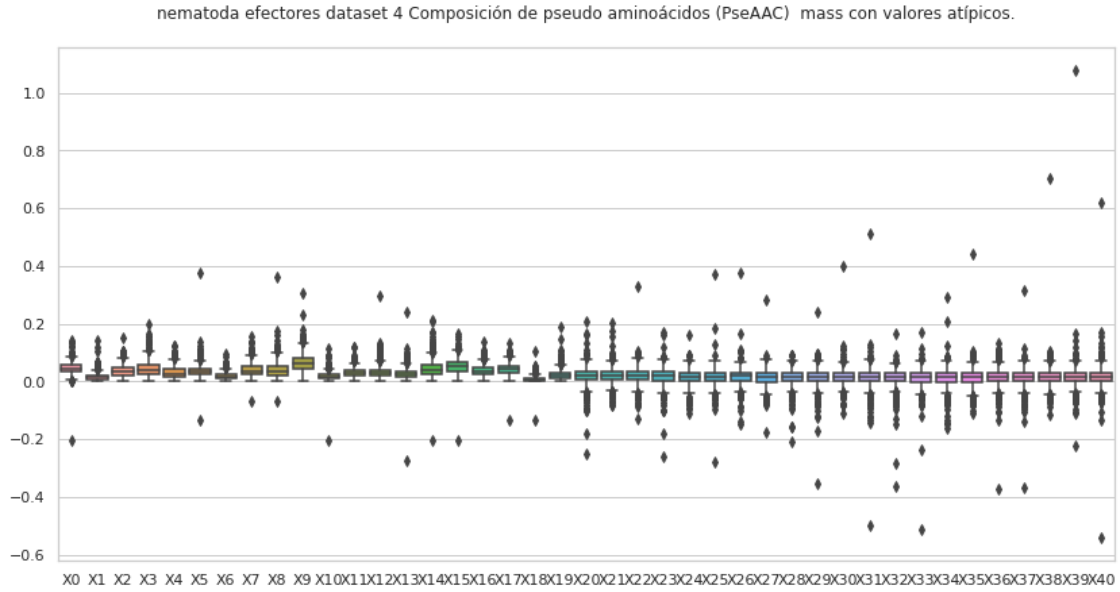
```
              X6          X7          X8          X9  …         X31  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.017998    0.040487    0.042583    0.063531  …    0.013287
std      0.013519    0.023346    0.031567    0.032582  …    0.044553
min      0.000000   -0.068127   -0.068127   -0.000000  …   -0.499552
25%      0.008581    0.025368    0.022393    0.042687  …    0.001559
50%      0.015508    0.036492    0.035640    0.061263  …    0.016154
75%      0.023580    0.052613    0.053935    0.078829  …    0.031143
max      0.097085    0.155350    0.361855    0.304534  …    0.509454

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011841    0.010462    0.012475    0.013431    0.013558    0.012232
std      0.035591    0.039107    0.034195    0.034626    0.035074    0.035734
min     -0.365109   -0.514880   -0.161048   -0.109462   -0.373318   -0.368806
25%      0.002285   -0.001743   -0.000971   -0.001568    0.000945    0.000144
50%      0.016549    0.014973    0.014810    0.016236    0.016286    0.016139
75%      0.028163    0.027769    0.029152    0.029215    0.028185    0.028903
max      0.163749    0.170563    0.289542    0.441415    0.133850    0.315308

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.014985    0.015762    0.013880
std      0.042200    0.056571    0.048041
min     -0.114169   -0.221106   -0.541268
25%      0.000085    0.000834   -0.000209
50%      0.015248    0.016227    0.014165
75%      0.029481    0.029227    0.029596
max      0.704171    1.077177    0.616717

[8 rows x 41 columns]
```

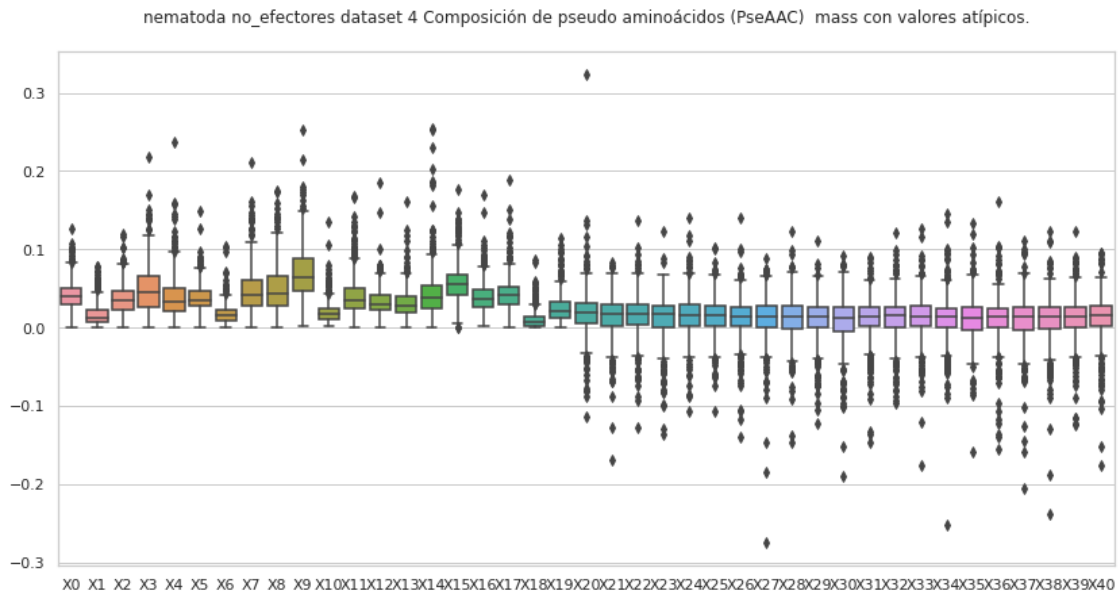
no_efectores
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.077797  0.000000  0.017288  0.043221  0.008644  0.043221  0.034576
1    0.028412  0.014206  0.049721  0.085236  0.071030  0.035515  0.007103
2    0.051197  0.051197  0.042664  0.025598  0.051197  0.042664  0.008533
3    0.033402  0.055670  0.033402  0.033402  0.044536  0.044536  0.022268
4    0.037300  0.012068  0.016456  0.026329  0.023038  0.034008  0.035105
..        …         …         …         …         …         …         …
495  0.054175  0.012039  0.018058  0.024078  0.030097  0.042136  0.006019
496  0.000000  0.000000  0.059320  0.088980  0.039547  0.009887  0.019773
```

```
497  0.043321  0.016246  0.037906  0.081228  0.048737  0.027076  0.037906
498  0.017065  0.003413  0.046075  0.040955  0.017065  0.025597  0.005119
499  0.052957  0.010591  0.046602  0.072021  0.061430  0.044484  0.019064

           X7        X8        X9  …       X32       X33       X34  \
0    0.034576  0.172882  0.017288  …  0.079838  0.011299  0.019006
1    0.028412  0.085236  0.085236  …  0.054352 -0.007614 -0.043819
2    0.093861  0.102394  0.068263  …  0.024423 -0.008958  0.014824
3    0.077938  0.144741  0.033402  … -0.022844  0.002730  0.054061
4    0.021941  0.015359  0.060338  …  0.020280  0.016992  0.020453
..        …         …         …  …       …         …         …
495  0.048156  0.030097  0.060195  … -0.036717  0.029467  0.071004
496  0.088980  0.118640  0.088980  …  0.029493  0.003241 -0.023991
497  0.054152  0.048737  0.124549  …  0.019366 -0.018891 -0.022269
498  0.037542  0.047781  0.058020  …  0.016623  0.007162  0.011716
499  0.101677  0.052957  0.080494  …  0.024037 -0.007457  0.012214

          X35       X36       X37       X38       X39       X40           X41
0    0.046231  0.014581  0.020536  0.052147  0.006909  0.015715  no_efectores
1   -0.039296  0.010535 -0.018789 -0.014734 -0.041325 -0.035198  no_efectores
2    0.071190 -0.063525  0.093061 -0.009362  0.008025  0.027086  no_efectores
3   -0.026868 -0.009529  0.043709  0.021457  0.043056  0.044356  no_efectores
4    0.023956  0.024232  0.022338  0.013648  0.032172  0.021233  no_efectores
..        …         …         …         …         …         …           …
495 -0.016438  0.025369  0.044135  0.058499 -0.042434  0.041241  no_efectores
496  0.019820  0.026390 -0.006248 -0.017980 -0.010066  0.012553  no_efectores
497 -0.013033  0.020898  0.020559  0.030228 -0.024261 -0.065545  no_efectores
498  0.002144  0.010509  0.019278  0.028331  0.019761  0.013054  no_efectores
499 -0.023556  0.024152 -0.012577  0.004700 -0.025102 -0.001396  no_efectores

[500 rows x 42 columns]


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

              X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.041592    0.016639    0.036357    0.048863    0.039400    0.037780
std      0.017935    0.014074    0.019148    0.028903    0.026950    0.017214
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.029698    0.006931    0.022925    0.026730    0.020397    0.027651
50%      0.040195    0.012712    0.034931    0.044324    0.033772    0.035548
75%      0.051019    0.023072    0.046540    0.065551    0.051160    0.047004
max      0.125872    0.077797    0.119958    0.217904    0.237470    0.148127

              X6          X7          X8          X9  …         X31  \
```

```
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.017049    0.047113    0.049420    0.069193  …    0.010797
std      0.012777    0.027883    0.029826    0.032436  …    0.027014
min      0.000000    0.000000    0.000000    0.002397  …   -0.147485
25%      0.009219    0.027179    0.028189    0.047403  …    0.001253
50%      0.015219    0.042031    0.044286    0.063503  …    0.013789
75%      0.022493    0.060494    0.065716    0.088367  …    0.025737
max      0.103946    0.211055    0.174706    0.252470  …    0.089618

              X32         X33         X34         X35         X36         X37  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.011373    0.012894    0.012410    0.010550    0.011155    0.009854
std      0.026797    0.026844    0.029284    0.026523    0.028908    0.030220
min     -0.095750   -0.175980   -0.252347   -0.159146   -0.155800   -0.204963
25%     -0.000520    0.001599    0.000653   -0.003871    0.000207   -0.003020
50%      0.015172    0.013689    0.014830    0.012681    0.013696    0.013785
75%      0.026226    0.028089    0.025202    0.025664    0.025138    0.026023
max      0.120698    0.125689    0.145358    0.133175    0.160360    0.111020

              X38         X39         X40
count  500.000000  500.000000  500.000000
mean     0.010094    0.011984    0.012007
std      0.030903    0.027231    0.028536
min     -0.238965   -0.123362   -0.176058
25%     -0.002033    0.000707    0.001549
50%      0.014030    0.013979    0.015843
75%      0.025777    0.026984    0.027527
max      0.122453    0.123213    0.095523

[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 4,
sin valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
1    0.049563  0.024782  0.019825  0.037172  0.017347  0.042129  0.024782
2    0.063488  0.019166  0.043124  0.047915  0.029947  0.023958  0.031145
3    0.030813  0.025677  0.024393  0.028245  0.046219  0.032096  0.016690
4    0.059807  0.010351  0.059807  0.082810  0.029904  0.036805  0.020703
```

```
5    0.039912  0.005702  0.032785  0.027083  0.021381  0.028508  0.004276
..        …         …         …         …         …         …
495  0.055731  0.010787  0.037753  0.057528  0.046742  0.030562  0.026966
496  0.039800  0.019900  0.059699  0.109449  0.034825  0.044775  0.024875
497  0.056048  0.006594  0.026375  0.029672  0.023078  0.049454  0.029672
498  0.044382  0.028243  0.028243  0.028243  0.048416  0.060520  0.028243
499  0.041170  0.002941  0.055873  0.067636  0.049992  0.055873  0.017644

           X7        X8        X9  …       X32       X33       X34  \
1    0.064432  0.027260  0.074344  …  0.029623  0.005525 -0.048522
2    0.050311  0.055103  0.085050  …  0.015025 -0.014201 -0.006890
3    0.041083  0.017974  0.056490  …  0.017134  0.017833  0.005448
4    0.060958  0.074759  0.086261  …  0.002217  0.025039 -0.003161
5    0.029934  0.029934  0.037061  …  0.017115  0.035424  0.023930
..        …         …         …  …         …         …         …
495  0.048540  0.044944  0.077304  …  0.003257  0.011664  0.001980
496  0.029850  0.114424  0.044775  …  0.007986 -0.004807 -0.010993
497  0.062641  0.052751  0.075829  …  0.039356 -0.018896  0.015645
498  0.032277  0.020173  0.064555  … -0.004592 -0.017043  0.010810
499  0.067636  0.076458  0.102925  …  0.068603 -0.001901  0.027613

           X35       X36       X37       X38       X39       X40         X41
1    0.022512  0.039377  0.004523  0.018443  0.065220  0.025420  efectores
2    0.019783 -0.001932  0.007936  0.008288  0.022861  0.006743  efectores
3    0.017829  0.020139  0.020475  0.024376  0.032200  0.006118  efectores
4    0.007612 -0.008852  0.013990  0.023293 -0.008666  0.020302  efectores
5    0.027092  0.030689  0.012638  0.016768  0.015891  0.023877  efectores
..        …         …         …         …         …         …         …
495  0.015257  0.000503  0.005190  0.030545  0.012183 -0.005779  efectores
496  0.027231 -0.005796  0.001902  0.007155 -0.016459 -0.010179  efectores
497  0.037919  0.047508  0.005813  0.009266  0.033820  0.012120  efectores
498 -0.002832  0.008804 -0.000417  0.011251  0.059252  0.005504  efectores
499 -0.009943  0.038463 -0.019438  0.027928 -0.013349  0.017144  efectores

[415 rows x 42 columns]
```

Composición de pseudo aminoácidos (PseAAC) mass efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  415.000000  415.000000  415.000000  415.000000  415.000000  415.000000
mean     0.044690    0.015010    0.033560    0.041138    0.028808    0.034886
std      0.015218    0.010234    0.016232    0.022252    0.017098    0.013475
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.035226    0.007607    0.021972    0.025560    0.016648    0.025233
50%      0.043859    0.014035    0.033396    0.037460    0.025909    0.034246
```

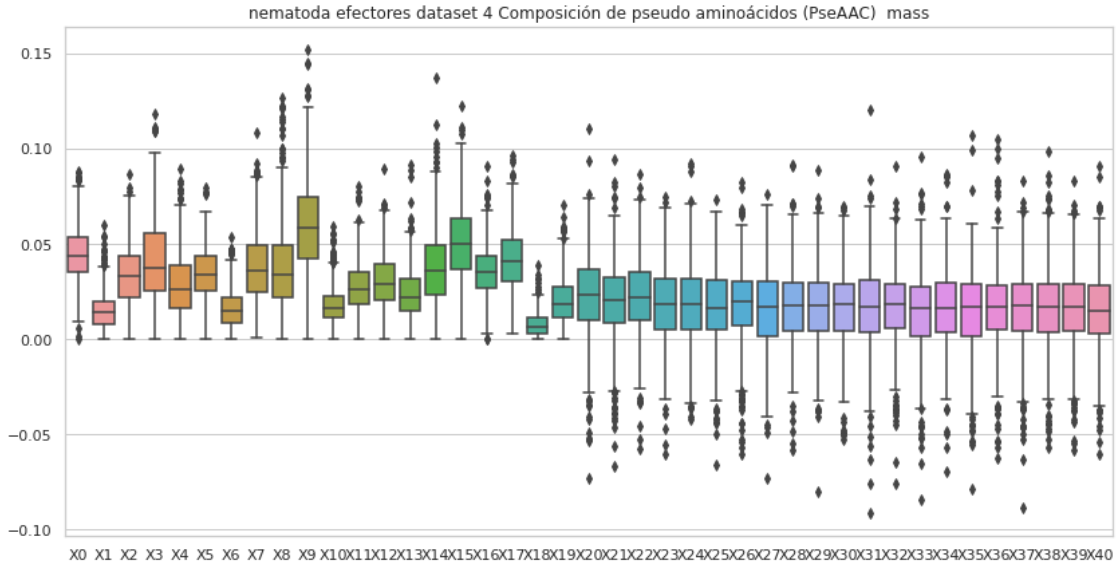
```
75%      0.053881    0.019883    0.043869    0.055454    0.038715    0.043782
max      0.088005    0.059731    0.086810    0.118344    0.089666    0.079495

                X6          X7          X8          X9  …         X31  \
count  415.000000  415.000000  415.000000  415.000000  …  415.000000
mean     0.016198    0.037906    0.038669    0.059563  …    0.016146
std      0.010377    0.018608    0.023465    0.026421  …    0.023309
min      0.000000    0.000593    0.000000    0.000000  …   -0.091301
25%      0.008313    0.025044    0.021716    0.042256  …    0.003406
50%      0.014676    0.035938    0.033966    0.058870  …    0.016790
75%      0.021862    0.049208    0.049491    0.075053  …    0.031034
max      0.053783    0.108318    0.126514    0.152029  …    0.120162

               X32         X33         X34         X35         X36         X37  \
count  415.000000  415.000000  415.000000  415.000000  415.000000  415.000000
mean     0.016943    0.014054    0.016104    0.014978    0.016902    0.015375
std      0.020636    0.023170    0.021851    0.022964    0.022730    0.022108
min     -0.075524   -0.084469   -0.069198   -0.078493   -0.062412   -0.088274
25%      0.005469    0.001836    0.003386    0.001494    0.004934    0.004212
50%      0.018693    0.016248    0.016198    0.016926    0.017309    0.017409
75%      0.028736    0.027346    0.029621    0.028909    0.028197    0.029260
max      0.090666    0.095871    0.086653    0.106720    0.104750    0.083314

               X38         X39         X40
count  415.000000  415.000000  415.000000
mean     0.016115    0.015907    0.014327
std      0.022185    0.021507    0.022567
min     -0.057110   -0.058073   -0.060190
25%      0.003882    0.004627    0.003020
50%      0.016919    0.017169    0.015049
75%      0.029102    0.029103    0.028244
max      0.098612    0.083192    0.091061

[8 rows x 41 columns]
```

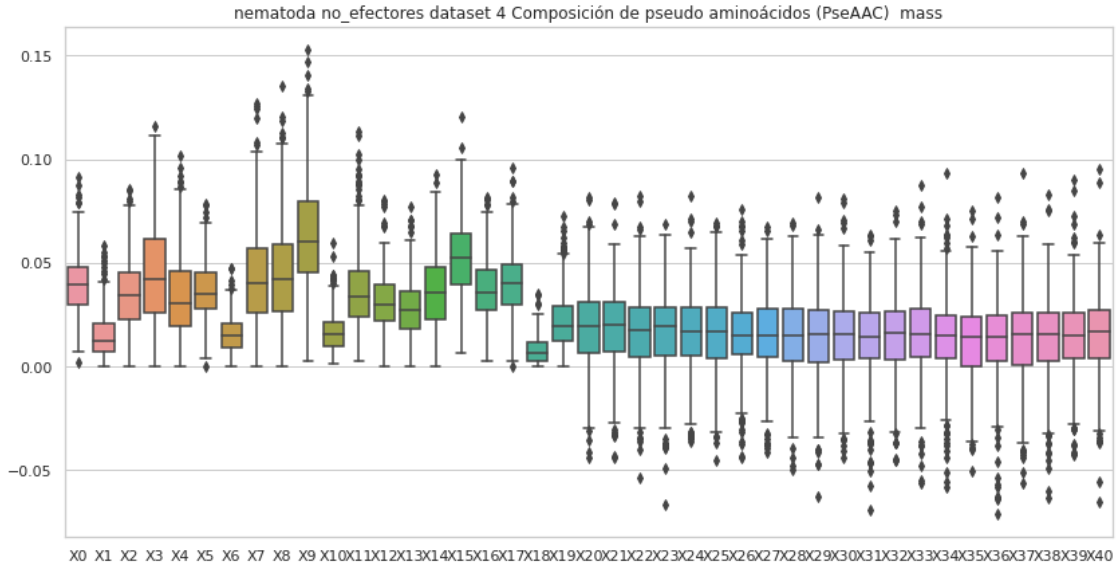
Composición de pseudo aminoácidos (PseAAC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

```
          X0        X1        X2        X3        X4        X5        X6  \
1    0.028412  0.014206  0.049721  0.085236  0.071030  0.035515  0.007103
2    0.051197  0.051197  0.042664  0.025598  0.051197  0.042664  0.008533
4    0.037300  0.012068  0.016456  0.026329  0.023038  0.034008  0.035105
5    0.091033  0.000843  0.045516  0.104519  0.005057  0.013486  0.017701
9    0.013599  0.013599  0.030597  0.054395  0.037396  0.030597  0.016998
..        …         …         …         …         …         …         …
```

```
494  0.054920  0.021123  0.004225  0.033797  0.025348  0.038021  0.008449
495  0.054175  0.012039  0.018058  0.024078  0.030097  0.042136  0.006019
497  0.043321  0.016246  0.037906  0.081228  0.048737  0.027076  0.037906
498  0.017065  0.003413  0.046075  0.040955  0.017065  0.025597  0.005119
499  0.052957  0.010591  0.046602  0.072021  0.061430  0.044484  0.019064

          X7        X8        X9  …       X32       X33       X34  \
1    0.028412  0.085236  0.085236  …  0.054352 -0.007614 -0.043819
2    0.093861  0.102394  0.068263  …  0.024423 -0.008958  0.014824
4    0.021941  0.015359  0.060338  …  0.020280  0.016992  0.020453
5    0.028659  0.059003  0.097776  …  0.011649  0.015355  0.011241
9    0.044196  0.047595  0.044196  … -0.007299  0.000256  0.010171
..        …         …         …  …         …         …         …
494  0.046470  0.025348  0.067593  … -0.014099  0.042493  0.016529
495  0.048156  0.030097  0.060195  … -0.036717  0.029467  0.071004
497  0.054152  0.048737  0.124549  …  0.019366 -0.018891 -0.022269
498  0.037542  0.047781  0.058020  …  0.016623  0.007162  0.011716
499  0.101677  0.052957  0.080494  …  0.024037 -0.007457  0.012214

          X35       X36       X37       X38       X39       X40           X41
1   -0.039296  0.010535 -0.018789 -0.014734 -0.041325 -0.035198  no_efectores
2    0.071190 -0.063525  0.093061 -0.009362  0.008025  0.027086  no_efectores
4    0.023956  0.024232  0.022338  0.013648  0.032172  0.021233  no_efectores
5    0.026389 -0.004645  0.019032  0.009457 -0.007216  0.016158  no_efectores
9    0.003278 -0.007332  0.005602  0.025163  0.011528  0.006362  no_efectores
..        …         …         …         …         …         …           …
494 -0.004670 -0.008639 -0.006252 -0.011744  0.006777  0.018227  no_efectores
495 -0.016438  0.025369  0.044135  0.058499 -0.042434  0.041241  no_efectores
497 -0.013033  0.020898  0.020559  0.030228 -0.024261 -0.065545  no_efectores
498  0.002144  0.010509  0.019278  0.028331  0.019761  0.013054  no_efectores
499 -0.023556  0.024152 -0.012577  0.004700 -0.025102 -0.001396  no_efectores

[394 rows x 42 columns]


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

                X0          X1          X2          X3          X4          X5  \
count   394.000000  394.000000  394.000000  394.000000  394.000000  394.000000
mean      0.039881    0.015216    0.035094    0.045993    0.034182    0.036376
std       0.014300    0.011150    0.016576    0.024304    0.019717    0.013411
min       0.001938    0.000000    0.000000    0.000000    0.000000    0.000000
25%       0.030015    0.007247    0.023025    0.026165    0.019396    0.027863
50%       0.039538    0.012242    0.034478    0.042369    0.030550    0.034905
75%       0.048236    0.021065    0.045375    0.061462    0.046289    0.045101
max       0.091033    0.058223    0.085678    0.115842    0.101702    0.078671
```

```
              X6          X7          X8          X9  …         X31  \
count  394.000000  394.000000  394.000000  394.000000  …  394.000000
mean     0.015561    0.043530    0.045466    0.064074  …    0.014096
std      0.008848    0.022734    0.025210    0.027223  …    0.019129
min      0.000000    0.000000    0.000000    0.002397  …   -0.069317
25%      0.009389    0.026219    0.026497    0.045093  …    0.004074
50%      0.014681    0.040281    0.041928    0.060435  …    0.014601
75%      0.020573    0.057287    0.059236    0.079500  …    0.025824
max      0.047473    0.127145    0.135105    0.152786  …    0.063413

              X32         X33         X34         X35         X36         X37  \
count  394.000000  394.000000  394.000000  394.000000  394.000000  394.000000
mean     0.014088    0.015271    0.014064    0.012315    0.012849    0.013009
std      0.019751    0.019046    0.019941    0.019235    0.019764    0.021721
min     -0.045219   -0.056338   -0.058435   -0.050584   -0.071120   -0.055918
25%      0.003176    0.004503    0.003651    0.000059    0.002774    0.000881
50%      0.016461    0.015934    0.015051    0.014031    0.014468    0.015300
75%      0.026679    0.027881    0.024861    0.024245    0.024908    0.026228
max      0.075189    0.087406    0.093355    0.075328    0.081693    0.093061

              X38         X39         X40
count  394.000000  394.000000  394.000000
mean     0.013822    0.014793    0.015247
std      0.020678    0.019200    0.019747
min     -0.063180   -0.042434   -0.065545
25%      0.002535    0.004002    0.003905
50%      0.015774    0.015168    0.016869
75%      0.026040    0.026029    0.027493
max      0.082804    0.090326    0.095523

[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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.011136  0.003182  0.006364  0.030227  0.017500  0.004773  0.012727
1    0.092567  0.046284  0.037027  0.069425  0.032399  0.078682  0.046284
2    0.083019  0.025062  0.056391  0.062656  0.039160  0.031328  0.040727
3    0.019294  0.016078  0.015274  0.017686  0.028941  0.020098  0.010451
4    0.042507  0.007357  0.042507  0.058855  0.021253  0.026158  0.014714
..        …         …         …         …         …         …         …
495  0.062844  0.012163  0.042572  0.064871  0.052708  0.034463  0.030408
496  0.008660  0.004330  0.012990  0.023815  0.007578  0.009743  0.005413
497  0.074936  0.008816  0.035264  0.039672  0.030856  0.066120  0.039672
498  0.050780  0.032314  0.032314  0.032314  0.055396  0.069245  0.032314
499  0.036475  0.002605  0.049501  0.059922  0.044290  0.049501  0.015632

           X7        X8        X9  …        X53       X54       X55  \
0    0.030227  0.036591  0.108181  …   0.004934  0.001574 -0.004453
1    0.120337  0.050912  0.138851  … -0.006544  0.059292  0.032058
```

```
2    0.065789  0.072055  0.111215  …  0.008120 -0.011489 -0.003575
3    0.025725  0.011255  0.035372  …  0.004121 -0.002145  0.002388
4    0.043324  0.053133  0.061308  …  0.009366 -0.010307  0.003890
..        …         …         …  …         …         …
495  0.054735  0.050680  0.087170  …  0.005392 -0.016871 -0.013994
496  0.006495  0.024898  0.009743  …  0.029677  0.016853  0.022779
497  0.083752  0.070528  0.101385  …  0.042724  0.060737  0.020951
498  0.036931  0.023082  0.073861  … -0.025008 -0.021939 -0.047187
499  0.059922  0.067738  0.091186  …  0.019699  0.000811  0.011838

          X56       X57       X58       X59       X60       X61        X62
0   -0.007728 -0.002726  0.027645  0.007278  0.060397  0.019580  efectores
1    0.072419 -0.020110  0.038217 -0.013973 -0.010349  0.012512  efectores
2   -0.037078  0.013175  0.051683  0.027123 -0.006072 -0.009567  efectores
3    0.000654 -0.001526  0.015407  0.013209  0.008582  0.003276  efectores
4    0.003831  0.037347 -0.012406  0.024746  0.001602  0.016222  efectores
..        …         …         …         …         …         …         …
495  0.019092  0.022111 -0.022248  0.000041  0.005464 -0.006274  efectores
496  0.016817  0.035912 -0.003120  0.022106  0.023367  0.025132  efectores
497 -0.116819 -0.033172 -0.040427 -0.005280 -0.052607 -0.060567  efectores
498 -0.011160 -0.022992 -0.048199 -0.019221 -0.016034 -0.044705  efectores
499  0.012881  0.047794  0.024928  0.049372 -0.015972  0.023362  efectores

[500 rows x 63 columns]
```

Composición de pseudo aminoácidos (PseAAC) hidro efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.055278    0.020399    0.039769    0.045905    0.037181    0.044631
std      0.076880    0.024023    0.030781    0.030842    0.070418    0.045540
min     -0.873086    0.000000    0.000000    0.000000    0.000000   -0.582058
25%      0.030668    0.006707    0.020174    0.026267    0.016248    0.022299
50%      0.048637    0.014278    0.036985    0.041644    0.030638    0.036462
75%      0.070119    0.025065    0.051458    0.059799    0.046653    0.060927
max      1.224777    0.244955    0.367433    0.367433    1.469733    0.285428

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.021963    0.044630    0.044443    0.072948  …   -0.002249
std      0.039659    0.036014    0.044568    0.091474  …    0.125939
min      0.000000   -0.291029   -0.291029   -0.000000  …   -2.064160
25%      0.008315    0.024556    0.023265    0.039064  …   -0.016696
50%      0.014900    0.039226    0.039056    0.063327  …    0.002635
75%      0.026274    0.058738    0.056675    0.089160  …    0.016704
```

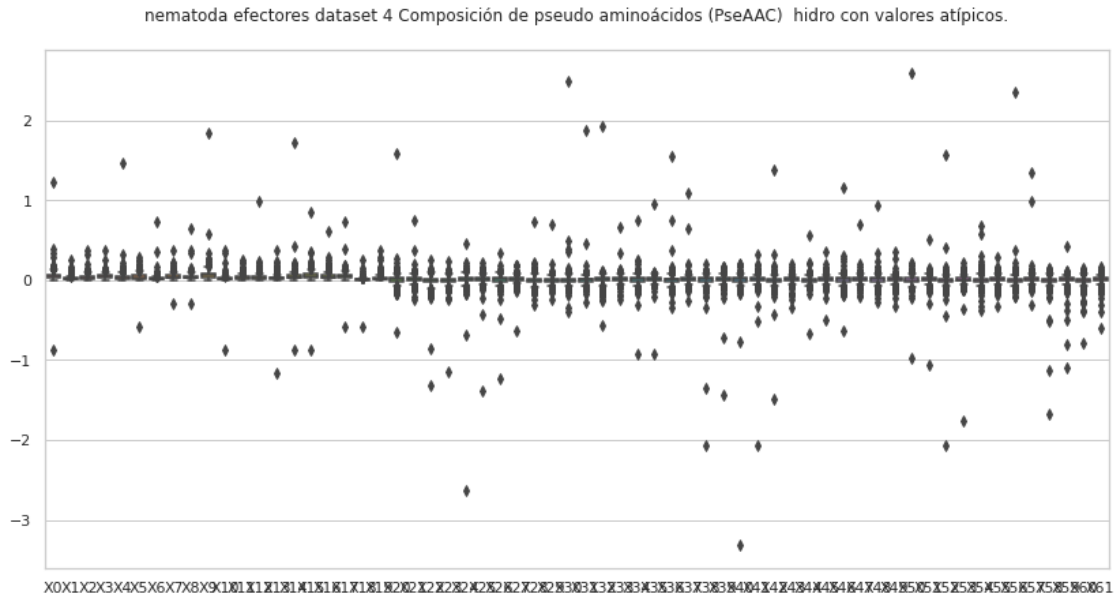
```
max      0.734866    0.367433    0.642214    1.837166  …    1.569460

              X53         X54         X55         X56         X57         X58  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.001904    0.004903    0.006951    0.010811    0.013783   -0.004119
std      0.088804    0.065232    0.042427    0.114598    0.082764    0.105033
min     -1.765688   -0.380087   -0.327894   -0.211280   -0.305383   -1.682176
25%     -0.008670   -0.015251   -0.009846   -0.012995   -0.007489   -0.014847
50%      0.005507    0.003964    0.010979    0.005284    0.008819    0.004125
75%      0.023450    0.020903    0.025703    0.022192    0.025863    0.022019
max      0.211848    0.676105    0.288155    2.352095    1.350001    0.149374

              X59         X60         X61
count  500.000000  500.000000  500.000000
mean     0.002169   -0.003490    0.003120
std      0.078844    0.061670    0.052805
min     -1.102970   -0.786721   -0.606452
25%     -0.008250   -0.017765   -0.009580
50%      0.010081    0.004731    0.008807
75%      0.024363    0.020426    0.023177
max      0.425519    0.142908    0.169832

[8 rows x 62 columns]



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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.030128  0.000000  0.006695  0.016738  0.003348  0.016738  0.013390
1    0.021814  0.010907  0.038174  0.065441  0.054534  0.027267  0.005453
2    0.034736  0.034736  0.028947  0.017368  0.034736  0.028947  0.005789
3    0.014553  0.024255  0.014553  0.014553  0.019404  0.019404  0.009702
4    0.036374  0.011768  0.016047  0.025676  0.022466  0.033164  0.034234
..        …         …         …         …         …         …         …
495  0.021555  0.004790  0.007185  0.009580  0.011975  0.016765  0.002395
496  0.000000  0.000000  0.015951  0.023927  0.010634  0.002659  0.005317
497  0.050067  0.018775  0.043809  0.093876  0.056325  0.031292  0.043809
498  0.014912  0.002982  0.040262  0.035789  0.014912  0.022368  0.004474
499  0.031611  0.006322  0.027818  0.042992  0.036669  0.026554  0.011380

           X7        X8        X9  …       X53       X54       X55  \
0    0.013390  0.066950  0.006695  …  0.010878 -0.017519  0.017228
1    0.021814  0.065441  0.065441  … -0.015336 -0.023880  0.010362
2    0.063683  0.069472  0.046315  … -0.011086 -0.030698 -0.071876
```

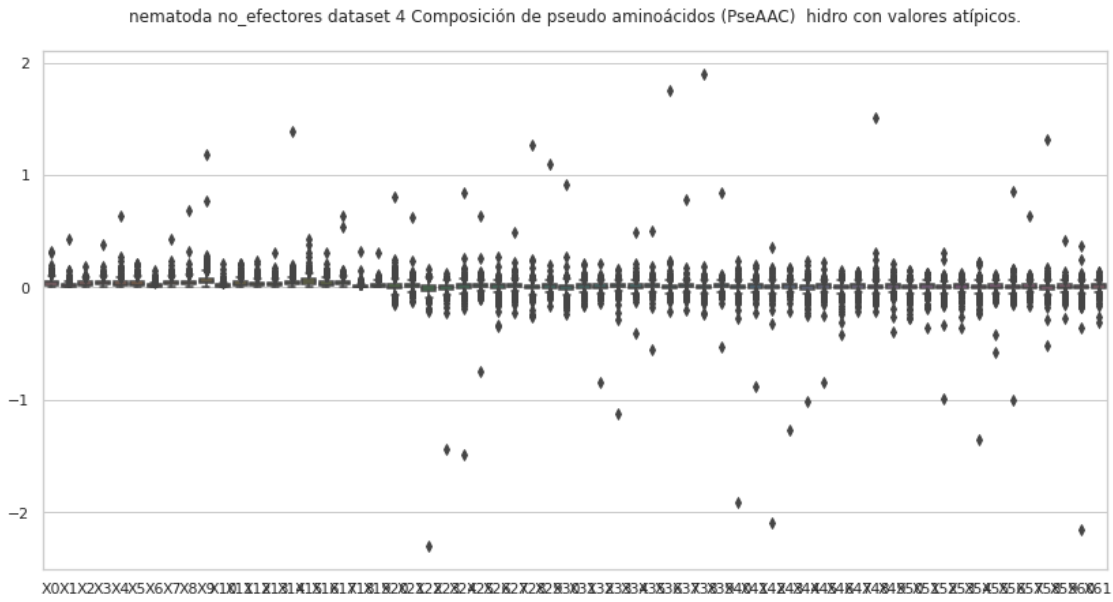
```
3    0.033957  0.063064  0.014553  … -0.004009 -0.006315 -0.009161
4    0.021396  0.014977  0.058840  … -0.003106  0.019982  0.006317
..        …         …         …  …         …         …
495  0.019160  0.011975  0.023949  …  0.005451  0.020811  0.012923
496  0.023927  0.031902  0.023927  …  0.020840  0.008895  0.019009
497  0.062584  0.056325  0.143943  … -0.064240 -0.043087  0.013377
498  0.032806  0.041753  0.050700  …  0.025512  0.022587  0.035678
499  0.060694  0.031611  0.048049  …  0.004805 -0.002342  0.006788

          X56       X57       X58       X59       X60       X61           X62
0    0.000065  0.025178  0.011266  0.022615  0.017649  0.028304  no_efectores
1   -0.006304  0.064515 -0.053551  0.005003 -0.064603  0.028079  no_efectores
2    0.008134  0.013572  0.013473 -0.010994 -0.021251 -0.028407  no_efectores
3    0.036880  0.029669  0.030050  0.001440 -0.029431 -0.018840  no_efectores
4    0.021733 -0.003907  0.006332 -0.007145  0.011246 -0.001063  no_efectores
..        …         …         …         …         …         …           …
495  0.042146  0.021378  0.034259  0.031713 -0.006078  0.000345  no_efectores
496  0.020635  0.042728 -0.004091  0.032092  0.033377  0.029960  no_efectores
497  0.034216  0.047690 -0.065973 -0.020824  0.044035  0.009385  no_efectores
498  0.011046  0.012327  0.008732  0.019683  0.006017  0.022145  no_efectores
499 -0.014068  0.011697 -0.010137 -0.009892 -0.002934  0.011915  no_efectores

[500 rows x 63 columns]


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

                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.043302    0.018165    0.035097    0.045124    0.038206    0.040928
std      0.031239    0.026059    0.020693    0.029398    0.039973    0.031208
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.022455    0.005704    0.019196    0.027209    0.018176    0.020102
50%      0.038262    0.011787    0.033946    0.041293    0.032250    0.033042
75%      0.055692    0.023160    0.047195    0.058603    0.047304    0.053210
max      0.320730    0.427640    0.187949    0.382488    0.641461    0.215768

                X6          X7          X8          X9  …         X52  \
count  500.000000  500.000000  500.000000  500.000000  …  500.000000
mean     0.017344    0.045097    0.046100    0.070123  …   -0.001357
std      0.016204    0.033449    0.040786    0.071468  …    0.060468
min      0.000000    0.000000    0.000000    0.005776  …   -0.991762
25%      0.007136    0.024729    0.025788    0.039505  …   -0.013012
50%      0.013383    0.040130    0.041769    0.058929  …    0.004434
75%      0.022891    0.057046    0.058827    0.082026  …    0.019255
max      0.152995    0.427640    0.688479    1.176011  …    0.310698
```

	X53	X54	X55	X56	X57	X58	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.006269	-0.001582	0.005248	-0.000045	0.009143	0.001269	
std	0.038316	0.073615	0.043059	0.069835	0.040761	0.074286	
min	-0.364456	-1.361055	-0.575039	-1.008468	-0.168795	-0.522830	
25%	-0.005456	-0.012932	-0.007119	-0.013097	-0.002979	-0.015806	
50%	0.010559	0.004980	0.010022	0.003474	0.009970	0.003683	
75%	0.024199	0.019263	0.024485	0.017771	0.024540	0.017258	
max	0.124052	0.227903	0.108588	0.850766	0.631166	1.315110	

	X59	X60	X61
count	500.000000	500.000000	500.000000
mean	0.006242	-0.001008	0.008496
std	0.038677	0.105986	0.036786
min	-0.272037	-2.152019	-0.305780
25%	-0.006270	-0.009756	-0.003701
50%	0.008390	0.006079	0.012914
75%	0.022996	0.020186	0.025734
max	0.415321	0.366882	0.143288

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

           X0        X1        X2        X3        X4        X5        X6  \
0    0.011136  0.003182  0.006364  0.030227  0.017500  0.004773  0.012727
1    0.092567  0.046284  0.037027  0.069425  0.032399  0.078682  0.046284
2    0.083019  0.025062  0.056391  0.062656  0.039160  0.031328  0.040727
3    0.019294  0.016078  0.015274  0.017686  0.028941  0.020098  0.010451
4    0.042507  0.007357  0.042507  0.058855  0.021253  0.026158  0.014714
..        …         …         …         …         …         …         …
495  0.062844  0.012163  0.042572  0.064871  0.052708  0.034463  0.030408
496  0.008660  0.004330  0.012990  0.023815  0.007578  0.009743  0.005413
497  0.074936  0.008816  0.035264  0.039672  0.030856  0.066120  0.039672
498  0.050780  0.032314  0.032314  0.032314  0.055396  0.069245  0.032314
499  0.036475  0.002605  0.049501  0.059922  0.044290  0.049501  0.015632

           X7        X8        X9  …       X53       X54       X55  \
0    0.030227  0.036591  0.108181  …  0.004934  0.001574 -0.004453
1    0.120337  0.050912  0.138851  … -0.006544  0.059292  0.032058
2    0.065789  0.072055  0.111215  …  0.008120 -0.011489 -0.003575
3    0.025725  0.011255  0.035372  …  0.004121 -0.002145  0.002388
4    0.043324  0.053133  0.061308  …  0.009366 -0.010307  0.003890
```

```
..        …       …       …  …       …       …       …
495  0.054735  0.050680  0.087170  …  0.005392 -0.016871 -0.013994
496  0.006495  0.024898  0.009743  …  0.029677  0.016853  0.022779
497  0.083752  0.070528  0.101385  …  0.042724  0.060737  0.020951
498  0.036931  0.023082  0.073861  … -0.025008 -0.021939 -0.047187
499  0.059922  0.067738  0.091186  …  0.019699  0.000811  0.011838

          X56       X57       X58       X59       X60       X61        X62
0   -0.007728 -0.002726  0.027645  0.007278  0.060397  0.019580  efectores
1    0.072419 -0.020110  0.038217 -0.013973 -0.010349  0.012512  efectores
2   -0.037078  0.013175  0.051683  0.027123 -0.006072 -0.009567  efectores
3    0.000654 -0.001526  0.015407  0.013209  0.008582  0.003276  efectores
4    0.003831  0.037347 -0.012406  0.024746  0.001602  0.016222  efectores
..        …       …       …       …       …       …       …
495  0.019092  0.022111 -0.022248  0.000041  0.005464 -0.006274  efectores
496  0.016817  0.035912 -0.003120  0.022106  0.023367  0.025132  efectores
497 -0.116819 -0.033172 -0.040427 -0.005280 -0.052607 -0.060567  efectores
498 -0.011160 -0.022992 -0.048199 -0.019221 -0.016034 -0.044705  efectores
499  0.012881  0.047794  0.024928  0.049372 -0.015972  0.023362  efectores

[448 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  efectores nematoda dataset 4, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  448.000000  448.000000  448.000000  448.000000  448.000000  448.000000
mean     0.050589    0.016677    0.036527    0.043286    0.030420    0.040121
std      0.028363    0.014644    0.020036    0.022142    0.018599    0.027101
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.029508    0.006514    0.019944    0.026418    0.016222    0.020960
50%      0.046373    0.013158    0.036090    0.040678    0.028919    0.033636
75%      0.066104    0.022650    0.049173    0.057991    0.043129    0.053825
max      0.189747    0.091742    0.105599    0.128543    0.111934    0.178688

                X6          X7          X8          X9  …         X52  \
count  448.000000  448.000000  448.000000  448.000000  …  448.000000
mean     0.018266    0.041372    0.040643    0.064120  …   -0.000874
std      0.014632    0.023039    0.022972    0.031978  …    0.034035
min      0.000000    0.000000    0.000000    0.000447  …   -0.165436
25%      0.008172    0.024362    0.023148    0.038747  …   -0.015443
50%      0.014175    0.038379    0.037495    0.061468  …    0.003181
75%      0.024714    0.056704    0.054151    0.085538  …    0.015706
max      0.084995    0.120337    0.141221    0.169307  …    0.118292

               X53         X54         X55         X56         X57         X58  \
```

count	448.000000	448.000000	448.000000	448.000000	448.000000	448.000000
mean	0.006874	0.004962	0.009733	0.005695	0.009262	0.005762
std	0.031626	0.035810	0.028600	0.037486	0.029053	0.035750
min	-0.123621	-0.133716	-0.092522	-0.205810	-0.097110	-0.110828
25%	-0.007185	-0.013159	-0.005620	-0.011162	-0.007149	-0.011677
50%	0.006554	0.004520	0.011655	0.005897	0.008582	0.005443
75%	0.023753	0.020490	0.025251	0.021618	0.024666	0.022650
max	0.153919	0.193082	0.133918	0.163835	0.130129	0.149374

	X59	X60	X61
count	448.000000	448.000000	448.000000
mean	0.009148	0.003878	0.008971
std	0.028833	0.034351	0.030454
min	-0.099980	-0.133356	-0.095197
25%	-0.004980	-0.014458	-0.007006
50%	0.010826	0.006172	0.010100
75%	0.024132	0.020603	0.023378
max	0.105638	0.137694	0.136295

[8 rows x 62 columns]

no_efectores
Composición de pseudo aminoácidos (PseAAC) no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.030128	0.000000	0.006695	0.016738	0.003348	0.016738	0.013390	
1	0.021814	0.010907	0.038174	0.065441	0.054534	0.027267	0.005453	
2	0.034736	0.034736	0.028947	0.017368	0.034736	0.028947	0.005789	
3	0.014553	0.024255	0.014553	0.014553	0.019404	0.019404	0.009702	
4	0.036374	0.011768	0.016047	0.025676	0.022466	0.033164	0.034234	
..	…	…	…	…	…	…	…	
495	0.021555	0.004790	0.007185	0.009580	0.011975	0.016765	0.002395	
496	0.000000	0.000000	0.015951	0.023927	0.010634	0.002659	0.005317	
497	0.050067	0.018775	0.043809	0.093876	0.056325	0.031292	0.043809	
498	0.014912	0.002982	0.040262	0.035789	0.014912	0.022368	0.004474	
499	0.031611	0.006322	0.027818	0.042992	0.036669	0.026554	0.011380	

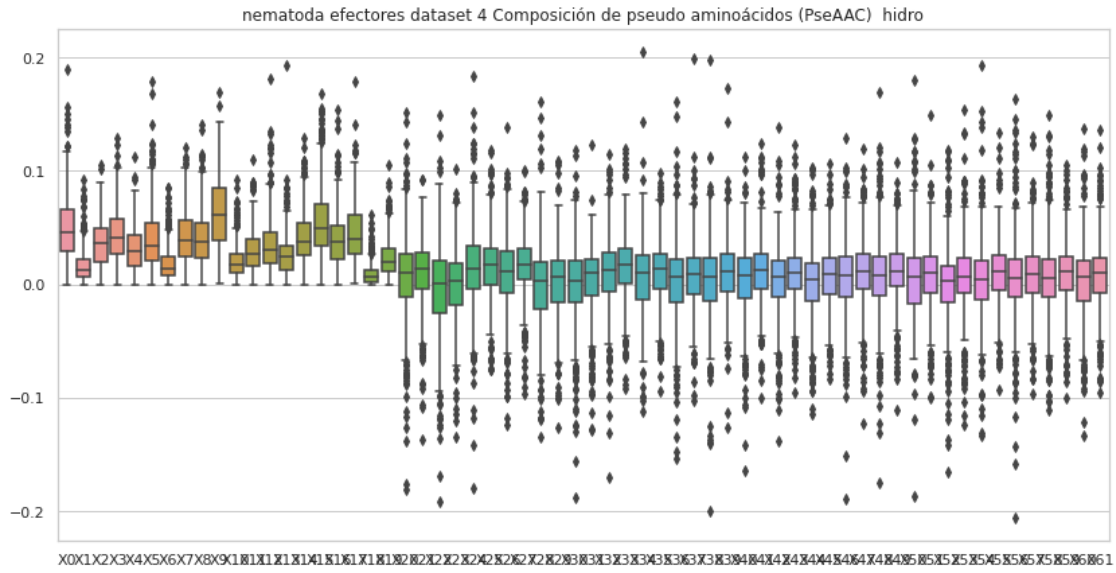
	X7	X8	X9	…	X53	X54	X55	\
0	0.013390	0.066950	0.006695	…	0.010878	-0.017519	0.017228	
1	0.021814	0.065441	0.065441	…	-0.015336	-0.023880	0.010362	
2	0.063683	0.069472	0.046315	…	-0.011086	-0.030698	-0.071876	
3	0.033957	0.063064	0.014553	…	-0.004009	-0.006315	-0.009161	
4	0.021396	0.014977	0.058840	…	-0.003106	0.019982	0.006317	
..	…	…	…	…	…	…	…	

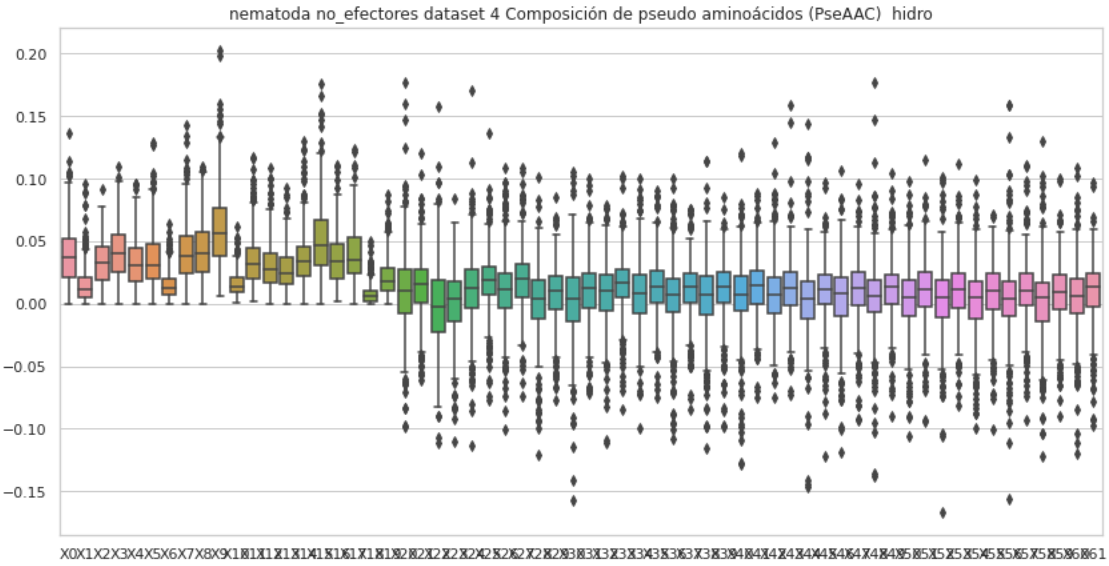
```
495  0.019160  0.011975  0.023949  …  0.005451  0.020811  0.012923
496  0.023927  0.031902  0.023927  …  0.020840  0.008895  0.019009
497  0.062584  0.056325  0.143943  … -0.064240 -0.043087  0.013377
498  0.032806  0.041753  0.050700  …  0.025512  0.022587  0.035678
499  0.060694  0.031611  0.048049  …  0.004805 -0.002342  0.006788

          X56       X57       X58       X59       X60       X61           X62
0    0.000065  0.025178  0.011266  0.022615  0.017649  0.028304  no_efectores
1   -0.006304  0.064515 -0.053551  0.005003 -0.064603  0.028079  no_efectores
2    0.008134  0.013572  0.013473 -0.010994 -0.021251 -0.028407  no_efectores
3    0.036880  0.029669  0.030050  0.001440 -0.029431 -0.018840  no_efectores
4    0.021733 -0.003907  0.006332 -0.007145  0.011246 -0.001063  no_efectores
..        …         …         …         …         …         …           …
495  0.042146  0.021378  0.034259  0.031713 -0.006078  0.000345  no_efectores
496  0.020635  0.042728 -0.004091  0.032092  0.033377  0.029960  no_efectores
497  0.034216  0.047690 -0.065973 -0.020824  0.044035  0.009385  no_efectores
498  0.011046  0.012327  0.008732  0.019683  0.006017  0.022145  no_efectores
499 -0.014068  0.011697 -0.010137 -0.009892 -0.002934  0.011915  no_efectores

[446 rows x 63 columns]


Composición de pseudo aminoácidos (PseAAC)  no_efectores nematoda dataset 4, sin
valores atípicos.
Estadísticas.

                X0          X1          X2          X3          X4          X5  \
count  446.000000  446.000000  446.000000  446.000000  446.000000  446.000000
mean     0.039301    0.015167    0.033221    0.041474    0.032467    0.035871
std      0.022160    0.014199    0.017787    0.020920    0.018334    0.022305
min      0.000000    0.000000    0.000000    0.000000    0.000000    0.000000
25%      0.021556    0.005406    0.018836    0.025360    0.017557    0.019597
50%      0.036854    0.011042    0.032711    0.039909    0.030412    0.031041
75%      0.052462    0.020953    0.046150    0.055791    0.044433    0.048357
max      0.136897    0.096157    0.091612    0.110288    0.096375    0.128677

                X6          X7          X8          X9  …         X52  \
count  446.000000  446.000000  446.000000  446.000000  …  446.000000
mean     0.015035    0.040692    0.041920    0.060089  …    0.002464
std      0.010699    0.022929    0.021416    0.030268  …    0.028654
min      0.000000    0.000000    0.000000    0.005776  …   -0.166745
25%      0.006927    0.023957    0.025677    0.038633  …   -0.010751
50%      0.012789    0.037989    0.040078    0.056319  …    0.004762
75%      0.020627    0.053864    0.057566    0.076450  …    0.018649
max      0.064055    0.142793    0.109411    0.202580  …    0.101628

               X53         X54         X55         X56         X57         X58  \
count  446.000000  446.000000  446.000000  446.000000  446.000000  446.000000
```

mean	0.010218	0.002237	0.008551	0.003609	0.009944	0.001630
std	0.024787	0.026768	0.023781	0.030758	0.025131	0.030040
min	-0.080956	-0.099141	-0.100030	-0.155861	-0.092661	-0.121847
25%	-0.003706	-0.012182	-0.004033	-0.010268	-0.001446	-0.013920
50%	0.011647	0.004898	0.010744	0.003962	0.010845	0.004806
75%	0.024544	0.017728	0.024370	0.018064	0.024304	0.017169
max	0.112113	0.098942	0.070949	0.159239	0.111078	0.130457

	X59	X60	X61
count	446.000000	446.000000	446.000000
mean	0.007951	0.005035	0.010548
std	0.024814	0.027722	0.024851
min	-0.089862	-0.119277	-0.097626
25%	-0.004576	-0.007861	-0.002032
50%	0.009508	0.006404	0.013338
75%	0.022935	0.019581	0.024531
max	0.100477	0.108664	0.097056

[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 4,
con valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023642  0.023329  0.013284  0.040666  0.026471  0.015236  0.051753
1    0.035649  0.018330  0.087357 -0.004611 -0.068975  0.010340 -0.008007
2   -0.005317 -0.012036  0.069958  0.005763 -0.003564 -0.022218  0.014115
3    0.014677  0.046700  0.066425 -0.071348  0.004474  0.005898 -0.053103
4    0.030113 -0.029540  0.010632 -0.016750  0.030376 -0.008255 -0.004074
..        …       …       …       …       …       …       …
495  0.066859  0.036295  0.070433  0.032700 -0.008911 -0.036207  0.035014
496  0.061628  0.062655  0.057156  0.057411 -0.018826 -0.071057  0.019859
497 -0.105467 -0.020914  0.061333 -0.087161 -0.002571 -0.075824  0.051977
498  0.061321  0.059642  0.018091 -0.015505  0.138756  0.067489 -0.067868
499  0.004734  0.030948 -0.009569  0.097735 -0.083865 -0.027612 -0.041137

           X7        X8        X9       X10       X11       X12        X13
0    0.024269  0.017839  0.055835  0.020615  0.013077  0.020485  efectores
1   -0.055377 -0.038105  0.093633  0.024200 -0.024381  0.053780  efectores
2    0.046834  0.019292  0.013929  0.007335  0.004366  0.016240  efectores
3    0.054807 -0.010043 -0.017734  0.028067  0.018582  0.011217  efectores
4    0.031913 -0.041663 -0.042714  0.012771 -0.029838 -0.009484  efectores
..        …       …       …       …       …       …        …
495  0.025654 -0.028501 -0.031832  0.032605 -0.017355 -0.013409  efectores
496 -0.044064 -0.011852  0.037272 -0.045361  0.026725  0.017098  efectores
497 -0.008349 -0.047218  0.020124  0.026053 -0.031018  0.068617  efectores
498 -0.008834  0.030187  0.017831 -0.079696 -0.066539 -0.059355  efectores
499 -0.053296 -0.044599 -0.007348  0.018089 -0.008402  0.146997  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro_mass efectores nematoda dataset 4,
con valores atípicos.
Estadísticas.

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

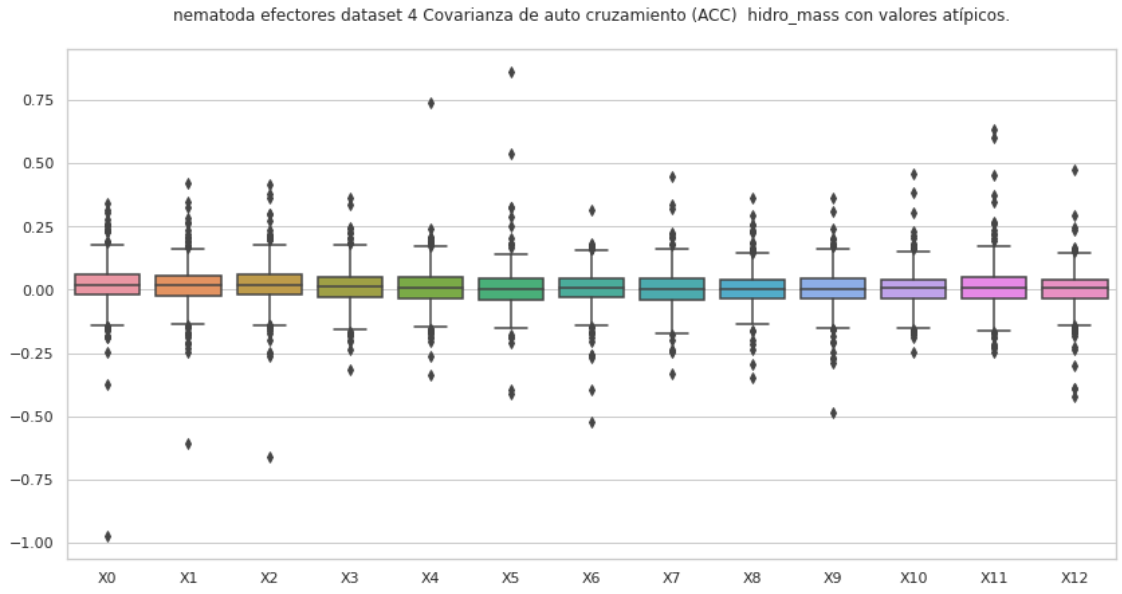
```
mean     0.017349    0.013427    0.018413    0.009457    0.006345    0.002949
std      0.089541    0.081756    0.082486    0.077526    0.077753    0.085851
min     -0.973303   -0.605199   -0.658721   -0.316062   -0.337254   -0.408999
25%     -0.020949   -0.024297   -0.020775   -0.032536   -0.033140   -0.039195
50%      0.019019    0.017385    0.017901    0.010622    0.008144    0.000198
75%      0.060676    0.052248    0.059027    0.051739    0.048115    0.041786
max      0.339290    0.421521    0.413073    0.364189    0.740553    0.859837

                X6          X7          X8          X9         X10         X11  \
count   500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean      0.002905    0.004037    0.003360    0.001236    0.005156    0.007733
std       0.074695    0.076328    0.072067    0.074440    0.071398    0.089071
min      -0.521300   -0.333285   -0.347344   -0.486290   -0.246654   -0.245671
25%      -0.030898   -0.041030   -0.034908   -0.036447   -0.036596   -0.037892
50%       0.004736    0.003459    0.001441    0.004373    0.005591    0.004435
75%       0.043148    0.045720    0.037478    0.041633    0.040600    0.047724
max       0.314018    0.445118    0.361326    0.363463    0.460151    0.633564

               X12
count   500.000000
mean      0.003112
std       0.077239
min      -0.421017
25%      -0.033314
50%       0.009097
75%       0.039468
max       0.472230



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

            X0        X1        X2        X3        X4        X5        X6  \
0   -0.111515  0.019754 -0.000176 -0.016985  0.073930 -0.134858 -0.000467
1    0.069525  0.075932 -0.004006  0.033160  0.031882  0.101721  0.046542
2   -0.137117 -0.050901 -0.035012  0.029662  0.093507 -0.032124 -0.069208
3    0.146028 -0.018566 -0.028978 -0.068683 -0.061945 -0.081317 -0.074043
4   -0.051331  0.124671 -0.004765  0.133567  0.032564  0.054623  0.010945
..        …         …         …         …         …         …         …
495  0.013776 -0.004682 -0.096022 -0.066234  0.009551  0.100483  0.116397
496  0.007620 -0.053047 -0.014487  0.022938  0.051835 -0.038313  0.018839
497  0.194208 -0.003555 -0.028271  0.059043  0.101451  0.096941  0.067167
498  0.102074  0.088756 -0.020687  0.018340 -0.018867 -0.012318 -0.033797
499  0.076414  0.031061 -0.061479 -0.002258 -0.077372 -0.006103  0.014075
```

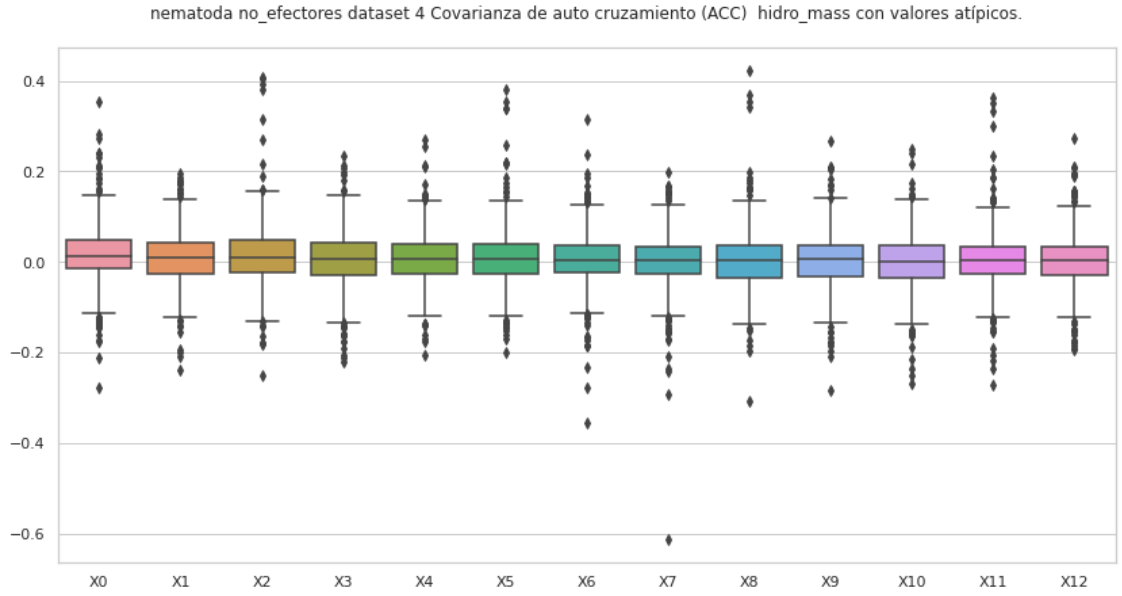
```
           X7        X8        X9       X10       X11       X12           X13
0   -0.127911 -0.012164  0.011154 -0.127371 -0.057354  0.194261  no_efectores
1   -0.052750 -0.064243 -0.108268  0.135297 -0.025711  0.103850  no_efectores
2   -0.013584 -0.116356  0.005888 -0.034093  0.044329  0.033424  no_efectores
3    0.000896  0.036371 -0.061117 -0.155283  0.030327 -0.044876  no_efectores
4    0.047644 -0.033995  0.060953  0.023873 -0.007808 -0.009675  no_efectores
..        …         …         …         …         …         …           …
495  0.048412  0.051016 -0.055788 -0.136253  0.086386 -0.138749  no_efectores
496  0.009826  0.007165 -0.072778  0.028932  0.077095  0.059111  no_efectores
497  0.040701  0.020770  0.001007 -0.024734 -0.025231  0.045784  no_efectores
498  0.011331  0.080197  0.052789  0.083814  0.001885  0.023073  no_efectores
499  0.060084  0.003329  0.019467 -0.017768 -0.091018  0.040127  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.016602    0.008684    0.013893    0.004777    0.006311    0.009818
std      0.066678    0.060686    0.070707    0.063284    0.058331    0.066473
min     -0.276882   -0.238490   -0.251597   -0.222571   -0.205070   -0.200748
25%     -0.015788   -0.025268   -0.022608   -0.028365   -0.026146   -0.026984
50%      0.012066    0.010699    0.010537    0.006967    0.007463    0.007507
75%      0.049438    0.042144    0.049227    0.041712    0.038157    0.039701
max      0.353960    0.195242    0.407087    0.234259    0.268951    0.382057

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.002998    0.000415    0.006428    0.002988   -0.001705    0.003887
std      0.063353    0.066410    0.069301    0.063624    0.064744    0.067038
min     -0.356904   -0.613455   -0.308092   -0.285042   -0.267716   -0.272173
25%     -0.024645   -0.027761   -0.034028   -0.032740   -0.034260   -0.027597
50%      0.003924    0.003105    0.003218    0.005610    0.001471    0.002764
75%      0.035458    0.033877    0.036466    0.036932    0.036172    0.034681
max      0.314908    0.197628    0.422461    0.265879    0.248764    0.364048

               X12
count  500.000000
mean     0.002885
std      0.063054
min     -0.194420
25%     -0.028722
50%      0.002244
75%      0.034848
```

```
max       0.272793
```


nematoda efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro_mass con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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 4,
sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.023642	0.023329	0.013284	0.040666	0.026471	0.015236	0.051753	
1	0.035649	0.018330	0.087357	-0.004611	-0.068975	0.010340	-0.008007	
2	-0.005317	-0.012036	0.069958	0.005763	-0.003564	-0.022218	0.014115	
3	0.014677	0.046700	0.066425	-0.071348	0.004474	0.005898	-0.053103	
4	0.030113	-0.029540	0.010632	-0.016750	0.030376	-0.008255	-0.004074	
..	…	…	…	…	…	…	…	
495	0.066859	0.036295	0.070433	0.032700	-0.008911	-0.036207	0.035014	
496	0.061628	0.062655	0.057156	0.057411	-0.018826	-0.071057	0.019859	
497	-0.105467	-0.020914	0.061333	-0.087161	-0.002571	-0.075824	0.051977	
498	0.061321	0.059642	0.018091	-0.015505	0.138756	0.067489	-0.067868	
499	0.004734	0.030948	-0.009569	0.097735	-0.083865	-0.027612	-0.041137	

	X7	X8	X9	X10	X11	X12	X13
0	0.024269	0.017839	0.055835	0.020615	0.013077	0.020485	efectores
1	-0.055377	-0.038105	0.093633	0.024200	-0.024381	0.053780	efectores
2	0.046834	0.019292	0.013929	0.007335	0.004366	0.016240	efectores
3	0.054807	-0.010043	-0.017734	0.028067	0.018582	0.011217	efectores
4	0.031913	-0.041663	-0.042714	0.012771	-0.029838	-0.009484	efectores
..	…	…	…	…	…	…	…
495	0.025654	-0.028501	-0.031832	0.032605	-0.017355	-0.013409	efectores
496	-0.044064	-0.011852	0.037272	-0.045361	0.026725	0.017098	efectores
497	-0.008349	-0.047218	0.020124	0.026053	-0.031018	0.068617	efectores
498	-0.008834	0.030187	0.017831	-0.079696	-0.066539	-0.059355	efectores
499	-0.053296	-0.044599	-0.007348	0.018089	-0.008402	0.146997	efectores

[461 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) hidro_mass efectores nematoda dataset 4,
sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	461.000000	461.000000	461.000000	461.000000	461.000000	461.000000	
mean	0.016518	0.011257	0.015744	0.006344	0.005832	0.000005	
std	0.068613	0.065570	0.062639	0.066197	0.061120	0.059539	
min	-0.248487	-0.208644	-0.201822	-0.202503	-0.175019	-0.209886	
25%	-0.019246	-0.023473	-0.019927	-0.032462	-0.029431	-0.037079	
50%	0.017780	0.015349	0.016326	0.008621	0.008511	-0.000377	
75%	0.059516	0.049538	0.054294	0.047868	0.047144	0.037235	
max	0.257680	0.232387	0.235181	0.225879	0.210755	0.205563	

```
                X6           X7           X8           X9          X10          X11  \
count   461.000000   461.000000   461.000000   461.000000   461.000000   461.000000
mean      0.005992     0.003498     0.001681     0.003088     0.003031     0.000791
std       0.062600     0.063628     0.057375     0.061663     0.059065     0.068012
min      -0.203901    -0.201537    -0.197325    -0.209759    -0.179640    -0.245671
25%      -0.029565    -0.035717    -0.033273    -0.034969    -0.036267    -0.037882
50%       0.004934     0.003462     0.000731     0.004650     0.005588     0.003163
75%       0.043401     0.041309     0.033744     0.038995     0.037767     0.044271
max       0.181643     0.220791     0.182610     0.202020     0.205775     0.233232

               X12
count   461.000000
mean      0.005435
std       0.060142
min      -0.183315
25%      -0.030389
50%       0.009766
75%       0.039352
max       0.166678
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.069525  0.075932 -0.004006  0.033160  0.031882  0.101721  0.046542
2   -0.137117 -0.050901 -0.035012  0.029662  0.093507 -0.032124 -0.069208
3    0.146028 -0.018566 -0.028978 -0.068683 -0.061945 -0.081317 -0.074043
4   -0.051331  0.124671 -0.004765  0.133567  0.032564  0.054623  0.010945
5    0.001314  0.049025  0.021784  0.055438  0.058698 -0.044424  0.049523
..        …         …         …         …         …         …         …
495  0.013776 -0.004682 -0.096022 -0.066234  0.009551  0.100483  0.116397
496  0.007620 -0.053047 -0.014487  0.022938  0.051835 -0.038313  0.018839
497  0.194208 -0.003555 -0.028271  0.059043  0.101451  0.096941  0.067167
498  0.102074  0.088756 -0.020687  0.018340 -0.018867 -0.012318 -0.033797
499  0.076414  0.031061 -0.061479 -0.002258 -0.077372 -0.006103  0.014075

           X7        X8        X9       X10       X11       X12           X13
1   -0.052750 -0.064243 -0.108268  0.135297 -0.025711  0.103850  no_efectores
2   -0.013584 -0.116356  0.005888 -0.034093  0.044329  0.033424  no_efectores
3    0.000896  0.036371 -0.061117 -0.155283  0.030327 -0.044876  no_efectores
4    0.047644 -0.033995  0.060953  0.023873 -0.007808 -0.009675  no_efectores
5    0.013137 -0.011459 -0.003372  0.048529 -0.006154  0.015843  no_efectores
..        …         …         …         …         …         …           …
495  0.048412  0.051016 -0.055788 -0.136253  0.086386 -0.138749  no_efectores
```

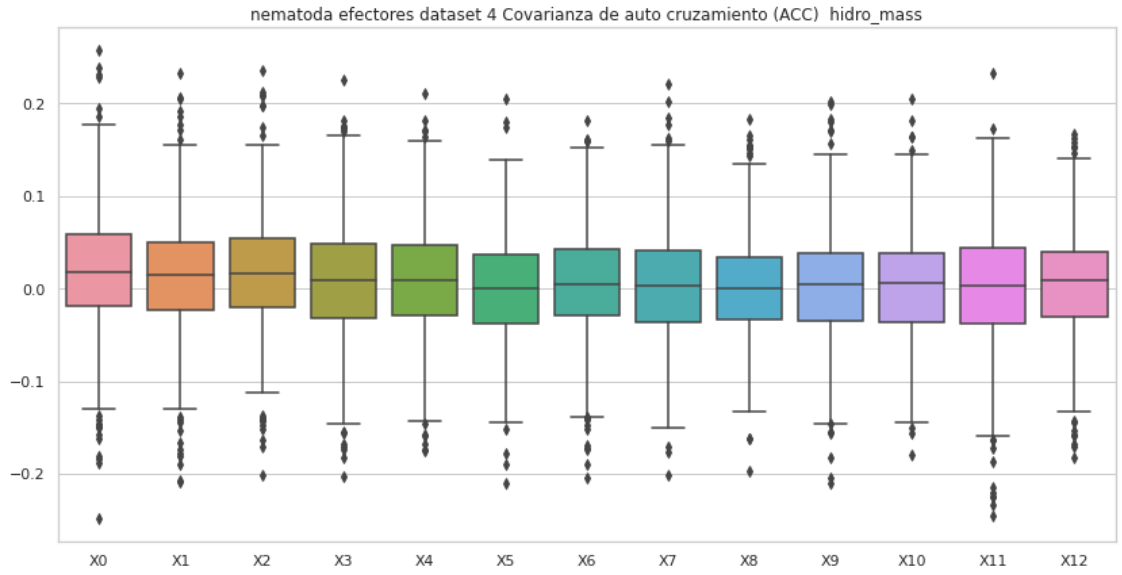
```
496  0.009826  0.007165 -0.072778  0.028932  0.077095  0.059111  no_efectores
497  0.040701  0.020770  0.001007 -0.024734 -0.025231  0.045784  no_efectores
498  0.011331  0.080197  0.052789  0.083814  0.001885  0.023073  no_efectores
499  0.060084  0.003329  0.019467 -0.017768 -0.091018  0.040127  no_efectores

[447 rows x 14 columns]
```

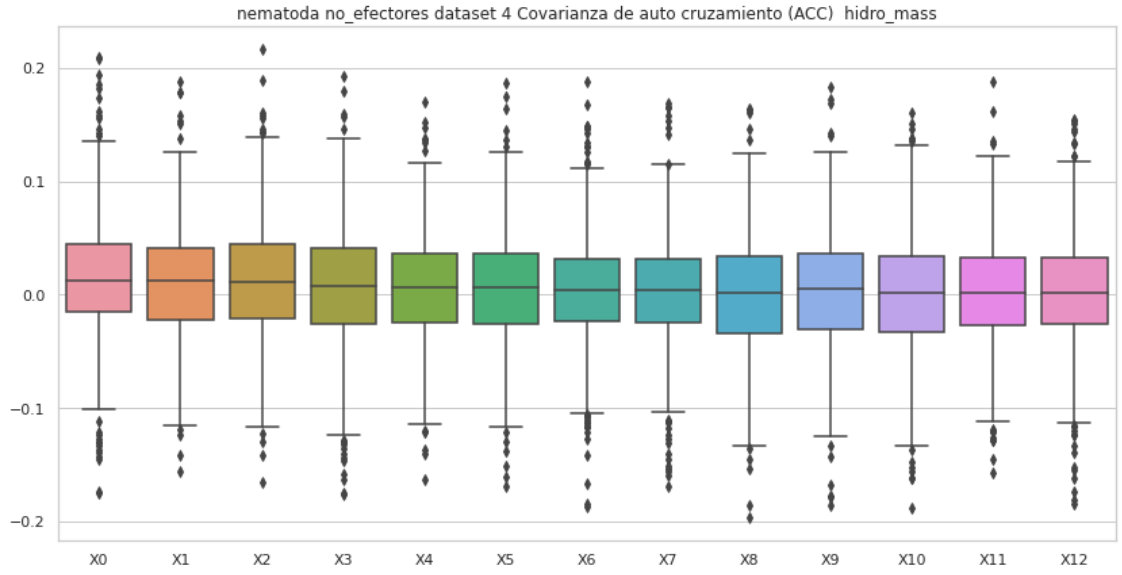
Covarianza de auto cruzamiento (ACC) hidro_mass no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	447.000000	447.000000	447.000000	447.000000	447.000000	447.000000	
mean	0.014922	0.009989	0.012099	0.006456	0.005208	0.006140	
std	0.057260	0.052982	0.052852	0.056465	0.049778	0.051949	
min	-0.175168	-0.155457	-0.165162	-0.175595	-0.162385	-0.169509	
25%	-0.015023	-0.021961	-0.020571	-0.025759	-0.023993	-0.025679	
50%	0.011883	0.012014	0.010938	0.007254	0.006285	0.006226	
75%	0.045156	0.041457	0.044360	0.041249	0.036190	0.036599	
max	0.209451	0.187548	0.216309	0.192805	0.170226	0.186795	

	X6	X7	X8	X9	X10	X11	\
count	447.000000	447.000000	447.000000	447.000000	447.000000	447.000000	
mean	0.003680	0.002656	0.001538	0.003057	-0.001521	0.001544	
std	0.051639	0.052021	0.053576	0.055152	0.055905	0.050639	
min	-0.186582	-0.168982	-0.196360	-0.185611	-0.188340	-0.156627	
25%	-0.023131	-0.024441	-0.033627	-0.030460	-0.033268	-0.026575	
50%	0.003631	0.003979	0.002183	0.005474	0.001421	0.001577	
75%	0.031616	0.031444	0.033541	0.035975	0.033609	0.032819	
max	0.187394	0.168781	0.163935	0.182632	0.160687	0.187432	

	X12
count	447.000000
mean	0.001034
std	0.054307
min	-0.184941
25%	-0.025650
50%	0.002013
75%	0.032203
max	0.154903





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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023642  0.023329  0.013284  0.040666  0.026471  0.015236  0.051753
1    0.035649  0.018330  0.087357 -0.004611 -0.068975  0.010340 -0.008007
2   -0.005317 -0.012036  0.069958  0.005763 -0.003564 -0.022218  0.014115
3    0.014677  0.046700  0.066425 -0.071348  0.004474  0.005898 -0.053103
4    0.030113 -0.029540  0.010632 -0.016750  0.030376 -0.008255 -0.004074
..        …         …         …         …         …         …         …
495  0.066859  0.036295  0.070433  0.032700 -0.008911 -0.036207  0.035014
496  0.061628  0.062655  0.057156  0.057411 -0.018826 -0.071057  0.019859
497 -0.105467 -0.020914  0.061333 -0.087161 -0.002571 -0.075824  0.051977
498  0.061321  0.059642  0.018091 -0.015505  0.138756  0.067489 -0.067868
499  0.004734  0.030948 -0.009569  0.097735 -0.083865 -0.027612 -0.041137

           X7        X8        X9       X10       X11       X12        X13
0    0.024269  0.017839  0.055835  0.020615  0.013077  0.020485  efectores
1   -0.055377 -0.038105  0.093633  0.024200 -0.024381  0.053780  efectores
```

```
2    0.046834  0.019292  0.013929  0.007335  0.004366  0.016240  efectores
3    0.054807 -0.010043 -0.017734  0.028067  0.018582  0.011217  efectores
4    0.031913 -0.041663 -0.042714  0.012771 -0.029838 -0.009484  efectores
..        …        …        …        …         …         …
495  0.025654 -0.028501 -0.031832  0.032605 -0.017355 -0.013409  efectores
496 -0.044064 -0.011852  0.037272 -0.045361  0.026725  0.017098  efectores
497 -0.008349 -0.047218  0.020124  0.026053 -0.031018  0.068617  efectores
498 -0.008834  0.030187  0.017831 -0.079696 -0.066539 -0.059355  efectores
499 -0.053296 -0.044599 -0.007348  0.018089 -0.008402  0.146997  efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.017349	0.013427	0.018413	0.009457	0.006345	0.002949	
std	0.089541	0.081756	0.082486	0.077526	0.077753	0.085851	
min	-0.973303	-0.605199	-0.658721	-0.316062	-0.337254	-0.408999	
25%	-0.020949	-0.024297	-0.020775	-0.032536	-0.033140	-0.039195	
50%	0.019019	0.017385	0.017901	0.010622	0.008144	0.000198	
75%	0.060676	0.052248	0.059027	0.051739	0.048115	0.041786	
max	0.339290	0.421521	0.413073	0.364189	0.740553	0.859837	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.002905	0.004037	0.003360	0.001236	0.005156	0.007733	
std	0.074695	0.076328	0.072067	0.074440	0.071398	0.089071	
min	-0.521300	-0.333285	-0.347344	-0.486290	-0.246654	-0.245671	
25%	-0.030898	-0.041030	-0.034908	-0.036447	-0.036596	-0.037892	
50%	0.004736	0.003459	0.001441	0.004373	0.005591	0.004435	
75%	0.043148	0.045720	0.037478	0.041633	0.040600	0.047724	
max	0.314018	0.445118	0.361326	0.363463	0.460151	0.633564	

	X12
count	500.000000
mean	0.003112
std	0.077239
min	-0.421017
25%	-0.033314
50%	0.009097
75%	0.039468
max	0.472230

no_efectores
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	-0.111515	0.019754	-0.000176	-0.016985	0.073930	-0.134858	-0.000467	
1	0.069525	0.075932	-0.004006	0.033160	0.031882	0.101721	0.046542	
2	-0.137117	-0.050901	-0.035012	0.029662	0.093507	-0.032124	-0.069208	
3	0.146028	-0.018566	-0.028978	-0.068683	-0.061945	-0.081317	-0.074043	
4	-0.051331	0.124671	-0.004765	0.133567	0.032564	0.054623	0.010945	
..	…	…	…	…	…	…	…	
495	0.013776	-0.004682	-0.096022	-0.066234	0.009551	0.100483	0.116397	
496	0.007620	-0.053047	-0.014487	0.022938	0.051835	-0.038313	0.018839	
497	0.194208	-0.003555	-0.028271	0.059043	0.101451	0.096941	0.067167	
498	0.102074	0.088756	-0.020687	0.018340	-0.018867	-0.012318	-0.033797	
499	0.076414	0.031061	-0.061479	-0.002258	-0.077372	-0.006103	0.014075	

	X7	X8	X9	X10	X11	X12	X13
0	-0.127911	-0.012164	0.011154	-0.127371	-0.057354	0.194261	no_efectores
1	-0.052750	-0.064243	-0.108268	0.135297	-0.025711	0.103850	no_efectores
2	-0.013584	-0.116356	0.005888	-0.034093	0.044329	0.033424	no_efectores
3	0.000896	0.036371	-0.061117	-0.155283	0.030327	-0.044876	no_efectores
4	0.047644	-0.033995	0.060953	0.023873	-0.007808	-0.009675	no_efectores
..	…	…	…	…	…	…	…
495	0.048412	0.051016	-0.055788	-0.136253	0.086386	-0.138749	no_efectores
496	0.009826	0.007165	-0.072778	0.028932	0.077095	0.059111	no_efectores
497	0.040701	0.020770	0.001007	-0.024734	-0.025231	0.045784	no_efectores
498	0.011331	0.080197	0.052789	0.083814	0.001885	0.023073	no_efectores
499	0.060084	0.003329	0.019467	-0.017768	-0.091018	0.040127	no_efectores

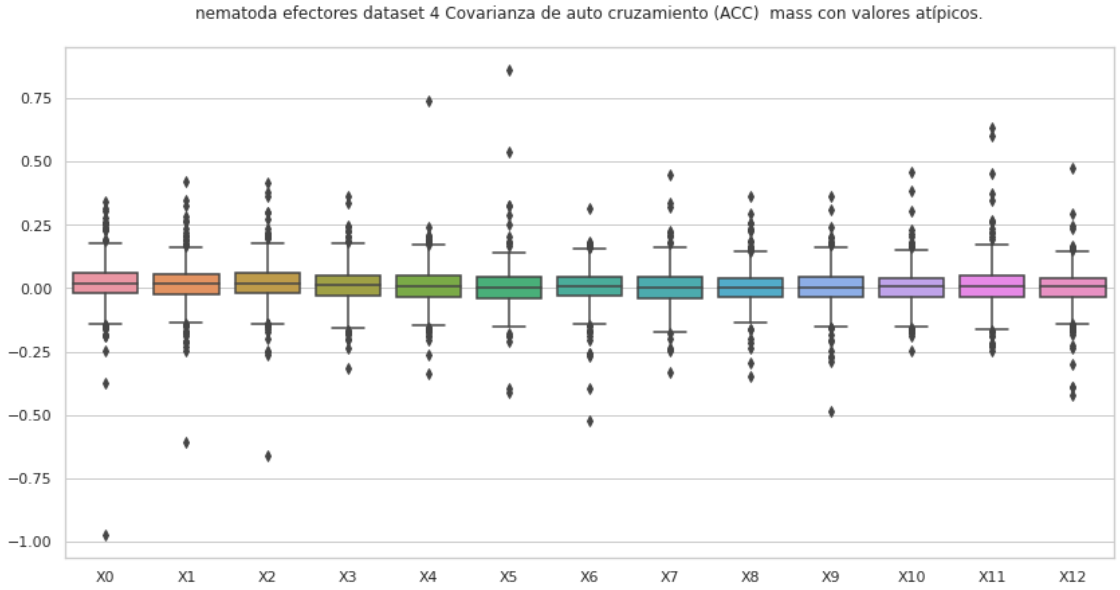
[500 rows x 14 columns]

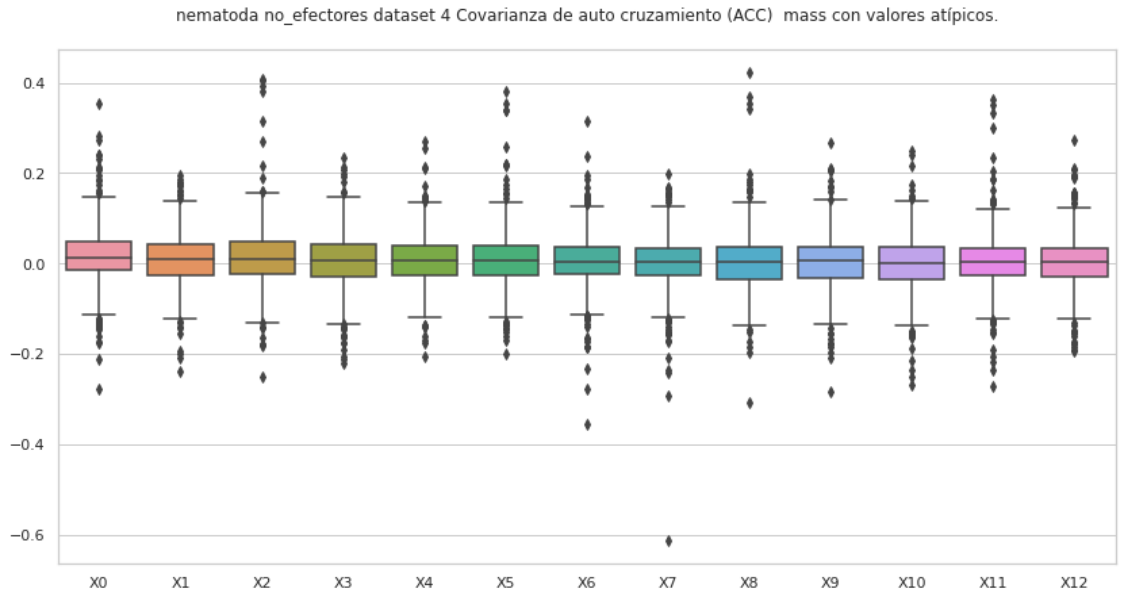
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.016602	0.008684	0.013893	0.004777	0.006311	0.009818	
std	0.066678	0.060686	0.070707	0.063284	0.058331	0.066473	
min	-0.276882	-0.238490	-0.251597	-0.222571	-0.205070	-0.200748	
25%	-0.015788	-0.025268	-0.022608	-0.028365	-0.026146	-0.026984	
50%	0.012066	0.010699	0.010537	0.006967	0.007463	0.007507	
75%	0.049438	0.042144	0.049227	0.041712	0.038157	0.039701	
max	0.353960	0.195242	0.407087	0.234259	0.268951	0.382057	

	X6	X7	X8	X9	X10	X11	\
count	500.000000	500.000000	500.000000	500.000000	500.000000	500.000000	
mean	0.002998	0.000415	0.006428	0.002988	-0.001705	0.003887	
std	0.063353	0.066410	0.069301	0.063624	0.064744	0.067038	
min	-0.356904	-0.613455	-0.308092	-0.285042	-0.267716	-0.272173	
25%	-0.024645	-0.027761	-0.034028	-0.032740	-0.034260	-0.027597	
50%	0.003924	0.003105	0.003218	0.005610	0.001471	0.002764	
75%	0.035458	0.033877	0.036466	0.036932	0.036172	0.034681	
max	0.314908	0.197628	0.422461	0.265879	0.248764	0.364048	

	X12
count	500.000000
mean	0.002885
std	0.063054
min	-0.194420
25%	-0.028722
50%	0.002244
75%	0.034848
max	0.272793





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 4, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.023642  0.023329  0.013284  0.040666  0.026471  0.015236  0.051753
1    0.035649  0.018330  0.087357 -0.004611 -0.068975  0.010340 -0.008007
2   -0.005317 -0.012036  0.069958  0.005763 -0.003564 -0.022218  0.014115
3    0.014677  0.046700  0.066425 -0.071348  0.004474  0.005898 -0.053103
4    0.030113 -0.029540  0.010632 -0.016750  0.030376 -0.008255 -0.004074
..        …         …         …         …         …         …         …
495  0.066859  0.036295  0.070433  0.032700 -0.008911 -0.036207  0.035014
496  0.061628  0.062655  0.057156  0.057411 -0.018826 -0.071057  0.019859
497 -0.105467 -0.020914  0.061333 -0.087161 -0.002571 -0.075824  0.051977
498  0.061321  0.059642  0.018091 -0.015505  0.138756  0.067489 -0.067868
499  0.004734  0.030948 -0.009569  0.097735 -0.083865 -0.027612 -0.041137

           X7        X8        X9       X10       X11       X12        X13
0    0.024269  0.017839  0.055835  0.020615  0.013077  0.020485  efectores
1   -0.055377 -0.038105  0.093633  0.024200 -0.024381  0.053780  efectores
2    0.046834  0.019292  0.013929  0.007335  0.004366  0.016240  efectores
3    0.054807 -0.010043 -0.017734  0.028067  0.018582  0.011217  efectores
4    0.031913 -0.041663 -0.042714  0.012771 -0.029838 -0.009484  efectores
..        …         …         …         …         …         …         …
```

```
495  0.025654 -0.028501 -0.031832  0.032605 -0.017355 -0.013409  efectores
496 -0.044064 -0.011852  0.037272 -0.045361  0.026725  0.017098  efectores
497 -0.008349 -0.047218  0.020124  0.026053 -0.031018  0.068617  efectores
498 -0.008834  0.030187  0.017831 -0.079696 -0.066539 -0.059355  efectores
499 -0.053296 -0.044599 -0.007348  0.018089 -0.008402  0.146997  efectores

[461 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) mass efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	461.000000	461.000000	461.000000	461.000000	461.000000	461.000000	
mean	0.016518	0.011257	0.015744	0.006344	0.005832	0.000005	
std	0.068613	0.065570	0.062639	0.066197	0.061120	0.059539	
min	-0.248487	-0.208644	-0.201822	-0.202503	-0.175019	-0.209886	
25%	-0.019246	-0.023473	-0.019927	-0.032462	-0.029431	-0.037079	
50%	0.017780	0.015349	0.016326	0.008621	0.008511	-0.000377	
75%	0.059516	0.049538	0.054294	0.047868	0.047144	0.037235	
max	0.257680	0.232387	0.235181	0.225879	0.210755	0.205563	

	X6	X7	X8	X9	X10	X11	\
count	461.000000	461.000000	461.000000	461.000000	461.000000	461.000000	
mean	0.005992	0.003498	0.001681	0.003088	0.003031	0.000791	
std	0.062600	0.063628	0.057375	0.061663	0.059065	0.068012	
min	-0.203901	-0.201537	-0.197325	-0.209759	-0.179640	-0.245671	
25%	-0.029565	-0.035717	-0.033273	-0.034969	-0.036267	-0.037882	
50%	0.004934	0.003462	0.000731	0.004650	0.005588	0.003163	
75%	0.043401	0.041309	0.033744	0.038995	0.037767	0.044271	
max	0.181643	0.220791	0.182610	0.202020	0.205775	0.233232	

	X12
count	461.000000
mean	0.005435
std	0.060142
min	-0.183315
25%	-0.030389
50%	0.009766
75%	0.039352
max	0.166678

Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Valores del documento csv.

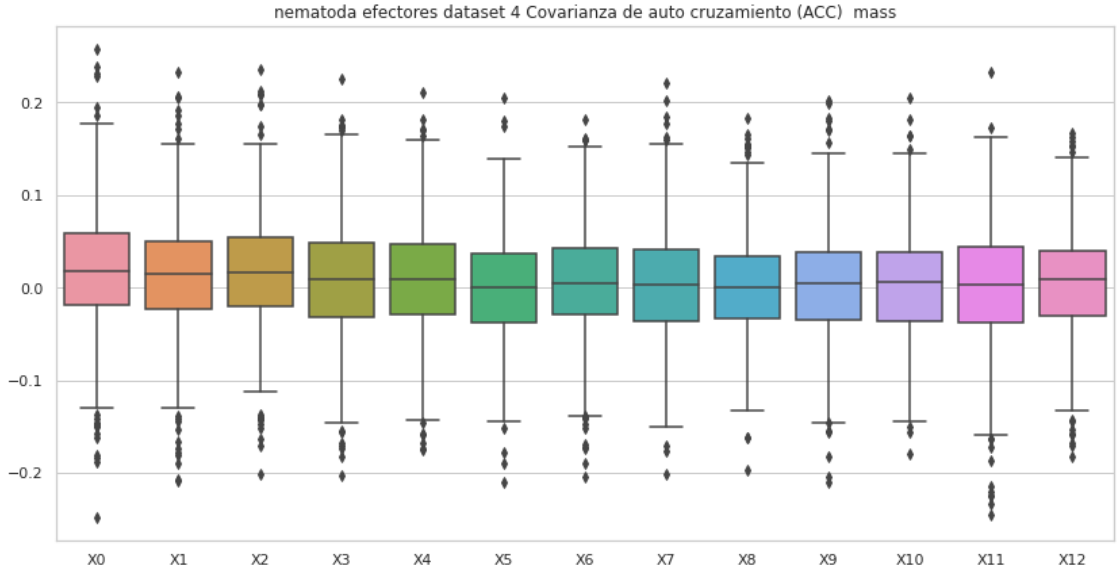
```
           X0        X1        X2        X3        X4        X5        X6  \
1    0.069525  0.075932 -0.004006  0.033160  0.031882  0.101721  0.046542
2   -0.137117 -0.050901 -0.035012  0.029662  0.093507 -0.032124 -0.069208
3    0.146028 -0.018566 -0.028978 -0.068683 -0.061945 -0.081317 -0.074043
4   -0.051331  0.124671 -0.004765  0.133567  0.032564  0.054623  0.010945
5    0.001314  0.049025  0.021784  0.055438  0.058698 -0.044424  0.049523
..        …         …         …         …         …         …         …
495  0.013776 -0.004682 -0.096022 -0.066234  0.009551  0.100483  0.116397
496  0.007620 -0.053047 -0.014487  0.022938  0.051835 -0.038313  0.018839
497  0.194208 -0.003555 -0.028271  0.059043  0.101451  0.096941  0.067167
498  0.102074  0.088756 -0.020687  0.018340 -0.018867 -0.012318 -0.033797
499  0.076414  0.031061 -0.061479 -0.002258 -0.077372 -0.006103  0.014075

           X7        X8        X9       X10       X11       X12           X13
1   -0.052750 -0.064243 -0.108268  0.135297 -0.025711  0.103850  no_efectores
2   -0.013584 -0.116356  0.005888 -0.034093  0.044329  0.033424  no_efectores
3    0.000896  0.036371 -0.061117 -0.155283  0.030327 -0.044876  no_efectores
4    0.047644 -0.033995  0.060953  0.023873 -0.007808 -0.009675  no_efectores
5    0.013137 -0.011459 -0.003372  0.048529 -0.006154  0.015843  no_efectores
..        …         …         …         …         …         …           …
495  0.048412  0.051016 -0.055788 -0.136253  0.086386 -0.138749  no_efectores
496  0.009826  0.007165 -0.072778  0.028932  0.077095  0.059111  no_efectores
497  0.040701  0.020770  0.001007 -0.024734 -0.025231  0.045784  no_efectores
498  0.011331  0.080197  0.052789  0.083814  0.001885  0.023073  no_efectores
499  0.060084  0.003329  0.019467 -0.017768 -0.091018  0.040127  no_efectores

[447 rows x 14 columns]
```

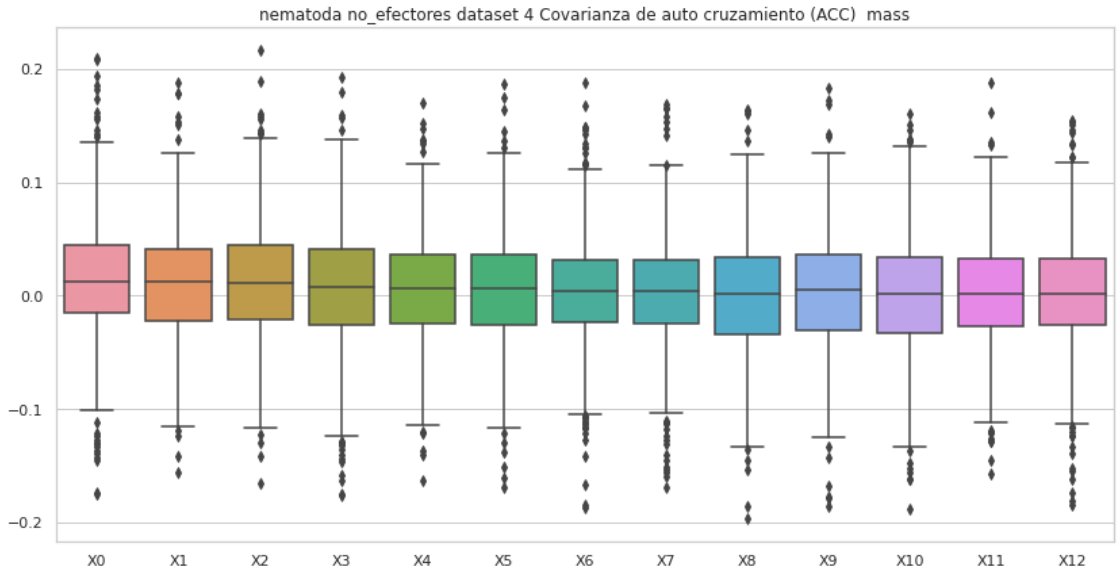
Covarianza de auto cruzamiento (ACC) mass no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  447.000000  447.000000  447.000000  447.000000  447.000000  447.000000
mean     0.014922    0.009989    0.012099    0.006456    0.005208    0.006140
std      0.057260    0.052982    0.052852    0.056465    0.049778    0.051949
min     -0.175168   -0.155457   -0.165162   -0.175595   -0.162385   -0.169509
25%     -0.015023   -0.021961   -0.020571   -0.025759   -0.023993   -0.025679
50%      0.011883    0.012014    0.010938    0.007254    0.006285    0.006226
75%      0.045156    0.041457    0.044360    0.041249    0.036190    0.036599
max      0.209451    0.187548    0.216309    0.192805    0.170226    0.186795

                X6          X7          X8          X9         X10         X11  \
count  447.000000  447.000000  447.000000  447.000000  447.000000  447.000000
mean     0.003680    0.002656    0.001538    0.003057   -0.001521    0.001544
std      0.051639    0.052021    0.053576    0.055152    0.055905    0.050639
```

min	-0.186582	-0.168982	-0.196360	-0.185611	-0.188340	-0.156627
25%	-0.023131	-0.024441	-0.033627	-0.030460	-0.033268	-0.026575
50%	0.003631	0.003979	0.002183	0.005474	0.001421	0.001577
75%	0.031616	0.031444	0.033541	0.035975	0.033609	0.032819
max	0.187394	0.168781	0.163935	0.182632	0.160687	0.187432

	X12
count	447.000000
mean	0.001034
std	0.054307
min	-0.184941
25%	-0.025650
50%	0.002013
75%	0.032203
max	0.154903





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 4, con
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.028711 -0.015801  0.072228  0.040547 -0.052455 -0.007357  0.020078
1   -0.036599 -0.026153 -0.010764  0.001333  0.003238 -0.062025  0.037190
2   -0.029981 -0.135378  0.036558  0.006781 -0.027149  0.026754  0.041512
3    0.080662  0.073778  0.090974  0.039758  0.066878  0.073850  0.045422
4   -0.027765 -0.067295  0.054629 -0.002677 -0.005438  0.000400  0.019138
..        …       …         …         …         …         …         …
495 -0.030087 -0.045489  0.033518  0.018402 -0.023900 -0.004714 -0.014491
496 -0.137377 -0.023717  0.057842  0.172777  0.047206  0.013638  0.085346
497 -0.013672 -0.087455 -0.050489 -0.041589 -0.097680  0.046167 -0.026502
498  0.116427  0.012649  0.115212 -0.008380  0.126774  0.156099 -0.052903
499 -0.055984 -0.167736  0.073530 -0.114125  0.018041  0.084930  0.006425

           X7        X8        X9       X10       X11       X12        X13
0   -0.074102 -0.035452  0.081557  0.054416  0.027528  0.034677  efectores
1    0.027850 -0.025965  0.086861 -0.063821 -0.087188 -0.059452  efectores
2    0.047561 -0.050205  0.027008  0.020462 -0.080180 -0.039155  efectores
3    0.143714  0.082443  0.063353  0.124786  0.052447  0.083028  efectores
4   -0.075391  0.027096  0.026459 -0.005397  0.093696 -0.023226  efectores
..        …       …         …         …         …         …         …
495  0.101363  0.092605 -0.067730  0.020144  0.013296 -0.055343  efectores
496  0.066323  0.094505  0.028592  0.044044  0.087972  0.036332  efectores
497  0.144762  0.086445 -0.024479  0.033327 -0.065997 -0.041003  efectores
498  0.065460  0.017706  0.072340 -0.001520  0.034559  0.046141  efectores
499  0.034823 -0.003672  0.003698  0.103003  0.059927 -0.042815  efectores

[500 rows x 14 columns]


Covarianza de auto cruzamiento (ACC)  hidro efectores nematoda dataset 4, con
valores atípicos.
Estadísticas.

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

```
mean     0.009902   -0.024740    0.026127    0.024025   -0.014302   -0.003167
std      0.089790    0.091976    0.093054    0.089174    0.086569    0.090854
min     -0.536985   -0.445911   -0.550900   -0.376130   -0.550519   -0.252429
25%     -0.038065   -0.075863   -0.032195   -0.021685   -0.065022   -0.053188
50%      0.011480   -0.023669    0.025080    0.017144   -0.015402   -0.005149
75%      0.062210    0.028089    0.083965    0.070647    0.034402    0.045987
max      0.445521    0.686959    0.310817    0.676982    0.377244    0.694214

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.023136    0.010315   -0.005134    0.007498    0.011051    0.006083
std      0.091083    0.086926    0.088597    0.086886    0.087315    0.084211
min     -0.557633   -0.286106   -0.523786   -0.435020   -0.531370   -0.331700
25%     -0.025086   -0.039258   -0.052371   -0.037669   -0.034072   -0.039888
50%      0.022350    0.014075   -0.002350    0.005987    0.010414    0.004558
75%      0.069443    0.058756    0.044221    0.053018    0.054627    0.044564
max      0.556925    0.619717    0.364580    0.651487    0.379292    0.637500

               X12
count  500.000000
mean     0.003485
std      0.087285
min     -0.526906
25%     -0.045764
50%      0.003520
75%      0.046564
max      0.516259
```

no_efectores
Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Valores del documento csv.

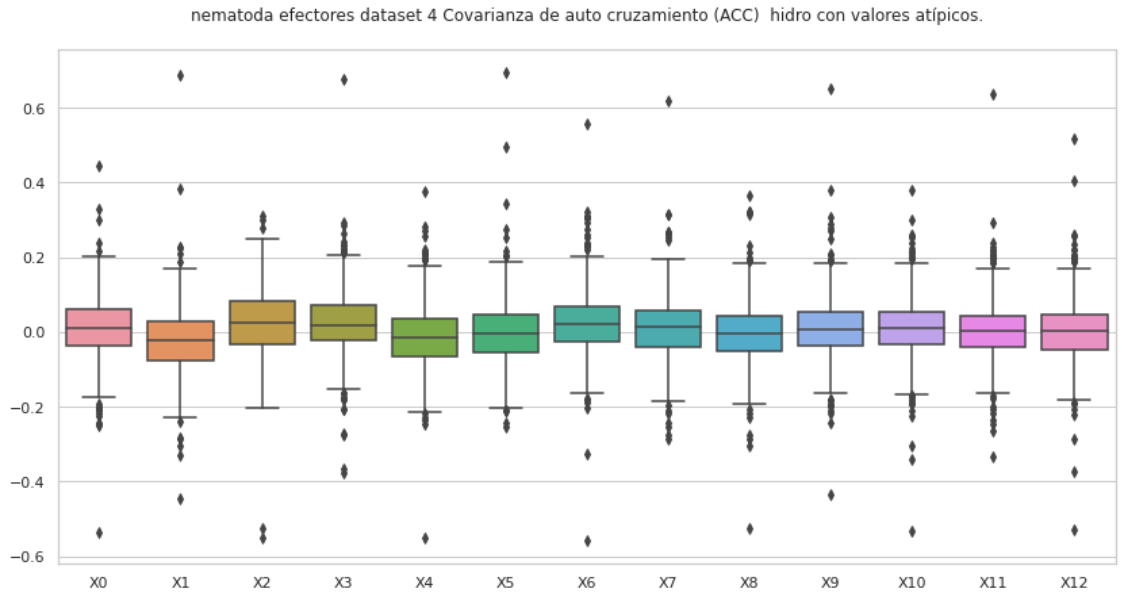
```
           X0        X1        X2        X3        X4        X5        X6  \
0    0.017217 -0.000186 -0.001335 -0.107420 -0.074637 -0.069802  0.124366
1    0.233717  0.009833 -0.172268 -0.140536 -0.137146  0.131456 -0.013840
2    0.031322  0.159508  0.006764  0.342762  0.175492  0.163503  0.026272
3    0.083595 -0.065101  0.130475  0.169807  0.053345  0.045782 -0.064739
4    0.050226  0.064606  0.124274  0.074895  0.042082  0.017217  0.071243
..        …         …         …         …         …         …         …
495  0.103152  0.099676  0.218625  0.001099  0.084220  0.031071  0.067768
496  0.165227  0.210263  0.180681  0.050588 -0.025762  0.177570  0.031750
497 -0.082277 -0.124967  0.017727  0.174228 -0.139317 -0.142785  0.069090
498 -0.052187 -0.047313  0.054709  0.052864 -0.078061 -0.031516 -0.036298
499  0.120675  0.019436  0.059216  0.067337 -0.000912 -0.008401  0.065489
```

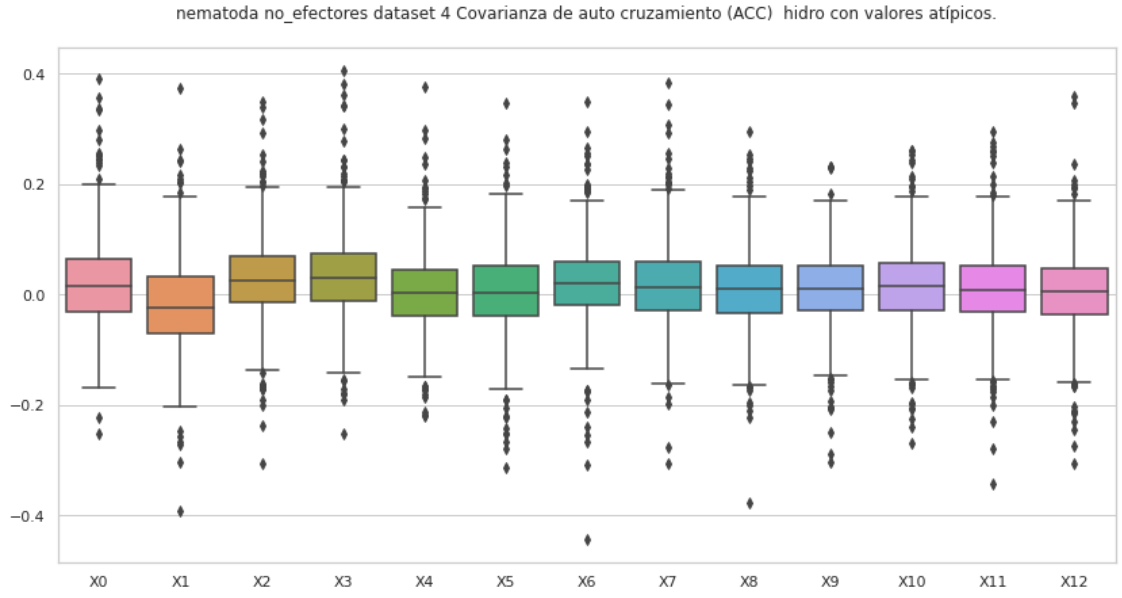
```
           X7        X8        X9       X10       X11       X12           X13
0    0.199445  0.048679 -0.083417  0.012377 -0.069264 -0.032600  no_efectores
1    0.101748 -0.012088 -0.053497 -0.073133 -0.133774 -0.052520  no_efectores
2    0.220003  0.067117  0.183416  0.066626  0.123090 -0.135925  no_efectores
3    0.081507  0.099614  0.101321  0.104649 -0.098531  0.052982  no_efectores
4   -0.021262  0.040934  0.021154  0.014439 -0.005251  0.007662  no_efectores
..        …         …         …         …         …         …             …
495  0.006993  0.006882  0.057183 -0.094742  0.138124  0.169704  no_efectores
496  0.118920  0.066244  0.060780  0.130516 -0.128408  0.142744  no_efectores
497  0.057416  0.078191 -0.205027  0.013801  0.163945  0.057994  no_efectores
498 -0.040099 -0.043500  0.088774  0.057586 -0.027883  0.035716  no_efectores
499  0.084956  0.027118  0.022774  0.027697  0.067439 -0.004438  no_efectores

[500 rows x 14 columns]
```

Covarianza de auto cruzamiento (ACC) hidro no_efectores nematoda dataset 4, con valores atípicos.
Estadísticas.

```
                X0          X1          X2          X3          X4          X5  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020376   -0.018781    0.028363    0.033883    0.003690    0.006360
std      0.084716    0.087225    0.079588    0.081671    0.076332    0.082921
min     -0.253155   -0.392002   -0.307104   -0.252623   -0.219664   -0.312871
25%     -0.031219   -0.069667   -0.013427   -0.012205   -0.037576   -0.038800
50%      0.016448   -0.023452    0.026043    0.029794    0.002082    0.004128
75%      0.064163    0.031836    0.069630    0.074227    0.045713    0.051047
max      0.390809    0.373044    0.348716    0.405606    0.376400    0.346569

                X6          X7          X8          X9         X10         X11  \
count  500.000000  500.000000  500.000000  500.000000  500.000000  500.000000
mean     0.020949    0.018922    0.009019    0.007596    0.014556    0.010868
std      0.079633    0.079026    0.074129    0.069484    0.075593    0.077284
min     -0.443690   -0.305882   -0.377733   -0.304305   -0.268777   -0.343650
25%     -0.018501   -0.029841   -0.034485   -0.028378   -0.028127   -0.030886
50%      0.019959    0.013157    0.010371    0.010427    0.014646    0.008224
75%      0.059777    0.058780    0.051855    0.051338    0.056568    0.052429
max      0.348763    0.384877    0.294713    0.230977    0.261206    0.294473

               X12
count  500.000000
mean     0.003918
std      0.075483
min     -0.306363
25%     -0.036185
50%      0.004481
75%      0.046407
```

```
max       0.359717
```


nematoda efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro con valores atípicos.
0.4
0.2
0.0
−0.2
−0.4
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

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 4, sin valores atípicos.
Valores del documento csv.

	X0	X1	X2	X3	X4	X5	X6	\
0	0.028711	-0.015801	0.072228	0.040547	-0.052455	-0.007357	0.020078	
1	-0.036599	-0.026153	-0.010764	0.001333	0.003238	-0.062025	0.037190	
2	-0.029981	-0.135378	0.036558	0.006781	-0.027149	0.026754	0.041512	
3	0.080662	0.073778	0.090974	0.039758	0.066878	0.073850	0.045422	
4	-0.027765	-0.067295	0.054629	-0.002677	-0.005438	0.000400	0.019138	
..	…	…	…	…	…	…	…	
495	-0.030087	-0.045489	0.033518	0.018402	-0.023900	-0.004714	-0.014491	
496	-0.137377	-0.023717	0.057842	0.172777	0.047206	0.013638	0.085346	
497	-0.013672	-0.087455	-0.050489	-0.041589	-0.097680	0.046167	-0.026502	
498	0.116427	0.012649	0.115212	-0.008380	0.126774	0.156099	-0.052903	
499	-0.055984	-0.167736	0.073530	-0.114125	0.018041	0.084930	0.006425	

	X7	X8	X9	X10	X11	X12	X13
0	-0.074102	-0.035452	0.081557	0.054416	0.027528	0.034677	efectores
1	0.027850	-0.025965	0.086861	-0.063821	-0.087188	-0.059452	efectores
2	0.047561	-0.050205	0.027008	0.020462	-0.080180	-0.039155	efectores
3	0.143714	0.082443	0.063353	0.124786	0.052447	0.083028	efectores
4	-0.075391	0.027096	0.026459	-0.005397	0.093696	-0.023226	efectores
..	…	…	…	…	…	…	
495	0.101363	0.092605	-0.067730	0.020144	0.013296	-0.055343	efectores
496	0.066323	0.094505	0.028592	0.044044	0.087972	0.036332	efectores
497	0.144762	0.086445	-0.024479	0.033327	-0.065997	-0.041003	efectores
498	0.065460	0.017706	0.072340	-0.001520	0.034559	0.046141	efectores
499	0.034823	-0.003672	0.003698	0.103003	0.059927	-0.042815	efectores

[463 rows x 14 columns]

Covarianza de auto cruzamiento (ACC) efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	463.000000	463.000000	463.000000	463.000000	463.000000	463.000000	
mean	0.009423	-0.024623	0.024517	0.022813	-0.014304	-0.004531	
std	0.080303	0.079013	0.080025	0.073229	0.076730	0.076903	
min	-0.251316	-0.287181	-0.170099	-0.206271	-0.235742	-0.252429	
25%	-0.037583	-0.073825	-0.029113	-0.018371	-0.062146	-0.048338	
50%	0.012774	-0.023517	0.021639	0.016720	-0.016878	-0.005230	
75%	0.059546	0.027348	0.075123	0.066016	0.032836	0.044175	
max	0.239365	0.227006	0.279668	0.263307	0.220609	0.252355	

```
                X6          X7          X8          X9          X10          X11  \
count  463.000000  463.000000  463.000000  463.000000  463.000000  463.000000
mean     0.020160    0.009529   -0.002949    0.002122    0.010485    0.003234
std      0.074072    0.073253    0.073952    0.071192    0.074038    0.070786
min     -0.203588   -0.215909   -0.226944   -0.242487   -0.211316   -0.236858
25%     -0.023641   -0.035769   -0.048071   -0.038203   -0.032497   -0.038445
50%      0.021917    0.013982   -0.002145    0.004354    0.010315    0.003197
75%      0.066690    0.054401    0.042617    0.049477    0.052295    0.042487
max      0.260110    0.267296    0.214444    0.247621    0.258373    0.219228

               X12
count  463.000000
mean     0.003839
std      0.073279
min     -0.207596
25%     -0.042374
50%      0.003814
75%      0.044983
max      0.259418



no_efectores
Covarianza de auto cruzamiento (ACC)  no_efectores nematoda dataset 4, sin
valores atípicos.
Valores del documento csv.

           X0        X1        X2        X3        X4        X5        X6  \
0    0.017217 -0.000186 -0.001335 -0.107420 -0.074637 -0.069802  0.124366
1    0.233717  0.009833 -0.172268 -0.140536 -0.137146  0.131456 -0.013840
3    0.083595 -0.065101  0.130475  0.169807  0.053345  0.045782 -0.064739
4    0.050226  0.064606  0.124274  0.074895  0.042082  0.017217  0.071243
5   -0.129447 -0.031576  0.035548  0.094640 -0.116466 -0.061568  0.198006
..        …       …       …       …       …       …       …
494  0.150806  0.042827 -0.014662  0.131200  0.103578  0.061071 -0.016260
495  0.103152  0.099676  0.218625  0.001099  0.084220  0.031071  0.067768
496  0.165227  0.210263  0.180681  0.050588 -0.025762  0.177570  0.031750
498 -0.052187 -0.047313  0.054709  0.052864 -0.078061 -0.031516 -0.036298
499  0.120675  0.019436  0.059216  0.067337 -0.000912 -0.008401  0.065489

           X7        X8        X9       X10       X11       X12           X13
0    0.199445  0.048679 -0.083417  0.012377 -0.069264 -0.032600  no_efectores
1    0.101748 -0.012088 -0.053497 -0.073133 -0.133774 -0.052520  no_efectores
3    0.081507  0.099614  0.101321  0.104649 -0.098531  0.052982  no_efectores
4   -0.021262  0.040934  0.021154  0.014439 -0.005251  0.007662  no_efectores
5   -0.114728 -0.040116  0.057862  0.174306 -0.146306 -0.048606  no_efectores
..        …       …       …       …       …       …           …
494 -0.035624  0.014646 -0.003017  0.057381 -0.078355  0.008595  no_efectores
```

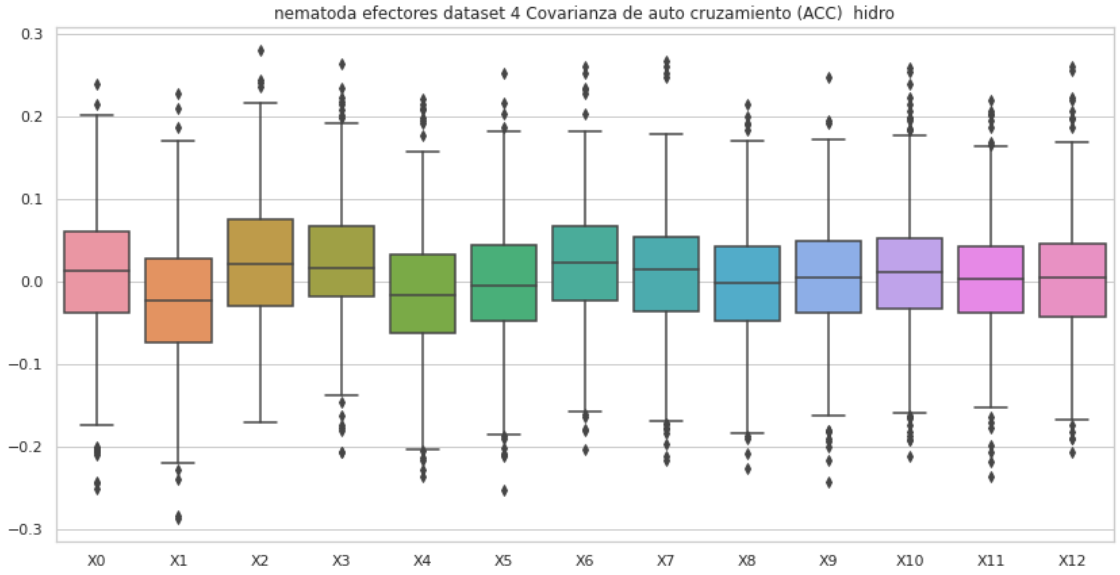
```
495  0.006993  0.006882  0.057183 -0.094742  0.138124  0.169704  no_efectores
496  0.118920  0.066244  0.060780  0.130516 -0.128408  0.142744  no_efectores
498 -0.040099 -0.043500  0.088774  0.057586 -0.027883  0.035716  no_efectores
499  0.084956  0.027118  0.022774  0.027697  0.067439 -0.004438  no_efectores

[455 rows x 14 columns]
```

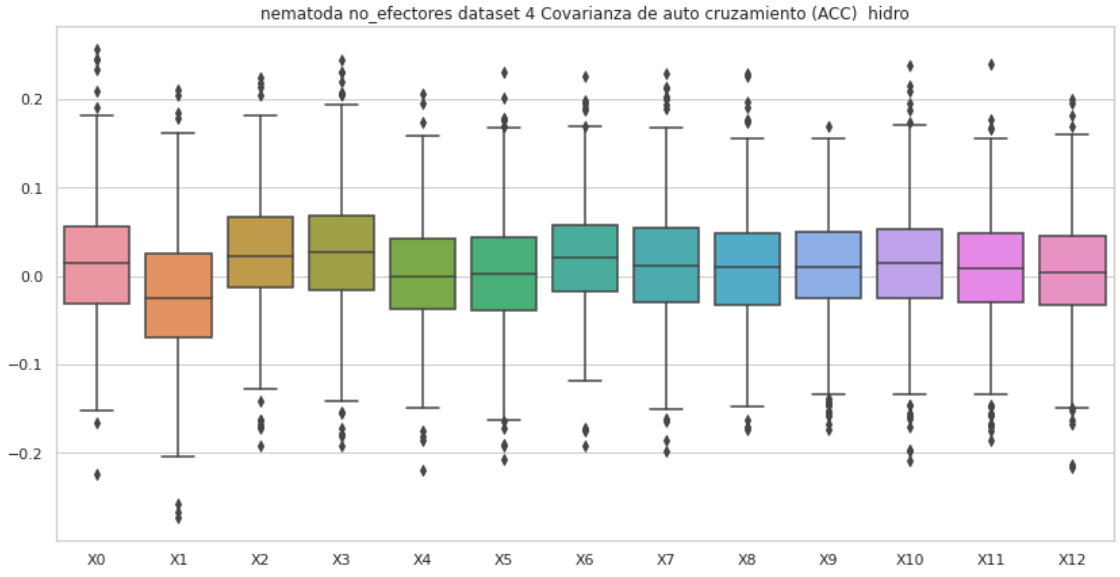
Covarianza de auto cruzamiento (ACC) no_efectores nematoda dataset 4, sin valores atípicos.
Estadísticas.

	X0	X1	X2	X3	X4	X5	\
count	455.000000	455.000000	455.000000	455.000000	455.000000	455.000000	
mean	0.013874	-0.022133	0.024795	0.026530	0.000524	0.004916	
std	0.071017	0.075885	0.068142	0.070219	0.065968	0.068838	
min	-0.223286	-0.272805	-0.191363	-0.191691	-0.219664	-0.206672	
25%	-0.030760	-0.069057	-0.013481	-0.015541	-0.036744	-0.038439	
50%	0.014904	-0.024846	0.022924	0.026693	-0.000200	0.002101	
75%	0.056341	0.025013	0.066148	0.068243	0.042337	0.044185	
max	0.256262	0.210263	0.224051	0.244489	0.206261	0.230426	

	X6	X7	X8	X9	X10	X11	\
count	455.000000	455.000000	455.000000	455.000000	455.000000	455.000000	
mean	0.021184	0.014908	0.007876	0.008572	0.012685	0.007255	
std	0.063131	0.066119	0.063163	0.059139	0.066736	0.064404	
min	-0.191009	-0.198155	-0.173587	-0.173855	-0.208512	-0.185373	
25%	-0.016788	-0.029227	-0.033309	-0.024554	-0.025599	-0.029047	
50%	0.020216	0.011694	0.009615	0.010398	0.014231	0.008009	
75%	0.057694	0.055202	0.047753	0.050081	0.053359	0.048528	
max	0.225940	0.228797	0.229244	0.168604	0.237843	0.239367	

	X12
count	455.000000
mean	0.005128
std	0.062900
min	-0.216006
25%	-0.032333
50%	0.004483
75%	0.044997
max	0.199644


nematoda efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12


nematoda no_efectores dataset 4 Covarianza de auto cruzamiento (ACC) hidro
0.2
0.1
0.0
−0.1
−0.2
X0
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12